Evidence Accumulation Model Insights Into Cognitive Processes

Prince Kouassi

A thesis submitted to the

Department of Psychology

City, University of London

For the degree of

Doctor of Philosophy

July, 2023

Declaration

I grant powers of discretion to the University Librarian to copy this thesis in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Acknowledgements

Firstly, I would like to deeply thank James Yearsley for his continued intellectual and moral support and encouragement throughout the duration of my PhD, especially during the Covid-19 pandemic. I would also like to thank Emmanuel Pothos for his support throughout my PhD and his many words of wisdom regarding my PhD and my future life endeavours. I would also like to thank City University and the relevant administrative team in the Psychology department, who have never failed to lift the administrative load off my shoulders when necessary.

I want to thank my family for their love and support throughout my PhD, especially my mother Caterina Pau, who has consistently comforted, praised and assured me.

I would also like to thank my fellow PhD students for have given me much appreciated and helpful advice during my PhD. I would also like to thank them for the many pleasant and unpleasant memories we have shared together.

Finally, I would like to thank myself, for being my biggest fan and critic.

"If there is no struggle, there is no progress."

- *Frederick Douglass*

Table of Contents

List of Figures

List of Tables

Abstract

This thesis covers eleven empirical studies across four topics, all relating to how the underlying information processing systems behind various behavioural effects and performances can be explored through a computational framework. Specifically, how an evidence accumulation model framework can be used to get insights into these behaviours. The thesis begins with a general introduction to the different topics and empirical studies that will be explored. After the general introduction, the second chapter explores whether the conjunction fallacy phenomenon can be transposed from its traditional descriptive scenario-based experimental context to the psychophysical domain, in order to produce datasets better suited to modelling procedures. Next an evidence accumulation model is applied to reveal the exact information processing structure that is argued to be responsible for producing the conjunction fallacy. The third chapter introduces another cognitive effect known as the interference effect. A series of experiments then explore the extent of this effect. The argument is then made that this effect functions as a constraint on theories and models attempting to explain non-normative behaviour, such as evidence accumulation models. The subsequent chapter extends the evidence accumulation model from chapter two, to the interference effect observed in chapter three. This chapter goes on to propose that a simple process account of the interference effect may provide an equally plausible explanation for this effect. The fifth chapter extends the evidence accumulation model framework to a visual search task. This is done to explore how evidence accumulation models provide insights into the main performance drivers of a task. This chapter proposes that such models can function as an additional layer of analysis, to more deeply understand performance drivers, even in tasks with simple objectives. Finally, a conclusion on the main point throughout this thesis is presented.

Chapter One: General Introduction

The literature on decision-making has shown that although a decision maker can be rational and make logically sound decisions, in the large majority of situations they behave irrationally (Kahneman & Tversky, 1979; Pothos, Waddup, Kouassi, & Yearsley, 2021; Tversky & Kahneman, 1992). These studies have shown that in situations ranging from medicine, finance, education, politics and everyday life, individuals make decisions that are frequently prone to errors and biases (Albar & Jetter, 2009; Dawson & Arkes, 1987; Stanovich & West, 1998). One example of these ubiquitous instances of errors and biases in decision-making is the conjunction fallacy.

The conjunction fallacy (CF) is a judgement bias that occurs when a decision-maker estimates the probability of the conjunction of two events to be greater than the probability of either of the conjuncts (Costello, 2009; Hertwig & Gigerenzer, 1999; Moro, 2007; Amos Tversky & Kahneman, 1983). The literature has equally shown this bias to be an extremely robust phenomena occurring over several domains (Moro, 2007). Despite the CF's robustness and the extensive empirical work conducted to understand the scope of the bias, assessing the underlying cognitive processing mechanisms has proven more difficult (Moro, 2007). This is because although the CF is a violation of the conjunction rule in probability theory, as experimental stimuli, it is instead presented as a descriptive scenario-based task. In these tasks, participants have to rank statements about descriptions according to their likelihood (Costello, 2009; Amos Tversky & Kahneman, 1983). Consequently, the experimental paradigm does not lend itself to exhaustive analytical procedures aimed at analysing underlying cognitive processes. One such class of procedures that do, however, is cognitive models.

However, cognitive modelling procedures that aim to provide insight into the different features of cognitive processes underlying decisions and judgments relay on more exhaustive datasets than the ones provided by descriptive tasks. Nonetheless, from a more general modelling perspective, there are other reasons why a descriptive scenario-based task is flawed. Firstly, the various conjunct and conjunction probabilities in the task are represented by written descriptions. This makes it difficult to determine how participants themselves represent these descriptions as probability. The vagueness in the way these written descriptions are interpreted as probabilities requires a separate modelling procedure in itself. Additionally, from a modelling perspective the amount of degrees of freedom allowed in changing the descriptive scenario without changing the fundamental probability structure underlying these descriptions is limited. As such, multiple trials with a similar scenario structure are not presented to participants. Essentially, you cannot present the descriptive scenario multiple times to get improved estimates. Instead, participants are presented with one-shot tasks. For the majority of modelling procedures, relatively large datasets are required with multiple iterations of the same task (Guest & Martin, 2021; Van Rooij & Blokpoel, 2020). Consequently, the experimental paradigm used to elicit and capture the CF must be fundamentally changed in order to gain insights from standardised modelling procedures. Creating an experimental paradigm that reduced putative undesirable higher-order effects from scenario-based tasks and allowed for larger datasets to be collected to better suit model fitting procedures (relevant to information processing models), were the main motivations for a new experimental paradigm outlined in this chapter. The CF provides evidence of one way in which the assignment of probabilities to conjunctions of events cannot be reduced to a process involving assignments of probabilities to individual events, and then combining these using classical logical operations. The CF is a particularly stark example of this, however presumably a more careful analysis, with a larger set of probability judgments, could demonstrate inconsistencies with classical probability theory without the need to observe a CF. In this way, another purpose of translating these tasks into a psychophysical domain is to allow us to explore a wider class of probability assignments that may provide evidence of reasoning incompatible with classical probability without necessarily demonstrating s full blown CF.

As such, the aim of the second chapter is to assess whether the CF task can be transposed to a different domain, in order to produce more constrained and larger datasets that better suit cognitive modelling. Attempts are made to move the CF task to the psychophysical domain in order to produce more iterations of the task and consequently larger datasets based on accuracy and response time (RT) data. A robust class of RT models are subsequently used to provide insights into the underlying cognitive processes of the CF phenomenon. Specifically, evidence accumulation models (EAMs) are the main cognitive models applied. Please refer to Figure 1.1 below for an illustration and brief introduction to EAMs. This is because of their conceptual simplicity and robustness in being able to very accurately capture behaviour in speeded RT tasks across a variety of domains (Evans, Dutilh, Wagenmakers, & van der Maas, 2020; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998).

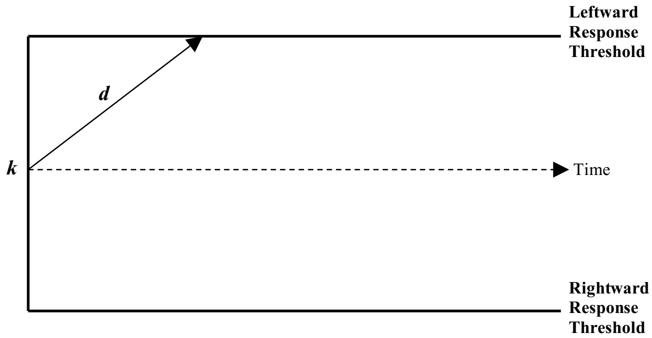

Figure 1.1. A simplified illustration of a binary response EAM and its main components. When a participant begins a binary response trial they have a starting amount of evidence in support of either response. This is represented as *k* in the figure and is assumed to be at equal distanced between both responses, such that the participant has no bias towards either response at the beginning of a trial. As the trial progresses the participant accumulates evidence in support of either response at some fixed rate, represented as *d*. Evidence here refers to any information derived from the task stimuli in support of either binary response. When the amount of accumulated evidence in support of either response reaches one of the two defined response thresholds, a response is triggered. In the above figure, the amount of accumulated evidence (the solid arrow) reaches the leftward response threshold and consequently triggers the leftward response in the task, e.g. the participant presses the left instead of the right response key.

I will go on to show that the classical CF task can be successfully transposed to the psychophysical domain, to produce datasets better suited to cognitive modelling procedures. I will subsequently show that EAMs can be fitted to these datasets to reveal that a bias in a specific processing order of conjunct probabilities appears to be responsible for the conjunction fallacy.

The most popular approach to understanding human decision-making behaviour has been based on the laws of classical probability theory (CPT) (Crupi, Fitelson, & Tentori, 2008; Franco, 2009; Reyna & Rivers, 2009). These laws assume that individuals behave as rational agents and make logically sound decisions (Gigerenzer & Gaissmaier, 2015). However, a large body of work over the last few decades has revealed that decision-makers are largely irrational at times and tend to make decisions that explicitly violate the laws of CPT. A classic example of this is the aforementioned CF. CPT proposes no direct explanation for the violation of this law within its framework. Instead, extensions within CPT are created to accommodate any violation within the framework, such as the EAM presented in chapter two to explain the underlying causes of the CF. However, using an alternative probability theory framework, quantum probability theory (QPT), instances of irrational behaviour can be logically expected and even predicted.

QPT is an alternative probability theory based on the mathematics of quantum physics (Pothos & Busemeyer, 2009, 2013, 2022; Pothos et al., 2021). As a probability theory applied to human cognition, it has been successful in accurately predicting multiple instances of irrational behaviour that are unobservable in CPT, such as the CF (Pothos et al., 2021). However, using an alternative probability theory, QPT, interference caused by the incompatibility in the original CF task questions can lead to a CF itself. Incompatibility in QPT means that two questions cannot be resolved concurrently. The decision-maker has to resolve one question after the other and resolving one question creates uncertainty for the other. Interference here refers to the way the evaluation of one question impacts the evaluation of subsequent questions. A CF can be computed in QPT through this interference which functions as a sort of order effect, but it needs to have a sequential form for incompatible questions.

In the third chapter I show how interference can be introduced between two disjoint scenarios to directly manipulate the assigned possibilities of various disjunctions by decision-makers. Furthermore, I show how interference can be used as a constraint on normative and non-normative decision-making theories. That is, how CPT makes no assumption or prediction of interference occurring in a disjunction involving pairs of scenarios (two-way interference), versus QPT that predicts interference. Additional experiments are then presented to show that interference can be manipulated to be either positive or negative, to in turn manipulate the degree of disjunction probabilities in decision-makers.

Finally, an argument is made that although CPT cannot account for the interference observed in the presented results, current extensions of CPT that are supposed to accommodate non-normative behaviour are also inadequate. Results are also presented that show how CPT and its currents extensions further fail at accounting for interference in a disjunction involving three scenarios (three-way interference).

Just as interference can be used as a constraint on normative and non-normative theories of decision-making, it can also be used as a constraint on EAM accounts. In chapter two an EAM is argued to be able to capture the CF and represent the bias in cognitive processes underlying the CF itself. However, interference as explained through QPT shows that classical approaches to explaining human decision-making behaviour are constrained by their assumption of interference. Results by Kvam, Pleskac, Yu, & Busemeyer (2015) have shown that this constraint can also be extended to EAMs. As such, if EAMs fail to account for interference occurring in a standard EAM paradigm where interference is indeed found to occur, this would further support the argument that QPT is the most appropriate framework for capturing genuine human decision-making behaviour. Additionally, the serial biased start point (S-BSP) model described in the second chapter describes the CF as being a result of some interference in the serial evidence accumulation process. The next step for this model is to assess what other non-normative decision-making effects it can be extended to. As the S-BSP model gives a serial account of non-normative decision-making behaviour, trying to capture non-normative behaviour that occurs due to a serial processing account would be the logical next step for this model. One example of this is the suggested quantum interference effect reported by Kvam et al. (2015). The fourth chapter in this thesis explores these points further.

The results reported by Kvam et al. (2015) are argued to show that an experimental quantum EAM paradigm can be created to produce interference between serial decisions. Additionally, it is argued that standard EAMs fail at being able to capture this effect. In this fourth chapter I focus on the interference effect reported by Kvam et al. (2015) in the presented quantum EAM paradigm. Issues with extrapolating the original experimental paradigm to standard EAMs are initially identified. Furthermore, the argument is made that the original experimental paradigm does not represent or capture the main features of standard EAMs. As such, a more suitable experimental EAM paradigm is proposed and tested to yield results that are largely consistent with the experimental results reported by Kvam et al. (2015). Suggesting that a simpler processing account may be sufficient in explaining the reported quantum effect.

Kvam et al. (2015) initially proposed an EAM that functioned within a QPT framework to explain the effect of interference that is argued to occur within the evidence accumulation process. The researchers argued that the main feature of EAMs, the way in which information is processed or “accumulated”, occurs in a “quantum” way. This alternate way of processing information is what permits the occurrence of an interference effect. I go on to show how the results that identify an effect of interference can be captured by extending standard EAMs in a similar way to what was proposed in chapter two. I propose that the extended EAM presented in chapter two that captures the CF can also capture the effect of interference observed by Kvam et al. (2015). I further argue that the effect can be captured by a non-quantum processing account. Similar to the cause of the CF presented in chapter two, the extended standard EAM shows that a bias or error in the processing order of the task questions can produce results congruent with an effect of interference. I infer that the proposed model shows that a simple processing account can also capture the results observed in the original paper, in a similar way to the quantum account.

A follow-up experiment is proposed with an experimental paradigm that still represents the main features of an EAM, but more closely aligns with the design of the original experiment. The rational for this second experiment was based on the idea that observing interference effects was dependant on participants having a limited amount of time to process the task information. This was different to the first experiment that allowed participants to respond to the task at their own discretion and therefore had an unlimited amount of time to process task information. The results from this experiment were largely consistent with the first experiment. A final experiment that largely replicates the original is conducted. The results of this experiment were nonetheless consistent with the pattern of results found in the first experiment.

I conclude this chapter by arguing that although standard EAM accounts of cognitive processing appear to overcome the quantum constraint on decision-making, they do not rule out a QPT approach. A quantum account of interference effects in a serial decision task may not be the only plausible explanation for the observed results. Instead, the results observed in the original experiment may be due to unexpected results brought on by response priming, due to the response options in the various conditions. Nonetheless, quantum models and classical models such as the extended EAM in chapter two can provide different, but equally compelling, accounts of these effects. I therefore argue that it is sensible to pursue both approaches in parallel.

The final chapter focuses on how the different features of EAMs can be used to represent and capture other cognitive phenomena, such as the features controlling visual search strategies. The literature on visual search tasks has revealed that visual search strategies remain largely varied across a host of different visual search tasks (Eckstein, 2011; Timmis, Turner, & Van Paridon, 2014; Wolber & Wascher, 2003). However, the literature has also shown that within a particular subset of inter-task visual searches (e.g. multiple patch foraging tasks), search strategies can be found to converge, versus intra-task visual searches (Boot, Kramer, Becic, Wiegmann, & Kubose, 2006). An initial argument is proposed that visual search strategies are driven by speed-accuracy trades-offs that are themselves controlled by properties best defined by EAMs. Specifically, that the underlying speed-accuracy trade-offs are controlled by participants' response conservativeness. The exact trade-off utilised by participants (whether participants want more accuracy vs speed or vice versa), is set by a participant's response conservativeness. Refer to Figure 1.2 below for an illustration of how the speed-accuracy trade-off is represented by EAMs. Therefore, from an EAM perspective, search strategies when there are multiple foraging regions are exactly the same as a speed accuracy trade-off for a single region. When search strategies are found to converge onto one dominant search strategy, the argument is proposed that the speed-accuracy trade-off still provides a more comprehensive account of task performance. Furthermore, that this shows that task performance is not dependant on search strategy, but the speed-accuracy trade-off. Nowhere is this more evident than in intra-task visual searches, where search strategies are found to not converge.

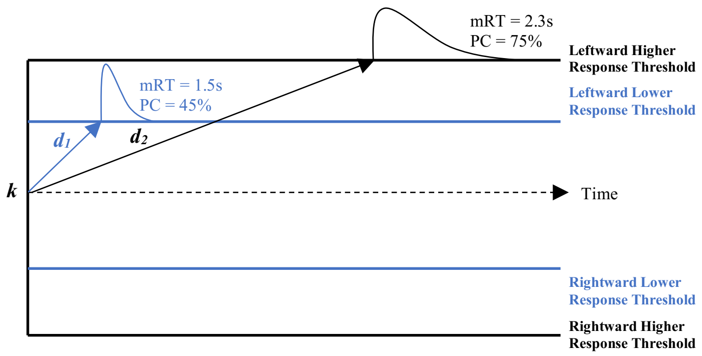

Figure 1.2. In the above example EAM there are two separate rates of evidence accumulation, d1 and d2. Each accumulator represents a different participant in this example. The horizontal blue lines represent the lower response thresholds for the first participant. The horizontal black lines represent the higher response thresholds for the second participant. At the end of each evidence accumulation process (the arrows) the curves show the distribution of RTs for similar d_1 and d_2 values, but different response threshold values. Associated hypothetical mean RTs (mRT) and accuracy rates (PC) are provided for illustrative purposes. The figure shows how an EAM represents a speed-accuracy trade-off, assuming that the rate of evidence accumulation is constant, but the response threshold (response conservativeness) varies.

The initial argument within this chapter proposes that while performance on visual search tasks is viewed as being driven by speed-accuracy trade-offs, an EAM perspective provides the most robust method through which to model and capture the feature responsible for controlling this trade-off. That is, response conservativeness. Therefore, the EAM perspective provides a testable assumption: under conditions where visual search strategies do not converge (intra-task searches), EAM parameters representing the speed-accuracy trade-off will show a clear pattern of manipulation.

The literature on optimal decision-making behaviour has additionally shown that while decision-makers are largely irrational, there are instances of optimality (Mcnair, 1982; McNamara & Houston, 1985; Nowakowska, Clarke, & Hunt, 2017). In these instances, participants can behave within a range that is statistically optimal across a variety of visual search tasks (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Najemnik & Geisler, 2009). In particular, in some visual search tasks participants have been able to demonstrate statistically optimal behaviour (Najemnik & Geisler, 2009). A final argument is proposed that within tasks with no convergence of search strategies, participants' speed-accuracy trade-off will not only show a clear pattern of manipulation, but will also reflect different levels of optimality.

I present an initial intra-task foraging experiment that is divided into two segments: participants were instructed to emphasise accuracy in one segment and speed in another segment. This was done to directly manipulate participants' speed-accuracy trade-off in a task not driven by search strategy. The aim was to assess whether manipulating speed-accuracy trade-offs caused corresponding changes in overall task performance. Initial results show that attempts at getting participants to emphasise different aspects of their speed-accuracy trade-off were unsuccessful. As such, a follow up experiment was conducted. Three different age groups were recruited (young, middle-aged and senior) to complete the task in experiment one, but without an emphasis on accuracy or speed. I argue that given the literature shows that as participants age their response conservativeness increases, the inherent characteristics of each age group should result in a different emphasis on the speed-accuracy trade-off (Starns & Ratcliff, 2010). This should therefore result in inherently distinct trade-offs.

The results show that the different age groups had inherently distinct trade-offs that corresponded with distinct performance patterns and levels of optimality. Additionally, EAM fits to the data reveal that the different age groups were represented by different levels of response conservativeness that were reflected in the speed-accuracy trade-off set by each group. The results are argued to show that, as expected, the different age groups made different trade-offs between speed and accuracy when searching a single region and that this single factor was also enough to explain differences in search strategy when searching multiple regions. Furthermore, in some cases, the exact visual search strategy or strategies used is largely dependent on the trade-off set by participants. Finally, the clinical implications of these results are discussed. Specifically, how the modelling results reveal decision-making behaviour in more senior individuals long thought to correspond to cognitive decline, instead correspond to intentional changes in response conservativeness.

Chapter Two: Modelling the Conjunction Fallacy in the Psychophysical Domain

Section 2.1 If Linda was a Square Grid

Introduction

Experimentally, people display apparent failures to reason rationally in various contexts. One celebrated example of such a failure is the *conjunction fallacy* (CF), whereby people violate a fundamental law of classical probability theory: the *conjunction rule* (Tversky and Kahneman, 1983). Mathematically, the conjunction rule states that the probability of a conjoint hypothesis, p(A&B), cannot exceed the probability of either of its constituent probabilities: p(A) or p(B). That is, $p(\text{A\&B}) \leq \min(p(\text{A}), p(\text{B}))$. Tversky and Kahneman (1983) showed that people violated the conjunction rule when ranking the likelihood of conjunct and conjunction statements about a particular character named Linda. Participants were presented with a scenario-based description of a women called Linda. Several character features of Linda were provided; statements were given saying Linda is active in the feminist movement and Linda is a bank teller. Participants then ranked conjunct and conjunction statements based on these features in order of likelihood: i.e. is it more likely that Linda is only a bank teller (conjunct probability), or a bank teller and feminist (conjunction probability)? Results displayed a CF by participants consistently ranking the conjunction statement as more probable than the conjunct statement.

The original explanation for participants committing the CF as observed by Tversky and Kahneman (1983) was the representativeness heuristic. The proposal is fairly intuitive and consists of the assumption that, instead of making probability judgments, participants consider the similarity/representativeness between the given instance (Linda) and the offered categories. Notably, Linda is hardly considered to be representative of the category of bank tellers, however, she would be seen as highly representative of the category of feminists and bank tellers in the descriptive task – hence, what appears to be a conjunction fallacy is, instead, a judgment concerning differences in representativeness.

In later work, there were attempts to explore in more detail similarity and the emergence of conjunction fallacies, with some supporting evidence (e.g., Smith et al., 1994). However, overall, the representativeness heuristic has been criticised by being fairly unspecific: without a particular model for computing representativeness/similarity and, given how flexible such judgments can be, it is hard to see how it can offer a principled account for the conjunction fallacy. Additionally, the representativeness heuristic has somewhat limited scope, in that it is unclear how it can apply to conjunction fallacies, which are the result of a particularly strong causal connection between the two conjuncts. To employ the example from the original Tversky and Kahneman (1983) work, the applicability of the representativeness heuristic is dubious in cases such as Prob(john has had a heart attack and is older than 50) vs. Prob (john has had a heart attack). Indeed, in four fairly well-known new proposals for the conjunction fallacy, the representativeness heuristic is not offered as a viable model: Tentori et al. (2013) note "To begin with, the representativeness heuristic was not advocated by Tversky and Kahneman as providing a general explanation of conjunction fallacy effects" and "Even in its intended domain of application, the representativeness account has met a remarkable degree of motivated caution and criticism. According to a recurrent complaint in the literature, the main limitation of the notion of representativeness, undermining its explanatory scope, lies in its broadly informal and fuzzy characterization." Zhu et al. (2020) do not mention it even once. Busemeyer et al. (2011) mention that their quantum model can be seen as a formalization of the representativeness heuristic, but without much additional detail. Finally, Costello and Watts (2014) note "Although the representativeness heuristic remains the routine explanation of the conjunction fallacy in introductory textbooks, a number of experimental results give convincing evidence against this account." These authors proceed to summarize several experimental results at odds with representativeness, even given the vague and unclear formulation of this account.

Since Tversky and Kahneman's (1983) seminal work showing that people can violate the conjunction rule in the Linda problem, a wide breadth of experimental work has focused on exploring the robustness of this violation (Tentori, Bonini, & Osherson, 2004). This resulted in a large body of literature concerned with reducing CF rates by reformatting the wording and presentation of the CF scenario, and the kind of probability judgements that participants had to make based on the scenario information (Hertwig & Gigerenzer, 1999; Tentori et al., 2004). More recent work has focused on attempting to define the CF as a result of some form of erroneous information processing of the scenario representing the conjunct and conjunction probabilities (Moro, 2007). This is argued to occur even in the presence of correctly understanding the probability judgements required to be made. For instance, some work has focused on reducing CF rates by emphasizing the independence between features within a scenario (Ahn & Bailenson, 1996; Maguire, Moser, Maguire, & Keane, 2018). The assumption is that information processing has a predilection for assuming subjective uncertainty (a tendency to interpret some causal relationship between scenario features), which causes the conjoint hypothesis to appear as the most probable answer. Overall, the experimental focus on exploring the underlying causes associated with committing the CF has changed since Tversky and Kahneman's (1983) initial findings on the cognitive fallacy.

However, the general method of determining whether people have violated the conjunction rule and committed a CF has remained largely untouched. Specifically, the CF has primarily been defined as a violation of the conjunction rule through some *descriptive* scenario-based task (Hertwig & Gigerenzer, 1999; Yearsley & Trueblood, 2018). Let us use Tversky and Kahneman's (1983) original scenario which facilitated the CF as an illustration. Although this method has been highly effective in initially detecting a CF, it functions less effectively as an exhaustive method of critically analysing the CF itself. Firstly, tasks based on descriptive scenarios are high level cognitive tasks, which are themselves susceptible to heuristics and biases in interpretation through strong framing effects (Stanovich & West, 1998). These are effects which themselves can facilitate misinterpretation and thereby erroneous perceptions of causal links between scenario features. Secondly, conventional frameworks do not allow for multiple stimulus feature combinations and involve lengthy descriptions representing probabilities. This largely limits the total number of trials which can be presented. This poses a further problem for the application of in-depth response time (RT) modelling, which can explore the underlying processing of the CF (Stafford, Pirrone, Croucher, & Krystalli, 2020). Therefore, determining how best to isolate unique CF-eliciting effects and facilitating extensive RT modelling are fundamental in critically analysing the CF as a robust phenomenon.

One method of tackling these problems is by applying a psychophysical framework to determining the occurrence of the CF. This new approach I propose to minimising these higher-level effects involves transferring the traditional CF task to an entirely different domain, that is not limited by the same contextual problems as the original task. Such an approach demands a more simplified representation of the conjunct and conjunction probabilities, in order to determine a more direct relationship between the stimulus representation of the probabilities and perceptions of it. This thereby requires a more rudimentary representation of probabilities and reduces the problem to a low-level cognitive task; an additional consequence is that this will reduce unwanted high-level cognitive effects associated with misinterpreting the task. One way of simplifying the representation of the conjunct and conjunction probabilities is by transposing their representation to another stimulus domain. This allows for an assessment of the domain generality of the CF phenomenon. Therefore, one aim of this paper is to demonstrate that a CF eliciting task can be viably transposed from the descriptive domain, into the psychophysical domain.

The psychophysical task proposed here presented participants with one or two square grids, each made up of a certain proportion of blue-to-orange squares. Please refer to Figure 2.1 below for an illustration of the stimuli. On each trial participants were tasked with determining whether they agree or disagree with a presented question, based on the proportion of blue-to-orange squares in one or two presented grids. For example, on trials where two grids are presented (paired grid trials), the trial question may be: are there more blue than orange patches in both grids? In this example blue is the target colour. Let us assume that one of the two grids in this example has more blue than orange squares in it and the other paired grid has more orange than blue squares in it. These specific trials are called CF-eliciting trials, because they are intended to elicit a CF response. Participants can agree with the conjunction position (both grids have more of the target colour) and in effect commit a CF when both grids in fact do not have more of the target colour. Alternatively, participants can disagree with the conjunction position (only one grid has more of the target colour), by answering no to the question in the trial and not commit a CF when both grids in fact do not have more of the target colour. Additionally, participants were presented with a single square grid on other trials (single grid trials). Again, participants are tasked with determining whether they agree or disagree with a presented question, based on the presented grid: e.g., are there more blue than orange patches in the grid? In this example, the target colour is again blue. Participants can agree (the grid has more of the target colour) or disagree (the grid has less of the target colour) with the conjunct proposition.

We can map the different stimuli and judgments in this task to a prototypical CF task such as the Linda problem in the following way (we will justify this in more detail below): The grid containing more blue than orange can be thought of as mapping to the property 'Is a Feminist', so that blue represents positive evidence and orange negative. In a trial involving just this grid the correct, and typical, response is to agree with the statement 'Does this grid contain more blue than orange?', which is the analogue of the question 'Is Linda a feminist?' In a similar way the grid containing more orange than blue can be mapped to the property 'Is a Bank

Teller', so that when presenting only this grid the correct, and typical, response is to disagree with the statement 'Does this grid contain more blue than orange?', which is the analogue of the question 'Is Linda a Bank Teller?'

As we have implied, typically when presented with a single grid, participants can correctly identify whether they contain more blue than orange patches. The trials in which two grids are presented at once are the analogue of the conjunction questions 'Is Linda a Feminist AND a Bank Teller?' In the usual Linda task a participant should answer yes to this question iff they judge than Linda is a Feminist AND Linda is a Bank Teller. For the case of the coloured grids, the conjunction question is "Are there more blue than orange patches in both grids?" Participants should respond yes if they judge that each grid separately contains a greater proportion of blue than orange patches.

Even though the present psychophysical task does not involve representativeness, it involves other features which are commonly associated with the emergence of conjunction fallacies. Notably, there is a conjunction comprised of a likely and unlikely event and the conjunction is established by comparing the conjunctive probability against the less likely marginal – I further explain this below.

There is an additional question of whether the perceptual judgments in this task are equivalent (in some sense) to the probabilistic inference judgments, typically employed in CF studies. That is, can we assume that participants assign likelihoods in response to questions about grid colours (and incomplete corresponding evidence, since we cannot assume that participants count all coloured grids individually). I argue that this is the case. As such, given a typical trial question and that a likelihood judgement must be made on two grids presented simultaneously on a trial, the likelihoods being assessed are:

1) both grids possess more of a target colour (the conjunction),
2) only one of the two grids has more of a target colour (the conjunct).

Therefore, what occurs during the processing of such questions and stimuli is a combination of probabilities, including as required to compute the probability of the conjunction.

Assigning probabilities in the present task is a general point concerning the nature of perceptual judgments. Specifically, it is suggested that there is a difference between these two questions (broadly speaking and of course allowing for variations along the lines below): First, “what is the likelihood of a patch being (>50%) a particular colour?”. Second, “is a patch a particular colour?”. One position is that these two judgments are different in nature, that is, they map onto different cognitive processes, with the former one corresponding to a question relating to probabilistic reasoning and the second question a perceptual judgment. If this is true, then the experiments in this chapter do not really tell us much about probabilistic reasoning and the assumed equivalence between the present paradigm and the Linda one (and conjunction fallacies in probabilistic reasoning in general) is wrong. The alternative possibility – the one I have assumed throughout this chapter – is that there is no essential difference between these two questions. That is, when we basically make any perceptual inference, even fairly trivial ones, such as, for example, “this pen is black”, we are implying a probabilistic judgment, that is, really, what we are saying is that “I think it is highly probable that this pen is black”.

As far as I know, there is no direct evidence between this assumed equivalence. However, there are models and theories indicative of the validity of my assumption. First, there have been proposals of influential categorization models, such as Anderson’s rational model of categorization, which essentially assume that category decisions are probabilistic. To frame in a way which is suggestive of my own position too, we do not really make judgments along the lines, for example, “this animal is a horse”, but rather “there is a high probability that this animal is a horse” (Anderson, 1991). Second, sequential sampling models have been employed both for perceptual judgments and decision-making ones. The latter kinds of tasks are closer to probabilistic inference. Even though, as far as I know, there are no published proposals of probabilistic reasoning based on sequential models, in fact there is some recent work by Busemeyer and colleagues, scheduled to appear later on this year (Busemeyer et al., in press). The fact that both perceptual judgments and probabilistic decisions can be modelled within the same framework is indication (though of course not proof) that the corresponding cognitive processes are similar. Finally, the quantum framework has allowed very similar proposals both for perceptual judgments (e.g., judgments concerning the similarity between colour patches, Epping et al., 2023) and, of course, probabilistic ones (e.g., Busemeyer et al., 2011). All these sources of evidence suggest that it may be a valid perspective to understand perceptual judgments in probabilistic terms. As noted, I do acknowledge that ‘hard’ proof of such an equivalence is lacking and that it might be the case that future work undermines my interpretation of the experiments in this chapter as equivalent (in the sense outlined elsewhere) to the conjunction fallacy.

To reiterate the points above, arguably, the ‘key’ feature of traditional CF paradigms, as exemplified in the Linda task, is that there is an unlikely characteristic for Linda (that she is a BT) paired with a very likely one (that she is a F). The combination of an unlikely and a likely characteristic in a conjunction is generally one of the ways in which a CF can emerge (relative to the marginal for the unlikely characteristic). The present psychophysical task also possesses these features. On specific trials, where two grids are presented simultaneously, there is an unlikely characteristic for the square grids in a trial (one of the two grids has less of the target colour and low visual discriminability) and a very likely characteristic (the other grid has more of the target colour and has high visual discriminability).

Probabilities and error rates

Participants’ answers were recorded in terms of error rates to the questions about colour proportions. Let’s consider the case of CF-eliciting trials. Such trials were composed of a pair of grids, so that one grid was easy and the other hard. The easy and the hard grids had colour proportions in the opposite direction, so that if the easy grid had more blue, the hard would have more orange and vice versa. CF-eliciting trials would be paired with single trials with just the easy grids. An example would be as follows:

(a) Do both grids have more orange than blue?
(b) Does the grid have more orange than blue?

For the (a) question, the correct answer is no and for the (b) question the correct answer is yes. These questions can be rephrased as:

(a’) What is the probability that both grids have more orange than blue?
(b’) What is the probability that the grid has more orange than blue?

Across the participant sample, we can assume that the proportion of yes/no responses to the a,b questions corresponds to the probabilities in the a’,b’ ones. The logic is that, for example, the more participants respond with a yes question to (a), the higher the implied probability that it is indeed the case that both grids have more orange than blue. Note, the approach of equating across participant patterns with within participant biases is an assumption, though a fairly common one in experimental psychology.

One final recasting of the questions is needed before the association with the conjunction fallacy set up becomes clear:

> (a'') What is the probability that participants make an error when responding that the left grid has more orange than blue *and* make an error when responding that the right grid has more orange than blue?
> (b'') What is the probability that participants make an error when responding that the single grids have more orange than blue?

Clearly, question a'' is a conjunction such that one of the conjuncts is question b''. Moreover, we expect a low probability for b'' (this question can be considered equivalent to the bank teller one in the Linda set up) and a high probability for a'' (this question can be considered equivalent to feminist and bank teller one). Therefore, if the error rate in a'' is higher than the error rate in b'', then we can take this as evidence that there is a conjunction fallacy. This can be represented by the following inequality:

$P_{ic} > S_{ic}$……………………………………………………………………………………(2)

where P_{ic} is the proportion of incorrect responses on CF-eliciting paired grid trials and S_{ic} is the proportion of incorrect responses on hard single grid trials. When a participant answers yes on CF-eliciting paired grid trials they believe that the probability of both grids having more of a target colour (conjunction) is greater than only one of the grids having more of the target colour (conjunct). As error rates are assumed to be substantially lower when assessing easy single grid trials, it is expected that error rates for CF-eliciting paired grid trials ought to be significantly higher, given that they possess a paired hard grid. However, it is an unexpected result to find that error rates on CF eliciting paired grid trials are significantly higher than error rates for hard single grid trials. As such, it is this inequality in the present task that defines the presence of a CF. Therefore, if error rates for CF-eliciting paired grid trials, P_{ic}, is greater than for hard single grid trials, S_{ic}, a CF has been committed. Put differently, if the inequality $P_{ic} > S_{ic}$ holds across the majority of trials for any given participant, then the participant would be defined as having committed a CF overall. The greater the inequality, the greater the strength of the CF.

Another important distinction to determining whether a CF occurred is whether error rates for CF-eliciting trials are greater than error rates for non CF-eliciting trials. Non CF-eliciting trials were intended to not produce error rates indicative of a CF and are useful in evaluating the design assumptions of the proposed task. Validating this prediction also serves to support the argument that error rates on CF-eliciting trials are unique to these trials and are not produced by randomly pairing single grids together.

To recap, in CF-eliciting trials, the logic of obtaining CFs is that the easy and hard grids, in each pair, have a different majority colour. It seems likely, therefore, that participants will be misled by the easy trial and respond erroneously. By contrast, in non-CF-eliciting trials, both the easy and the hard grids have the same majority colour and in these cases we expect as low an error rate, as for the easy grid individually. Note, despite our characterizations of paired trials as CF-eliciting and non-CF-eliciting, the only consideration which matters is that in some cases the paired grid trials have a higher error rate than the corresponding single grid trials (since when this occurs, by the above logic, we have a CF).

Transposing violations of the conjunction rule into the psychophysical domain has the added advantage of allowing the application of response time (RT) models. Although attempts can be made to interpret performance in these tasks via a speed-accuracy trade-off, such an approach does not hold when the relationship between accuracies and RTs is non-symmetrical (Wagenmakers, Van Der Maas, & Grasman, 2007). In this instance, speed and accuracy are not traded but combined in some meaningful manner. One way to uncover this relationship is by investigating the unobserved variables underlying task performance through cognitive models. One such class of models is evidence accumulation models (EAMs). EAMs are a robust class of two-choice RT models (Brown & Heathcote, 2008). In their most basic form (EZ-Diffusion model) they have three variables assumed to drive performance on RT tasks. Firstly, the rate at which evidence for one of the two-choice responses is accumulated: drift rate. Here evidence simply means all the relevant information for one of the two responses. Secondly, the amount of response conservativeness for each of the two-choice responses: response threshold boundary. Thirdly, the amount of time spent on motor responses and pre-stimulus processing: non-decision time. Collectively, these variables model accuracies according to the first response threshold to accumulate the necessary amount of evidence to trigger a response (Wagenmakers, van der Maas, Dolan, & Grasman, 2008). RTs are modelled as a sum of the non-decision and decision components of processing (Wagenmakers et al., 2008). Analysing these differences amongst participants allows for a more distinct quantification of performance differences. As such, these models provide a well-developed method of exploring the information processing system underlying the occurrence of the CF.

Violations of the conjunction rule have largely been determined through a restricted methodological approach, which fails to address ways of minimizing framing effects linked to the CF and questions the domain generality of existing findings. One possible way of dealing with these issues is by applying a psychophysical framework to determining violations of the conjunction rule and therefore the presence of a CF. Applying a psychophysical framework would additionally allow the utilization of perceptual decision-making models that permit a more extensive exploration of the information processing systems associated with the CF.

To briefly summaries, the aim of this section is to examine whether under certain experimental conditions participants display a consistent violation of the conjunction rule, given its representation as psychophysical stimuli. There are therefore two experimental hypotheses:

H1: Proportion of incorrect responses for CF-eliciting trials should be higher than for non-CF-eliciting trials. This is equivalent to comparing the probability that Linda is a feminist and bank teller vs. Linda is a feminist and not a bank teller.

H2: Proportion of incorrect responses for CF-eliciting trials should be higher than for corresponding hard single grid trials, $P_{ic} > S_{ic}$.

2.1.1 Method

Participants

I recruited 12 participants through City, University of London's internal participant recruitment platform. All participants had normal or corrected to normal vision. Participants were all compensated £10 for their participation. As this is a psychophysical experiment, the standard procedure of recruiting a small number of participants to provide a large amount of data per participant for later modelling analyses was applied here.

Stimuli and stimulus sets

The stimuli consisted of 20x20 grids of 0.3cm x 0.3cm squares superimposed on a light grey background on each trial. Each grid was made up of a proportion of blue-to-orange coloured squares. Throughout the experiment there were four separate blue-to-orange colour proportions: either a grid had 65/35 percent more blue than orange squares in a grid, 60/40 percent more blue than orange squares in a grid, 54/46 percent more blue than orange squares in a grid or 52/48 percent more blue than orange squares in a grid. As a counterbalancing manipulation, a grid could have a greater percentage of orange-to-blue squares in it, instead of blue-to-orange. All stimulus grids were dynamic, with the location of each coloured square in the grid randomly changing at a rate of 15 frames per second. This was done simply so as to prevent the task from being trivial.

The grid colour proportions of 65/35 and 60/40 were defined as the easy colour proportions, because it was easier to determine what the dominant colour was in a grid (high visual discriminability). The grid colour proportions of 54/46 and 52/48 were defined as the hard colour proportions, because it was harder to determine what the dominant colour was in a grid (low visual discriminability).

The stimuli were divided into two main sets: either a single grid was presented on a trial (single grid trials), or two grids were simultaneously presented on a trial (paired grid trials). On trials where two grids were simultaneously presented, the grids were both shown in the centre of the screen, parallel to each other at either 5cm (close separation) or 14cm (moderate separation) from one another. Similar to the original Linda task where different feature descriptions of Linda could be introduced to assess their effects on CF rates, this manipulation was implemented as a way of introducing additional features in the task, which could later be controlled to assess their effect of CF rates.

For single grid trial sets, 12 trials were presented with the stimulus grid having colour proportions corresponding to each of the four grid colour proportions. This resulted in a total of 12*4=48 trials. Of these 48 trials, half the trials would have blue as the dominant colour and the other half orange.

For paired grid trial sets, participants were presented with two stimulus grids next to one another. Paired grid trials, for which blue was the dominant colour, were constructed so that one grid was hard and the other easy (as above, 60/40 vs 54/46 and 65/35 vs 52/48) and according to whether:

> the trials were CF-eliciting, in which case, the hard grid always had fewer squares of the target colour in a trial (e.g. if the trial question is "do both grids have more blue than orange in them", the hard grid always had more orange and the easy grid had more blue) vs. non-CF-eliciting, in which case in both grids the dominant colour was the same; grids were closely separated vs. moderately separated; placed in one of the two possible configurations on the screen (left-right or right-left). Refer to Figure 2.1 below for an illustration of the stimulus pairs.

We describe the CF-eliciting trials first. There were 96 trials, 48 for close separation and 48 for moderate separation. The 48 close separation trials consisted of two sets of 24 trials where the hard grid and the easy one, in each trial, were presented in a certain left-right configuration and another 24 trials for which this configuration was reversed. The non-CF-eliciting trials were constructed analogously, for another 96 trials. Overall, so far we have described 48 single grid trials and 96+96 paired grid trials, for a total of 240 trials. Another 240 trials were constructed so that the target colour in the question was switched, so that the experiment consisted of, in total, 480.

Design and procedure

The experiment was based on a within-participants design with four variables: the number of grids presented on a trial (V1), the distance between grids on paired grid trials (V2), the proportion of blue-to-orange squares in each grid (V3), the target colour of the trial questions (V4) and whether the grid with the easy colour proportion was shown on the left or right hand side on paired grid trials (V5). V1 had two levels: 1 or 2 grids presented on screen (single vs paired grid trials). V2 had two levels: 5cms or 14cms of distance between grids on paired grid trials. V3 had four levels: hard colour proportions (52/48 and 54/46) and easy colour proportions (65/35 and 60/40). V4 had two levels: "Are there more blue than orange patches in the grid(s)?" and "Are there more orange than blue patches in the grid(s)?" V5 had two levels: left-hand side or right-hand side.

The experiment was presented in four successive blocks (B1, B2, B3 and B4). Blocks B1 and B2 contained a set of single grid trials and paired grid trials respectively. Blocks B3 and B4 were identical to blocks B1 and B2, except that the target colour in the trial questions for these blocks was different (e.g. is there more blue vs is there more orange). At the start of each block with single grids trials, participants were presented with the question they had to agree or disagree with on all subsequent trials for that block. The question was either: “Are there more orange than blue patches in the grid?”, or “Are there more blue than orange patches in the grid?” At the start of each block with paired grids trials, participants were also presented with the question they had to answer on all subsequent trials. The question was either: “Are there more orange than blue patches in both grids?”, or “Are there more blue than orange patches in both grids?” The presentation order of the questions was randomized, such that if block 1 asked participants to determine the proportion of blue-to-orange patches, block 2 asked a similar question for paired grids. Blocks 3 and 4 subsequently asked participants to determine the proportion of orange patches in the grid(s).

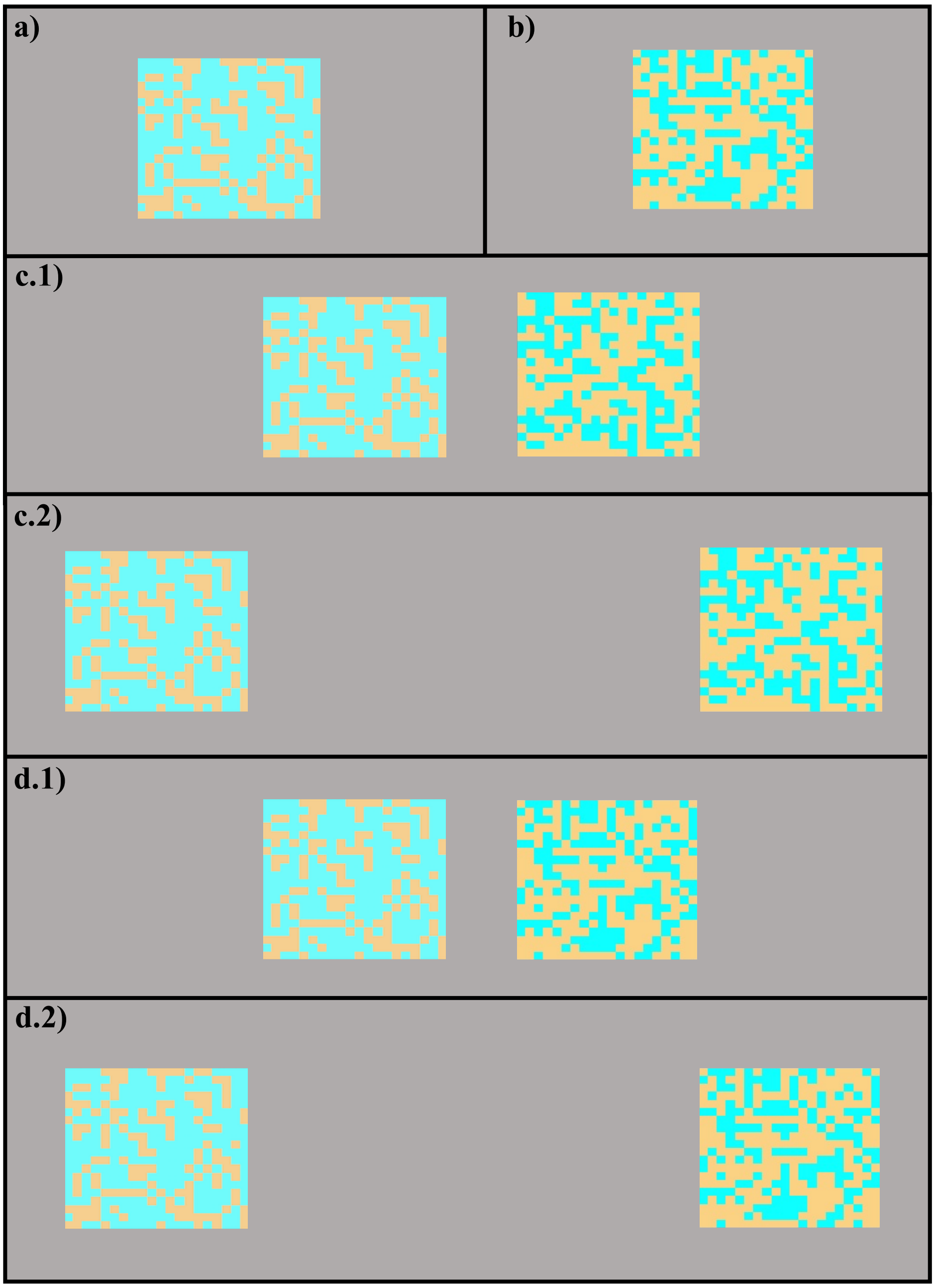

Figure 1.1. A static screenshot of the dynamic stimulus for the different conditions in the experiment. This figure shows conditions for trials where Blue is the target colour. A) A single easy grid trial with a colour proportion of 65/35 more blue than orange squares. b) A single hard grid trial with a colour proportion of 52/48 more blue than orange squares. c.1) A CF-eliciting paired grid trial with close separation, with the left grid having a colour proportion of 65/35 more blue than orange squares and the right grid having a colour proportion of 52/48 more orange than blue squares. c.2) An Cf-eliciting paired grid trial with moderate separation, with the left grid having a colour proportion of 65/35 more blue than orange squares and the right grid having a colour proportion of 52/48 more orange than blue squares. d.1) A non CF-eliciting grid trial with close separation, with the left grid having a colour proportion of 65/35 more blue than orange squares and the right grid having a colour proportion of 52/48 more blue than orange squares. d.2) A non CF-eliciting paired grid trial with moderate separation, with the left grid having a colour proportion of 65/35 more blue than orange squares and the right grid having a colour proportion of 52/48 more blue than orange squares.

2.1.2 Results

The mean RT for all participants across every condition in the experiment never exceeded 1.6s. Only on 0.4% of all trials were there RT outliers of 5s or more. As such, RTs over 5s were removed from the analysis. The proportion of incorrect responses for experimental and control trials across all participants for different trial types is shown in Table 2.1.

Table 2.1. Proportion of incorrect responses across all participants.

	Easy Single	Hard Single	CF-Eliciting	Non CF-Eliciting
Percent Incorrect	3%	23%	66%	16%

As previously mentioned, an examination of whether participants are committing a CF or not can be assessed by comparing accuracies on CF-eliciting trials, to accuracies on hard single grid trials. Between paired grid trials and single grid trials we see a clear difference in averaged responses. Participants on average largely performed well on single grid trials that represented conjunct probabilities. However, on trials representing conjunction probabilities, participants on average performed in the opposite direction and responded with a higher number or errors. Most importantly, the equality $P_{ic} > S_{ic}$, indicative of a CF as previously defined, appears to hold. To check this behaviour, we performed a series of inferential analyses described next.

As the data was found to not be normally distributed, non-parametric analyses were conducted. A 2(condition: CF-eliciting trials vs non CF-eliciting trials) x 2(targe colour: blue vs orange) x 2(distance: far vs close) x 2(side: easy proportion grid is on the left vs right) Friedman test was performed on participants' proportion of correct responses during paired grid trials. This was done to assess H1. A significant main effect of condition was found, with the proportion of correct responses on CF-eliciting trials (*Mdn* = .33) being significantly lower than in non CF-eliciting trials (*Mdn* = 0.92), $X^2(1) = 51.654$, $p < .001$. A non-significant main effect of distance was found. A non-significant main effect of side was also found. Additionally, a non-significant man effect of target colour was found. These results show that the quantity of errors associated with CF-eliciting trials are differentiable to non CF-eliciting trials in a manner consistent with H1. Consequently, these initial findings provide evidence in support of H1.

I examined the difference in the proportion of incorrect responses between CF-eliciting trials and hard single grid trials, to determine if error rates for CF-eliciting trials were significantly different, consistent with H2. To do this, a non-parametric Wilcoxon Signed Rank Test was performed on error rates for CF-eliciting trials and hard single grid trials. The results showed that participants in the CF-eliciting trials had a significantly higher proportion of incorrect responses (*Mdn* = 64.84) compared to hard single grid trials (*Mdn* = 27.08), $W_s = 78$, $p < .05$. These results uphold the inequality $P_{ic} > S_{ic}$ and are consistent with the definition of a CF being committed, as previously outlined. Recall, this is H2, according to which error rates on CF-eliciting trials should be higher than error rats on hard single grid trials, in order for a CF to have been committed.

EZ-Diffusion Modelling

The present framework permits large scale iterations of the CF problem. In turn, this allows for the collection of more exhaustive RT data. As a result, this facilitates the use of RT modelling procedures that may provide insight into the processing systems underlying the CF; one such class of models is EAMs. Fitting an EAM to the present data provides estimates for parameters associated with task-related and participant-related features such as drift rates, boundary threshold and non-decision times. These parameters are more general features of the information processing system: the rate of information processing, response conservativeness and stimulus encoding time, respectively. In the present study, participants' drift rate, response threshold (boundary separation) and non-decision time were derived via EZ-diffusion model estimates (Wagenmakers et al., 2007). As the EZ-diffusion model is an analytical model that simplifies the more commonly used EAMs, it does not use a modelling fitting procedure. Instead, the three model parameters are estimated based on three data features: the proportion of correct responses, the mean RT for correct responses and the variance of RTs for correct responses. Drift rates are derived from the proportion of correct responses and variance in the RTs of correct responses. Boundary separation values are then derived from the proportion of correct responses and drift rates. Nondecision times are finally derived from the proportion of correct responses, drift rates and boundary separation values. Refer to Figure 2.2 below for a schematic illustration of the EZ-Diffusion model inputs and outputs. The estimated model parameters representing the rate of information processing, response conservativeness and non-decision time were then analysed to access their contribution to committing the CF.

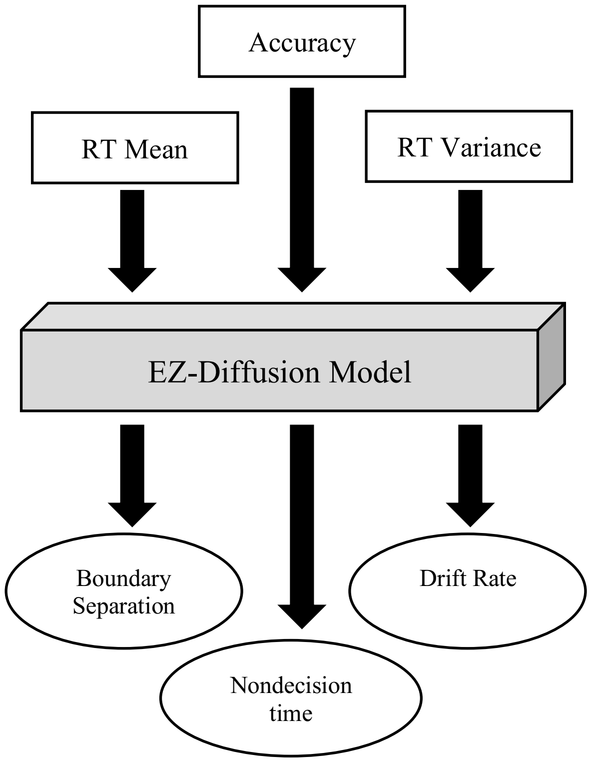

Figure 2.2. A schematic representation of the EZ-Diffusion model inputs and the three model parameters estimated as a result of the model computations.

Please refer to Table 2.2 below for the estimated drift rates for the four main different trial types (easy single grid, hard single grid, CF-eliciting and non CF-eliciting trials) across all the participants.

Table 2.2. Estimated drift rates for the four main different trial types across participants.

CF-eliciting Trials	Non CF-eliciting Trials	Hard Single Trials	Easy Single Trials
0.22	0.22	0.27	0.27
0.18	0.18	0.27	0.38
0.18	0.18	0.18	0.19
0.23	0.18	0.21	0.31
0.22	0.27	0.20	0.39
0.20	0.26	0.23	0.28
0.21	0.19	0.20	0.31
0.19	0.21	0.23	0.24
0.20	0.20	0.34	0.25
0.22	0.18	0.26	0.27
0.21	0.22	0.37	0.38
0.21	0.20	0.26	0.27

A one-way Greenhouse-Greisser corrected RM ANOVA on one factor with four levels (easy single grid, hard single grid, CF-eliciting and non CF-eliciting trials) was conducted on participants' drift rates. A significant main effect of trial type was found: $F(2.389, 26.281) = 12.246$, $p = <.001$, $\eta_p^2 = 0.53$. To further assess if there was a significant difference in drift rates between trial types, I conducted post-hoc Bonferroni–Holm *t*-tests comparing the different trial types, please refer to Table 2.3 for the results of the post-hoc analyses.

Table 2.3. Post-hoc analyses on the significant main effect of trial type on drift rates.

		Mean Difference	SE	t	p
CF-eliciting	Non CF-eliciting	-0.001	0.017	-0.068	1.000
	Hard single	-0.046	0.017	-2.684	0.068
	Easy single	-0.088	0.017	-5.189	< .001
Non CF-eliciting	Hard single	-0.045	0.017	-2.617	0.080
	Easy single	-0.087	0.017	-5.121	< .001
Hard single	Easy single	-0.043	0.017	-2.505	0.104

Overall, these results suggest that the rate of information processing varied significantly between trial types. Most significantly, between CF-eliciting and non CF-eliciting trials the rate of information processing was not significantly different, as were drift rates between single grid trials. However, the rate of information processing was significantly slower during paired grid trials (conjunction probabilities) compared to single grid trials (conjunct probabilities).

Given that the stimuli for single grid trials are the same as for paired grid trials, the difference in processing rates between single and paired grid trials highlight possible processing peculiarities in the case of conjunctive probabilities.

Please refer to Table.2.4 below for the estimated boundary separation values for the four main different trial types (easy single grid, hard single grid, CF-eliciting and non CF-eliciting trials) across all participants.

Table 2.4. Estimated boundary separation values for the four main different trial types across participants.

CF-eliciting Trials	Non CF-eliciting Trials	Hard Single Trials	Easy Single Trials
0.16	0.18	0.16	0.20
0.169	0.17	0.16	0.24
0.165	0.16	0.15	0.24
0.14	0.13	0.20	0.23
0.13	0.16	0.20	0.15
0.18	0.16	0.20	0.15
0.13	0.14	0.10	0.22
0.12	0.12	0.12	0.19
0.19	0.19	0.10	0.20
0.16	0.17	0.22	0.24
0.16	0.17	0.14	0.19
0.17	0.17	0.22	0.21

A one-way Greenhouse-Greisser corrected RM ANOVA on one factor with four levels (easy single grid, hard single grid, CF-eliciting and non CF-eliciting trials) was conducted on participants' response boundary separation values. A significant main effect of trial type was found: $F(1.886, 20.747) = 7.420$, $p = .004$, $\eta_p^2 = 0.40$. To further assess if there was a significant difference in boundary separation values between trial types, I conducted post-hoc Bonferroni–Holm *t*-tests comparing the different trial types, please refer to Table 2.5 for the results of the post-hoc analyses.

Table 2.5. Post-hoc analyses on the significant main effect of trial type on boundary separation values.

		Mean Difference	SE	t	p
CF-eliciting	Non CF-eliciting	-0.005	0.012	-0.413	1.000
	Hard single	-0.008	0.012	-0.686	1.000
	Easy single	-0.049	0.012	-4.177	0.001
Non CF-eliciting	Hard single	-0.003	0.012	-0.273	1.000
	Easy single	-0.045	0.012	-3.764	0.004
Hard single	Easy single	-0.041	0.012	-3.491	0.008

Overall, paired grid trials elicited lower mean response thresholds compared to single grid trials. However, this difference was only significant when comparing paired grid trials to the easy single grid trial with a high (easy) level of visual discriminability, not with hard single grid trials with a low (hard) level of visual discriminability. One would expect response conservativeness (i.e. requiring more information than usual before making a judgement) to be higher for the more complex paired grid trials (conjunction probabilities). However, the opposite was found here. These results suggest that participants may have erroneous or biased interpretations for what it means to combine probabilities in a conjunction, that result in lower levels of evidence being required in conjunction judgement before a decision is triggered.

A one-way RM ANOVA with four factors (easy single gird, hard single grid, experimental and control paired grid trials) was conducted on participants' non-decision time values. Results found no significant main effect.

2.1.3 Discussion

Results show that, relative to trials representing conjuct probabilities (single grids), trials representing conjunction probabilities (paired grids), were characterised by significantly higher error rates. Examining the rate at which participants were processing the different stimuli (drift rates) and response conservativeness (boundary separation values) through RT modelling, reveals that a possible bias in the information processing system may be related to the observed CF in this task.

Response boundary separation values are defined as a self-determined measure of response conservativeness (Brown & Heathcote, 2008). In this sense, a higher threshold indicates a participant's belief that more information needs to be gathered from the stimuli to trigger a certain response. Modelling results show that easy single grid trials had significantly lower response thresholds to paired grid trials. From an information processing perspective, this makes sense. As easy single grid trials had high visual discriminability, they were of significantly lower difficulty and thereby supposedly require less stimulus processing time compared to the more difficult hard single grids. However, when this logic is applied to the difference in response thresholds between hard single grid trials and paired grid trials, the results are surprising. It can be assumed that if a higher response threshold is used by participants on paired grid trials compared to easy single grid trials, then an even higher response threshold ought to be found for paired grid trials compared to hard single grid trials. This is because the increased perceptual difficultly associated with hard single grid trials that have a close blue-to-orange colour proportion should result in participants requiring more information before deciding on a response. However, surprisingly, neither of these results were found. Firstly, response thresholds were lower on paired grid trials compared to easy single grid trials. This implies that participants required less information to arrive at a decision on more complex (paired grid) trials compared to simpler (single grid) trials. Secondly, participants had response thresholds on paired grid trials that were not significantly different compared to hard single grid trials, despite these trials involving a trial question involving two grids instead of one and therefore presumably, a higher level of task difficultly that would require more information to be collected before a decision can be made, compared to single grid trials.

I argue that these contradictory results suggest that the information processing system associated with the present CF task appears to be erroneously combining information from single grid trials into paired grid trials or is experiencing some difficulty in doing so. There is some indication that these results do not arise from participants misunderstanding the task. Firstly, a similar pattern of results is found in the drift rate results in further experiments discussed below. Secondly, it is important to note that paired grid trials are comprised of single grid trials. As Table 2.1 above shows, error rates on these trials were low on average. As such, if error rates on paired grid trials are substantially higher, as was found, this demonstrates difficulty in combining single grid trials into paired grid trials. Furthermore, as single grid trials are argued to represent conjunct probabilities and paired grid trials conjunction probabilities, this contradictory pattern of results concerning response thresholds is viewed as capturing the erroneous combining of conjunct probabilities into conjunction probabilities, in a way broadly analogous to the process which ostensibly causes the CF in the original Linda task.

Drift rates are defined as the rate of information processing. To recap, a drift rate is primarily influenced by the quality of the stimulus and task difficulty, whereby higher drift rates indicate lower task difficultly and lower drift rates indicate higher task difficulty (Brown & Heathcote, 2008; Wagenmakers et al., 2007). Results show that drift rates were significantly lower for paired grid trials, compared to single grid trials. As paired grid trials are produced by combining grids presented on single grid trials, one would not expect the rate at which either single grid is processed to change simply because they are presented as a pair. This unexpected finding provides some insight into the possible information process system used by participants in this task. For example, if we assume that participants process the stimuli in parallel, then maybe there should be a difference in drift rates between single and paired grid trials. This is because participants may be processing both grids on paired grid trials simultaneously and subsequently processing two grids as one. As a result, information processing rates (drift rates) may fluctuate to represent this change in the way the stimuli are being processed. However, if participants process the stimuli serially, then each grid on paired grid trials should be treated individually and therefore drift rates on these paired grid trials should be identical to those for single grid trials.

In either case, a serial or parallel processing structure does not provide an immediate answer to some of the other findings. For example, it is not clear why drift rates are only significantly different when comparing paired grid trials and easy single grid trials and not hard single grid trials. I instead argue that the somewhat contradictory pattern of results found in the modelling results are indicative of some yet undiscovered phenomena in the information processing system. This may reflect a kind of error or bias that occurs when combining single grid trials (conjunct probabilities) into paired grid trials (conjunction probabilities). That is, this may be an error or bias related to information processing, by unusual and unexpected changes in the drift rates and response thresholds between paired and single grid trials. This resulting error or bias that occurs during the information processing of the stimuli may be responsible for increased error rates on paired grid trials that produce the observed CF in this task.

Since Tversky and Kahneman's (1983) findings on the CF, the literature has been heavily concerned with exploring the robustness of this phenomenon (Hertwig & Gigerenzer, 1999;

Tentori et al., 2004). However, the general framework used for accessing a CF has used one method of representing conjunct and conjunction probabilities: descriptive scenarios. Exploring the CF through one methodological approach restricts the scope through which the CF itself can be exhaustively analyzed. Conventional experimental methods utilize high level scenario-based tasks to determine a CF; such high level tasks are themselves prone to strong framing effects (Stanovich & West, 1998) and other possible biases. The result of such effects is to inadvertently facilitate misinterpretation and potentially cause an erroneous perception of a causal link between scenario features and thereby lead to overweighting of the conjoint probability. As such, what may be occurring is that the framework used to determine a CF may in itself be eliciting a CF, by facilitating erroneous information processing. More recent work has focused on attempting to explore the information processing system underlying the CF. What such findings have revealed is that people may be processing the features within a CF-eliciting scenario, under the erroneous assumption that some causal link exists between independent features within it (Ahn & Bailenson, 1996; Maguire et al., 2018). These findings thereby allude to a more general problem concerned with discovering a method that allows a critical exploration of the CF phenomenon, without facilitating the phenomenon itself.

In this experiment, I have tried to provide a method of tackling this issue by transposing the traditional framework used to determine the CF to the psychophysical domain. Furthermore, I have attempted to show that by doing so it allows for the use of cognitive modelling procedures which can provide more insight into the underlying processing architecture of the CF, compared to the conventional framework.

The results of this experiment show that a psychophysical framework can be set up to assess violations of the conjunction rule. Accuracy on probability judgements for conjunct probabilities (single grid trials) was very high. However, combining conjunct probabilities into conjunctions (paired grid trials), yielded significantly lower accuracy levels. The most pertinent and unexpected finding was that error rates on CF-eliciting trials were significantly higher than on hard single grid trials, despite CF-eliciting trials being comprised of easy and hard single grids. Recall, this is H2 and is the definition of a CF in the present task: $P_{ic} > S_{ic}$. These results thereby shed more light on the domains over which the CF is observable. For instance, the present psychophysical framework reduced the representation of conjunct and conjunction probabilities to a low-level cognitive task, by removing high order framing effects associated with descriptive representations. This in effect functioned to suppress the assumption that participants may have been adopting some form of causal link between the features in the descriptive representation of the conjunct and conjunction probabilities. Although some literature supports the idea that such assumptions are the main determinates of the CF, present findings do not. Instead, our findings reveal that in the near absence of any conceptual causal link between features in the conjunct and conjunction probabilities (by representing probabilities as abstract psychophysical stimuli) a CF is still observable.

A possible bias driven by an error in information processing becomes clearer when considering the symmetrical relationship between drift rates and response thresholds. Our results show that paired grid trials were categorized by the lowest drift rate and response threshold. However, it would be expected that if drift rates are low (high task difficulty), then response thresholds would in turn increase to accommodate the increased uncertainty in determining the correct response. Contrary to this, our results show that paired grid trials co-occur with the lowest response thresholds (decreased uncertainty in determining the response answer), compared to single grid trials. I propose that this unexpected and contradictory pattern of results is due to some yet undefined bias, driven by an error in the information processing system involved in combining conjunct probabilities into conjunction probabilities. I also propose that this bias is not due to a tendency to assume subjective uncertainty in the conjunct and conjunction probabilities, as probabilities were given an abstract representation through our psychophysical framework. This procedure thereby functioned to minimize any causal links between probabilities, which would otherwise be viable in more feature descriptive representations of probabilities. As such, I believe that some error in the information processing involved in combining conjunct probabilities into conjunction probabilities occurs, which then triggers a bias facilitating the CF.

Another unexpected finding is that response thresholds were found to be significantly different between easy single and hard single trials, even though participants had no prior experience with stimulus orders or the type of stimuli. Overall, the results indicate a possible erroneous response strategy or bias of some sort. It is possible that variability in response strategies led to variability in modelling results (e.g., variability in response thresholds), though equally I cannot preclude the possibility that the model simply fits poorly the current empirical data.

The EZ-diffusion model that I chose is the simplest class of EAMs. This is because model parameters are not “fitted” to the data in the classical modelling sense. Instead, model parameters are estimated based on three characteristics of the dataset: accuracy, mean RT and RT variance. This allows for a very general assessment of how certain features of the cognitive system may be operating (drift rate, response threshold and non-decision-time). It also allows for a general assessment of these features that bypasses the, at times, extremely lengthy modelling procedures associated with fitting EAMs to data. However, this approach is not without limitations, associated with not fitting a model to the data, as standard (Wagenmakers, Van Der Maas, & Grasman, 2007). Nonetheless, I argue that given that the modelling presented here largely functioned as an exploratory analysis of the underlying information processing system for the present CF task, it did not warrant a more laborious modelling procedure. Furthermore, I was only concerned with two processing features at this stage: the rate of information processing (drift rate) and response conservativeness (response threshold). The EZ-diffusion model was primarily designed to provide an estimate for these features, but was not supposed to offer conclusive modelling results for how these features function in such a task. For that, more elaborate modelling procedures are required.

I believe that a possible bias in information processing is not unique to our stimuli. Rather, it is reflective of a more fundamental bias in the processing involved in combining conjunct probabilities into conjunction probabilities generally. This is because the abstract psychophysical representation of conjunct and conjunction probabilities in our study is a simpler and more accurate representation of probabilities themselves. Conjunct and conjunction probabilities are visually represented as proportions occupying some finite space in our experiment: the number of blue-to-orange squares in a grid. The number of blue-to-orange squares presented in each grid is a direct visual representation of the corresponding probabilities. As such, estimating if two simultaneously presented grids have more or less blue than orange squares is a probability judgement directly related to the proportion of blue-to-orange squares. This proportion is visually displayed in our framework, unlike in descriptive scenario-based tasks.

The present results show that CF-eliciting trials are, overall, characterised by a slower rate of information processing and, unexpectedly, a lower response threshold. Specifically, there are peculiarities in the rate of information processing between CF-eliciting trials and hard single grid trials that imply, in part, that information processing may be occurring in a serial manner. Furthermore, response thresholds between CF-eliciting trials and single grid trials further emphasise an error or bias in the information processing system associated with the present task. However, whether this is primarily the result of an error or bias in the processing of conjunct probabilities into conjunctive probabilities or some other aspect of information processing, such as the general processing order, is unclear. As the present experiment only provided an exploratory analysis into the underlying information processing system. To further explore this in more detail, the next step is to determine the exact information processing order underlying the present CF task. This can be done by applying systems factorial technology (SFT) to the present psychophysical CF framework. This will allow for a more elaborate analysis of the serial and parallel information processing structures that underlie the processing of CF-eliciting paired grid trials that result in an overall CF (Townsend, 1990). This will in turn shed more light on the exact information processing structure and order that underlie and cause the CF.

Section 2.2 The Conjunction Fallacy and Systems Factorial Technology

Introduction

My initial aim in this chapter was to determine whether the CF could be observed in the perceptual domain. I additionally aimed to provide a more accurate modelling account of CFs by transposing the CF into the perceptual domain and applying well developed perceptual decision-making models to understand the underlying architecture. One such class of models is EAMs. EAMs have provided accurate depictions of RT distributions and accuracy rates associated with making single stimulus decisions in the perceptual domain. These models have also allowed for an elaboration of the non-linear relationship in speed/accuracy trade-offs. They also provide a means through which to quantify the task related and participant related features of a task, which influence and possibly determine the decision-making process. As such, these models provide a viable way of exploring perceptual stimulus decision-making processes. However, the more informative and elaborate versions of these models apply to single stimuli tasks and have not yet been extended to dual stimuli tasks. One aim of this second experiment is to extend this approach from single to dual stimuli tasks. This will allow for more insights to be revealed about the underlying cognitive processing behind the CF phenomenon.

Results from the first experiment showed that overall participants did commit a CF during the perceptual CF decision-making task. However, there was no significant effect of target colour (orange vs blue), left vs right (easy proportion grid being on the left vs right), or distance between grids (close vs moderate separation).

A CF in the first experiment was interpreted as higher error rates when simultaneously judging the colour proportions in CF-eliciting trials versus judging the colour proportions in hard single grid trials. The specific pattern of results from Experiment 1 could be due to task-related features (drift rates) or participant-related features (response thresholds). Therefore, I used the EZ diffusion model to determine the properties of the task (drift rate) or individual (response threshold), that contributed to participants committing the CF. I found a significant main effect of condition (single vs paired grids) on drift rates. I also found a significant main effect of condition on response thresholds, however this difference was not significant between all trial types.

However, higher level cognitive architecture cannot be determined by assuming that evidence accumulation processes for dual stimuli are independent of one another. Both accumulators in the EAM must have a method of functionally interacting with processed information, be it serial, parallel or both. One method for detailing information processing architecture is SFT. Encompassing a series of analyses, SFT can diagnose and discriminate between five types of information processing architectures that possibly underlie a mental process. These include serial, parallel and joint serial and parallel processing architectures named coactive models. Applying a SFT analysis to data is dependent on the experimental paradigm having a factorial manipulation of stimuli detectability, named double factorial paradigm (DFP). This involves systematically manipulating the discriminability between pairs of stimuli on various trials such that it is increased, decreased or mutual.

Systems Factorial Technology

This experiment's objective is to identify the processing order of the information processing system behind the CF task that contribute to committing the CF. The method used to determine this is SFT. SFT is a suite or toolbox of methods and statistical tools which expand the ability to use RTs to discover important properties of the information processing system (Townsend & Nozawa, 1995). SFT makes use of the Double Factorial Paradigm (DFP) to isolate and capture individual processing mechanisms associated with a cognitive task.

Double Factorial Paradigm

SFT uses a factorial manipulation of target detectability in order to separate and isolate the information processing mechanisms underlying a cognitive task. The aim is to determine whether the processing order of the information processing architecture is serial or parallel. A typical DFP uses a "redundant-target" set-up to allow the factorial manipulation of the detectability (or saliency) of the stimuli. This type of setup is best shown through a simple detection task. In a brightness detection task, participants have to detect the presence of a stimulus that varies in brightness. The standard finding is that increasing brightness of a stimulus shortens the detection time. The detection time is thought to be composed of several subcomponents (identification time, decision time, and motor execution time).

Now consider a dot-detection task in which the aim is to detect the presence of one of the two dots which can be presented in one of the two spatially separated locations (location 1 or location 2). In this instance each dot can be presented at either high (H) or low (L) brightness. In this case, we assume that the cognitive system must process two channels: one for each dot in either location 1 or location 2. Specifically, stimuli can appear as a *high* detectability target in both locations (HH), a high detectability target in the first location but a low detectability target in the second location (HL), the opposite of this setup (LH), and as a low detectability target in both locations (LL). Refer to Figure 2.3 below. Such methods are based on the assumption of selective influence. This implies that there exists a strict relationship between the experimental manipulation and the effects of the manipulation on the processes of interests. In other words, an experimental manipulation affects only one single channel (sub-process) within a cognitive architecture. As in the dot-detection example, it is assumed that one channel (sub-process) exists for each of the two stimuli locations.

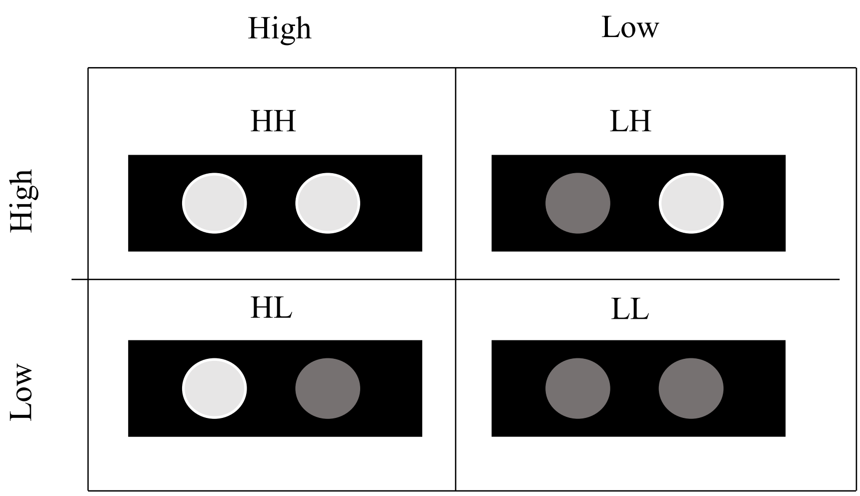

Figure 2.3. An illustration of the how two separate stimuli with different brightness levels in a dot detection task are represented by two separate processing channels in SFT.

When testing such cognitive mechanisms, one would expect the mean RTs for the high brightness dot in a given channel to be lower (faster) than the mean RTs for the low brightness dot in the same channel. This is because stimulus brightness is assumed to increase stimulus detectability and thereby decrease target identification time (RTs). One statistic that can be computed using the factorial manipulation is the survivor interaction contrast ($SIC(t)$). The survivor function $S(t)$, a statistical tool used in survival analysis (Elandt-Johnson & Johnson, 1999), is a function that indicates the probability that a particular process has terminated at time t.

First, we collect the vectors for the RTs from the LL, LH, HL, and HH experimental conditions for each individual participant. Averaging the data over participants is possible. However, there are statistical and philosophical issues associated with averaging across participants (Ashby, Maddox, & Lee, 1994; Estes, 1956). One disadvantage is that group averages can obscure important trends arising for an individual participant to such an extent that the average does not resemble any of the individuals. As a rule of thumb, responses for each experimental condition should each contain a relatively large number of trials (i.e., N>100). Next, from each of these vectors we calculate the normalized probability density function $f(t)$. Then, we obtain the empirical cumulative distribution function of the $f(t)$ values, $F(t)$. A simple transformation of $1 - F(t)$ yields the survivor function. $SIC(t)$ can therefore be defined as

$$SIC(t) = [SLL(t) - SLH(t)] - [SHL(t) - SHH(t)] \quad (1)$$

The survivor functions should be plotted on the same plot to ensure that they are ordered and that the assumption of selective influence holds (Houpt, Blaha, McIntire, Havig, & Townsend, 2014). A parallel exhaustive model predicts a SIC(t) that is entirely negative (revealing RT under-additivity). Refer to Figure 2.4 below for an illustration of the different SIC(t) curves produced by the different processing architectures in SFT. This exhaustive stopping rule is required in cases where all channels must reach completion before it is certain that a correct response can be made. The intuition for why a parallel exhaustive model predicts a negative SIC(t) is because the SLL(t) − SLH(t) term is always smaller than the SHL(t) − SHH(t) term across t. This is because, in a parallel exhaustive model, the RT for a redundant stimulus is the maximum time necessary to complete any of the target channels. Hence, the processing time for the LL, LH, and HL stimuli will be much slower than for the HH stimulus.

The SIC(t) functions for serial self-terminating and exhaustive processing take on very different shapes. When processing is serial and self- terminating, the SIC(t) is flat and equal to 0 at every point in time (Townsend & Nozawa, 1995). When processing is serial and exhaustive, the SIC(t) is an S-shaped curve with a negative region for early processing times and a positive region for later processing times. The negative and positive regions of the curve are equal to each other in the serial exhaustive model. Hence, the SIC function delivers strikingly distinct signatures for the important architectures and their stopping rules.

Coactive models form a class of serial and parallel models in which the information from each channel is pooled, typically by being added together into a single channel. The survivor interaction contrast function for the coactive model is negative at the beginning for the fast RTs and becomes positive later on and/or for slower RTs. This is similar in shape to the serial exhaustive SIC(t). However, the initial negative deflection is smaller than the later positive deflection in the coactive model. Assessing a system's capacity helps answer the question as to whether there is a significant cost, benefit, or no change in processing efficiency as a function of workload. Processing efficiency is essentially determined by comparing processing when multiple processing channels are operating, relative to an unlimited-capacity system.

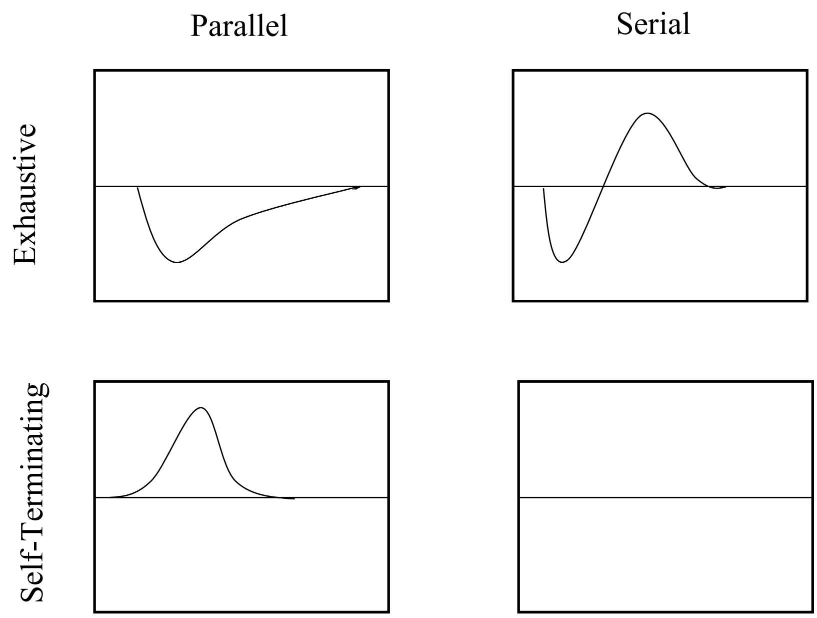

Figure 2.4. An illustration of the SIC(t) curves associated with the different processing architectures in SFT.

SFT and CF

The objective of this experiment is to identify the processing order of the information processing architecture associated with committing the CF in the psychophysical version of the

CF task presented in the previous section. SFT will be used to determine the processing order and stopping-rule that constitute the cognitive architecture associated with committing the CF. I will first demonstrate through simulation results that the inherent logic of SFT cannot produce a CF if applied to Experiment 1. This will show the inapplicability of the standard SFT framework to Experiment 1. I will then demonstrate that applying an effect of order to standard SFT logic can yield a CF. As such, the two hypotheses of Experiment 1 are also relevant here and a new third hypothesis is presented:

H1: Proportion of incorrect responses for CF-eliciting trials should be higher than for non-CF-eliciting trials. This is equivalent to comparing the probability that Linda is a feminist and bank teller vs. Linda is a feminist and not a bank teller.

H2: Proportion of incorrect responses for CF-eliciting trials should be higher than for corresponding single hard grid trials, $P_{ic} > S_{ic}$.

H3: Performance on the task will be driven by either a serial or parallel processing architecture, but not both.

2.2.1 Method

Participants

For Experiment 2 a total of 20 participants were recruited using the online recruitment platform Prolific. Participants were paid approximately £10 each for their participation. A similar logic to Experiment 1 is applied to participant recruitment in this experiment. Specifically, the standard procedure of recruiting a small number of participants to provide a large amount of data per participant for later modelling analyses is also applied here.

Design and procedure

The design of this second experiment was identical to Experiment 1, except a calibration stage was introduced at the start of the experiment.

In Experiment 1, the four colour proportions for easy and hard grids were selected after a series of pilot experiments. We nonetheless observed that error rates for designated hard trials were not within a clearly identifiable range for participants, indicating that participants found these trials challenging. As such, we introduced a calibration stage at the start of this second experiment. Participants were shown an indefinite number of single grid trials (i.e., the number of trials was not specified in advance) where the trial question remained the same: "*Are there more Blue than Orange patches in the grid?*" The first trial presented participants with a single square grid with a colour proportion of 66-to-34 more blue than orange squares. The majority colour in the grid randomly alternated between blue or orange on each trial.

If participants incorrectly judged the colour proportion, the majority colour proportion decreased by 2 on the subsequent trial (e.g. from 66/34 to 64/38). The calibration stage only terminated after participants produced an incorrect-correct-incorrect response sequence. At this point, the colour proportion on the last trial was set as one of the two colour proportions for the hard (visual discriminability) grids. The second colour proportion was determined by simply adding 2 to the majority colour in this colour proportion: e.g., if the hard colour proportion was 56/44 at the end of the calibration stage, the second hard colour proportion was 58/42. In addition to this, before the calibration stage ended, the colour proportion had to have been calibrated to a proportion where the majority colour was under 58% but above 52% This was done to prevent the termination of the response sequence and the calibration stage, when the discriminability of the colour proportion in the grid was still too easy or too hard for participants. We found through pilot testing in Experiment 1 that a colour proportion where the majority colour was above 60% produced very low error rates and therefore we set this as the upper limit for a hard grid colour proportion.

2.2.2 Results

SFT simulations. Preliminary data simulations were conducted to assess whether a standard SFT serial or parallel information processing architecture could produce error rates indicative of a CF, as in Experiment 1. Specifically, if I simulate performance for easy single and hard single grid trials in Experiment 1 and then combine performances using an SFT architecture to simulate paired grid trials, can I get an averaged error rate similar to Experiment 1 indicative of a CF? Also, I wanted to assess if such simulations can yield reasonably realistic RT data.

Data simulations were based on the simplest and most complete EAM choice-RT model: the linear ballistic accumulator (Brown & Heathcote, 2008). The LBA model represents a choice between N alternatives (N=2,3, ...) using N different evidence accumulators, one for each response. Each evidence accumulator begins the trial with a starting amount of evidence (k), drawn from a uniform distribution on the interval [0, A], that increases at a speed given by the "drift rate" (d). This drift rate is drawn from a normal distribution with standard deviation s. Accumulation continues until a response threshold (b) is reached. The first accumulator to reach the threshold decides the overt response, and the time taken to reach the threshold decides the RT (plus some extra constant time for non-decision processes, t_0). This gives five key parameters: k, d, b, $t0$ and s. Any EAM that possesses these five parameters is considered a complete EAM, in the sense that there exist other, more restricted, EAMs that possess fewer model parameters (Brown & Heathcote, 2008; Evans, 2019). For example, the EZ-diffusion model presented in Experiment 1 only possesses the d, b, and $t0$ parameters.

I first fitted the linear ballistic accumulator (LBA) model to RT and accuracy data for single easy (high visual discriminability) and single hard (low visual discriminability) grid trials from Experiment 1, using the quantile maximum probability estimator (QMPE) fitting procedure (Donkin, Averell, Brown, & Heathcote, 2009). QMPE creates a likelihood function in terms of the quantiles of the observed data and not the observed data themselves. Then parameters are searched for that maximize the likelihood by differentiating the log-likelihood of the parameters and applying a standard optimization method. The model was fitted to the averaged data to produce best fitting LBA parameters to simulate single patch responses for the hard grid colour proportions (52/48 and 54/46), and the easy grid colour proportions (65/35 and 60/40). As Experiment 1 had a non-significant effect of colour, all simulation results for paired grid trials refer to instances when either blue or orange is the target colour. This gave a total of two best fitting parameter sets: one for easy single grid trials and one for hard single grid trials. The fitting procedure was conducted on all single grid RT and accuracy data from all 12 participants in Experiment 1.

Simulations were initially performed to yield results for the easy and hard grid trials separately. The results were in line with those found in Experiment 1, refer to Table 2.6 below. We can see from Table 2.1 that the model fits for easy and hard single grid trials are similar to those from Experiment 1 (Table 2.6).

Table 2.6. Proportion of incorrect responses across all participants in Experiment 1 and for simulated easy and hard single patch trials. Each grid type was simulated for 1000 trials.

	Easy Single	Hard Single	CF-Eliciting	Non CF-Eliciting
Experiment 1 - Percent Incorrect	3%	23%	66%	16%
Simulations – Percent Incorrect	3%	29%	-	-

An algorithm for CF-eliciting trials was then written to simulate a RT and trial response (correct or incorrect) for a single trial, by combining simulated performances on easy single and hard single grid trials, using a serial and parallel SFT information processing architecture. The algorithm was therefore constructed by combining the separate simulation results for easy and hard single grid trials according to SFT serial or parallel processing logic. The simulated trial question being answered was: "*Are there more Blue than Orange patches in both grids?"* However, the target colour, blue in this case, is irrelevant since the simulating algorithm is insensitive to colours. Figure 2.5 below offers an illustration of the logic flow used to determine RTs and trial responses on simulated CF-eliciting trials, using a serial self-terminating architecture.

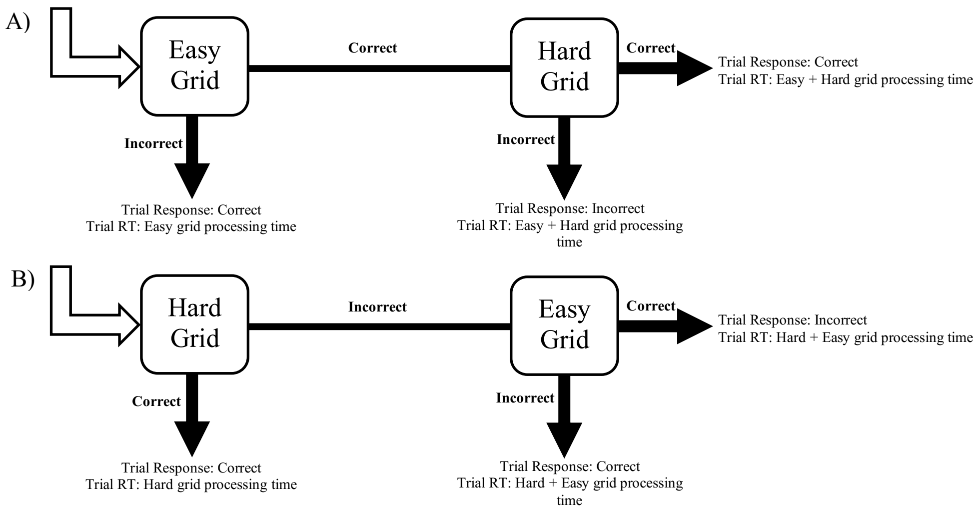

Figure 2.5. A) This figure illustrates the simulation logic behind the serial self-terminating processing order of Easy then Hard grid on a CF-eliciting trial. Statements at the end of arrows indicate end of trial responses and RTs. Correct/incorrect arrows and lines indicate the processing order and results depending on the simulated response provided for the preceding grid. B) This figure illustrates a similar simulation logic, but for the Hard then Easy processing order. A random processing order is defined as a random selection between the two processing orders.

The logic for a serial-exhaustive architecture is similar to that of the serial self-terminating one, except that RTs are always a sum of the total time spent processing both grids. This is because an exhaustive architecture assumes that both grids are processed regardless of whether or not a response is determined after processing the first grid.

Although a serial exhaustive architecture requires that both stimuli be processed in a serial order, in the context of the present experimental paradigm, it still holds that, even if a serial exhaustive architecture is being utilised, the first grid that is processed can still yield sufficient information on its own to determine a trial response. It may be the case that the first processed grid is determined to not have more squares with the target colour in it. Therefore, this yields enough information to respond to the trial without processing the second paired grid. This is because paired grid trials ask a question that must be true for both grids: e.g., *are there more Blue than Orange patches in both grids?* Regardless of whether the judgement is correct or incorrect, if the first grid to be processed in a serial order is viewed as violating the premise of the question, the trial answer can be determined without processing the paired second grid.

Figure 2.6 below offers an illustration of the parallel self-terminating logic flow used for paired grid trial simulations. The parallel exhaustive architecture is similar to the parallel self-terminating architecture, expect that RTs are equal to the total time spent on the last grid to finish being processed.

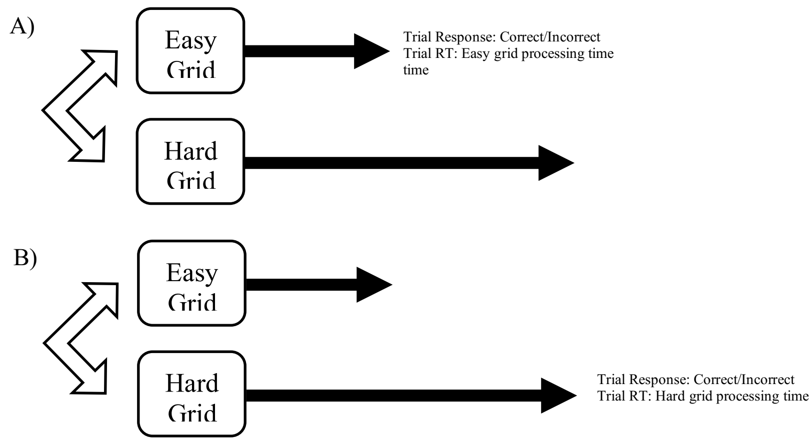

Figure 2.6. A) This figure illustrates the simulation logic behind the parallel self-terminating processing order of a CF-eliciting trial, where the processing of both grids begins at the same time, but the easy grid finishes being processed before the hard grid and triggers the response. The illustration here assumes that the easy single grid has a short processing time (denoted by a shorter black arrow) and the paired hard grid has a longer processing time (denoted by the longer black arrow). Statements at the end of arrows indicate end of trial responses and RTs. B) This figure illustrates a similar simulation logic but here the easy single grid does not trigger a response and depends on the outcome of the paired hard grid.

Simulation results for CF-eliciting trials based on the serial self-terminating processing architecture show that for the various processing orders in this architecture, the results do not yield a CF. More specifically, looking at the proportion of incorrect responses in the simulation results for each processing order reveals that the inequality $P_{ic} > S_{ic}$ does not hold, as shown in Figure 2.7 below. RT distributions are shown for all simulations to initially assess the plausibility of simulation results and will be used later to further evaluate the different simulation results.

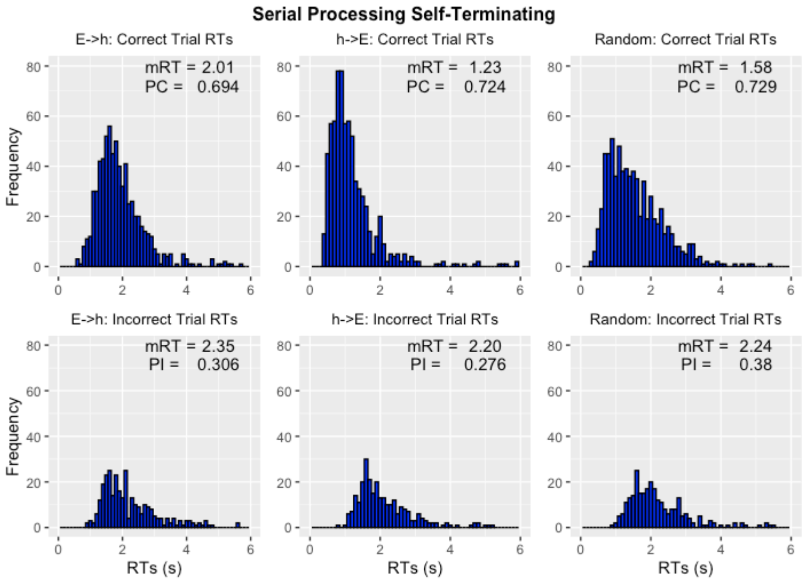

Figure 2.7. Each column contains the RT distribution for correct and incorrect trials, for the three processing orders of easy then hard grid (E->h), hard then easy grid (h->E) and random. Each processing order was simulated for 1000 CF-eliciting trials. The mean response time (mRT) is given for correct and incorrect trials. PC and PI refer to the overall proportion of correct and incorrect responses respectively for each processing order.

As such, a serial self-terminating architecture does not appear to replicate the experimental results which show a consistent CF. It is important to note that although the inequality does hold when compared to the hard single grid results of Experiment 1, it does not consistently hold when compared to its own simulated hard single grid results in Table 2.6. I next simulated the results for a serial exhaustive architecture, as illustrated in Figure 2.8 below.

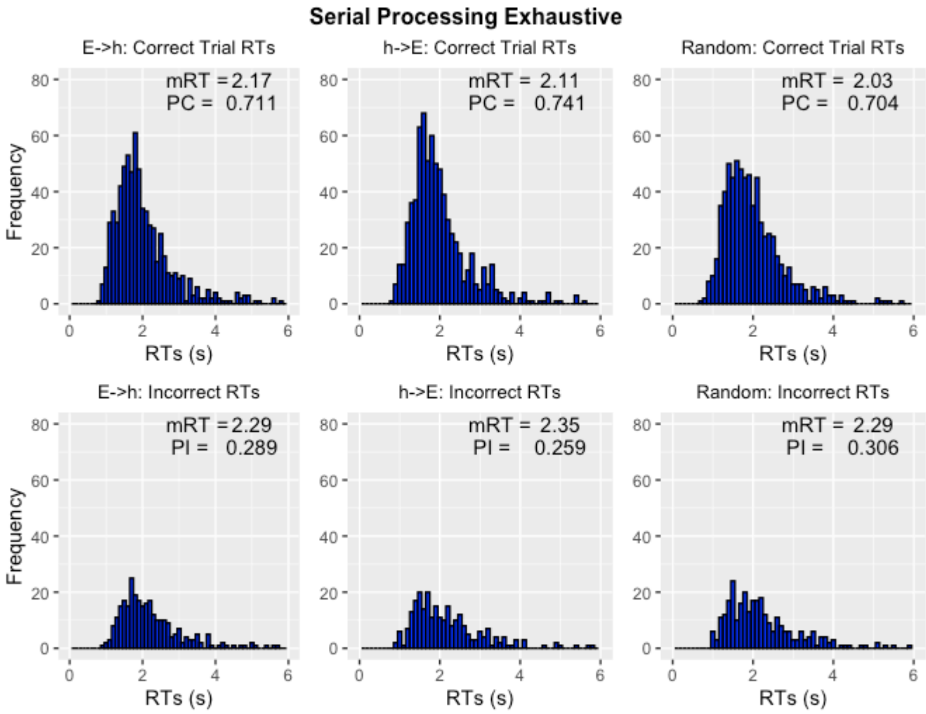

Figure 2.8. Each column contains the RT distribution for correct and incorrect trials, for the three processing orders of easy then hard grid (E->h), hard then easy grid (h->E) and random. Each processing order was simulated for 1000 CF-eliciting trials. The mean response time (mRT) is given for correct and incorrect trials. PC and PI refer to the overall proportion of correct and proportion of incorrect responses respectively for each processing order.

The results of the simulation show that the inverse of the inequality used to define a CF holds for the serial exhaustive architecture. As such, serial exhaustive simulation results do not appear to replicate the experimental results which show a CF. Similar to the serial self-terminating simulations, serial exhaustive simulations do not consistently show an overall CF for the three processing orders. Note, as expected due to the exhaustive nature of the serial exhaustive architecture, RTs are longer for exhaustive rather than self-terminating architectures.

Next the parallel self-terminating architecture was simulated, illustrated in Figure 2.9 below.

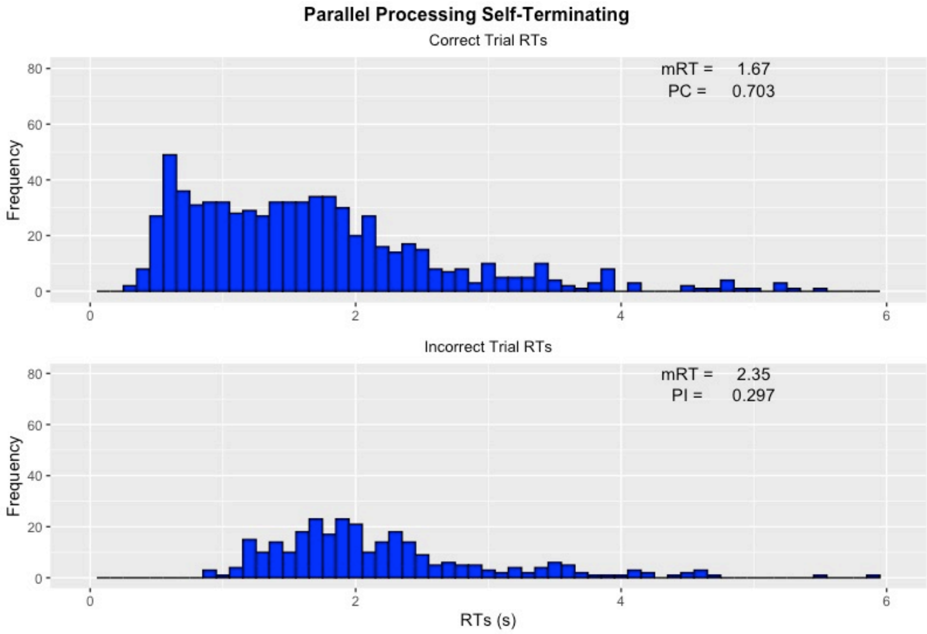

Figure 2.9. The first and second row contain the RT distribution for correct and incorrect response trials respectively. The trial was simulated for 1000 CF-eliciting trials. The mean response time (mRT) is given for correct and incorrect trials. PC and PI refer to the overall proportion of correct and incorrect responses respectively.

Simulation results for CF-eliciting trials using a parallel self-terminating architecture show that the inequality used to define a CF, again, does not hold. Parallel self-terminating simulations do not appear to show the occurrence of a CF. Paired grid trial simulation results for the parallel exhaustive architecture yields similar results, refer to Figure 2.10 below.

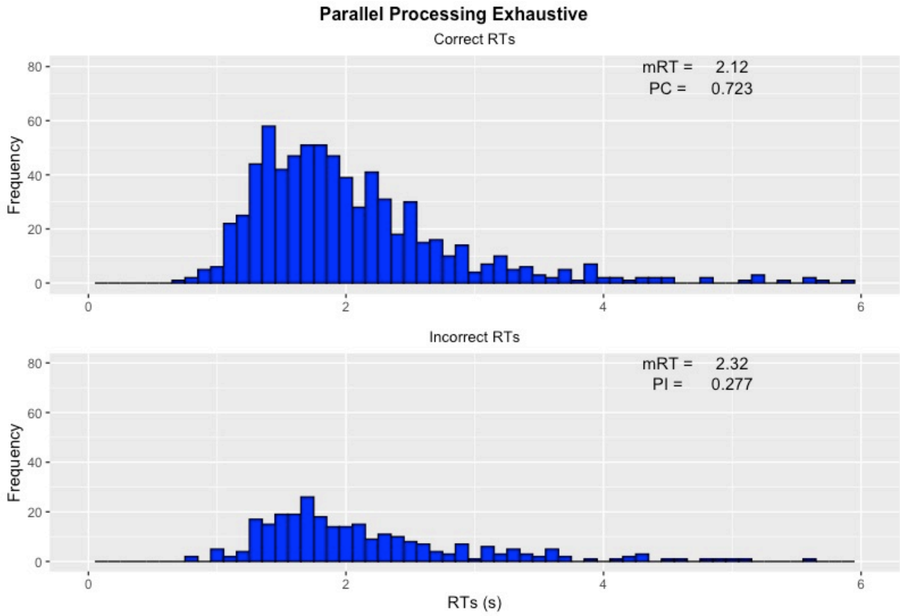

Figure 2.10. The first and second row contain the RT distribution for correct and incorrect response trials respectively. The trial was simulated for 1000 CF-eliciting trials. The mean response time (mRT) is given for correct and incorrect trials. PC and PI refer to the overall proportion of correct and incorrect responses respectively.

Results for the parallel exhaustive simulations also show that the inequality used to define a CF does not hold. Similar to the parallel self-terminating architecture, results do not yield a CF. Again, as expected due to the exhaustive nature of the parallel exhaustive architecture, RTs are longer for exhaustive rather than self-terminating architectures. Although some simulations displayed similarities in RTs to the experimental data, the most significant quantitative feature of the aforementioned simulations is their ability or not to yield a CF, as defined by the inequality $P_{ic} > S_{ic}$. None of the simulations for the two main architectures (serial or parallel) consistently displayed a CF. In the few instances where the inequality $P_{ic} > S_{ic}$ held for some serial self-terminating processing orders, the difference between P_{ic} and S_{ic} was negligible. Additionally, results from Experiment 1 show that, for CF-eliciting trials, the proportion of incorrect responses is greater than the proportion of correct responses within these specific trials. This is another feature not evidenced in any of the simulation results.

The reason why standard SFT cannot produce a CF is because of its inherent logic. Serial and parallel processing architectures assume that stimuli are processed through sperate channels, independent of one another. Therefore, a trial response on paired grid trials is based on some classical combination of the separate outcomes of independently processing both grids. This is best shown in Figures 2.5 and 2.6, which show how the trial response on a paired grid trial is derived and combined from the result of individually processing both grids, serially or in parallel. To commit a CF on CF-eliciting trials, an incorrect response on the colour proportions within one of the two grids presented must be given. As these specific trials present one grid with high visual discriminability (easy grid) and one with low visual discriminability (hard grid), providing a correct response hinges on correctly identifying the colour proportion in the grid with low visual discriminability. However, unless the trial responses are combined in some other way, the probability of providing an incorrect response on this paired grid trial cannot significantly exceed the probability of providing an incorrect response when the hard grid is presented by itself. Therefore, the inequality $P_{ic} > S_{ic}$ cannot hold within standard SFT architectures. As such, simulations based on standard SFT processing architectures do not merit further exploration.

It is also important for the reader to note at this stage that standard SFT analysis is a largely comparative process (Harding et al., 2016). The survivor function is calculated to give an RT curve that is then compared to template RT curves, which represent one of several processing architectures. The closer the calculated survivor function RT curve is to one of the template RT curves for the different processing architectures, the more probable it is that the cognitive processing order behind the task is represented by that architecture (Harding et al., 2016; Townsend, 1990). The process undertaken thus far in this project is similar. I used results from the Experiment 1 and a logic flow based on serial and parallel processing architectures to simulate what performance would look like on CF-eliciting trials, if participants were using a serial or parallel processing architecture as defined by standard SFT. The results show that the main qualitative feature of the first experiment, committing the CF as defined by the inequality $P_{ic} > S_{ic}$, cannot be replicated by such simulations. As such, standard SFT architectures are not plausibly representative of the underlying information processing of our CF task. Therefore, standard SFT is not able to provide a comprehensive account of the underlying information processing architecture behind the CF task and possibly CFs in general. However, expanding on the specific processes occurring in these serial or parallel information processes can provide a more accurate picture of how SFT architectures could be extended to accommodate CF results.

Modified SFT simulations

In the first CF experiment (Experiment 1), serial exhaustive models of processing information can overlap with serial self-terminating ones. That is, participants have the freedom to choose when they provide a response in a trial. On all paired grid trials in Experiment 1, participants can arrive at a conclusion to the trial question after observing only one grid. If this occurs, participants can provide a trial response and terminate the trial. As such, a serial self-terminating processing architecture conceptually better captures the processing undertaken by participants in the experiment, if a general serial processing order is used by participants. I believe this also holds for the parallel architectures. As such, I will focus subsequent modelling work on serial and parallel self-terminating architectures specifically. In addition to this, I will propose that within the broad serial and parallel processing architectures for CF-eliciting trials, there is a further effect of order.

In the serial case, on Cf-eliciting trials I propose that the first grid to be processed directly biases the processing of the second grid. More specifically, if participants are presented with two grids, one with much more blue than orange squares and another with a little more orange than blue squares, if they conclude that the first grid has more blue than orange squares, the second grid has the biased perception of having more blue than orange squares. The reason for the sequential order in this bias is to keep this effect within the EAM framework of independent evidence accumulators, while allowing some degree of interaction between them. Furthermore, this interaction between the outcome of one accumulation process on another accumulation process is the distinctly novel idea proposed here. I define the strength of this bias as φ. The higher the value of φ, the greater the strength of the bias on the second grid to be processed.

Present simulations are based on LBA simulations for individual grids, fitted to data from Experiment 1. The LBA model represents a choice between N alternatives (N=2,3, ...) using N different evidence accumulators, one for each response. Each evidence accumulator begins the decision trial with a starting amount of evidence (k), drawn from a uniform distribution on the interval $[0,A]$, that increases at a speed given by the "drift rate" (d). Refer to Figure 2.11 below. This drift rate is drawn from a normal distribution with standard deviation s. Accumulation continues until a response threshold (b) is reached. The first accumulator to reach the threshold decides the overt response, and the time taken to reach the threshold decides the RT (plus some extra constant time for non-decision processes, t_0). Therefore, the time taken for an accumulator to reach the threshold is

$$\frac{b-k}{d} \qquad (3)$$

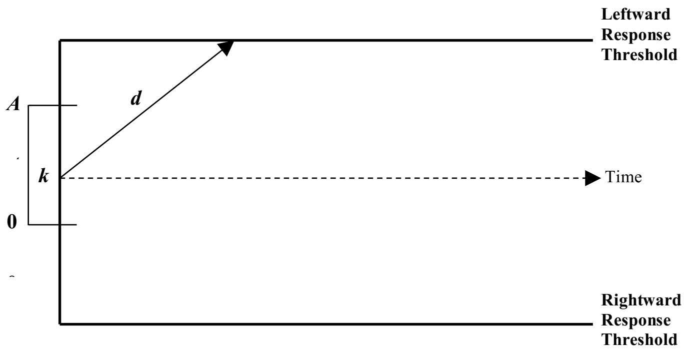

Figure 2.11. A simplified illustration of a binary response EAM and its main components. When a participant begins a binary response trial they have a starting amount of evidence in support of either response. This is represented as k in the figure and is assumed to be at equal distanced between both responses, such that the participant has no bias towards either response at the beginning of the trial. However, k can range between 0 and A. As the trial progresses the participants accumulates evidence in support of either response at some fixed rate, represented as *d*. Evidence here refers to any information derived from the task stimuli in support of either binary response. When the amount of accumulated evidence in support of either response reaches one of the two defined response thresholds, a response is triggered. In the above figure, the amount of accumulated evidence (the solid arrow) reaches the leftward response threshold and consequently triggers the leftward response in the task, e.g. the participant presses the left instead of the right button.

The aforementioned simulated trials were paired grid trials, specifically, CF-eliciting trials, where trial responses were based on simulating responses to an easy single grid and a hard single grid, then combining these separate responses using standard SFT architecture to determine an overall trial response. Using an EAM framework, the judgement that the colour proportion of each grid on a paired grid trial is congruent or incongruent with the trial question is represented by two evidence accumulators: one for the "yes" response and one for the "no" response. The accumulator to reach its respective response threshold first triggers the response for that grid. If a serial processing order is being used on paired grid trials, then a response bias can be introduced by assuming that the decision as to whether the first grid to be processed has a colour proportion congruent or incongruent with the trial question biases our decision on the second grid. For example, assume that the paired grid trial question that must be answered is *"Are there more Blue than Orange patches in both grids?"* Then, if after observing the first grid the accumulator for the congruent response reaches the response threshold first, the decision will be made that this grid has more blue than orange in it. The participant then observes the second paired grid. However, under these biased conditions, the participant's initial decision that the first grid's colour proportions are congruent with the trial question has biased the new accumulator for the second paired grid being observed, by increasing the starting amount of evidence already in support for the accumulator triggering the congruent response. This in effect reduces the amount of time needed for this accumulator to reach its respective response threshold. This is implemented in the model by redefining the time taken for this accumulator to reach its response threshold to

$$\frac{b-k-\varphi}{d} \qquad (4)$$

This simple redefining of the time to response threshold for the second processed grid allows us to easily change the previous simulations to yield results based on this idea of a response bias (termed *biased start point* hench forth), for the second grid being processed under a serial architecture. I provide simulation results for a serial self-terminating processing architecture with a biased start point, because as previously mentioned, the self-terminating architecture more accurately captures the underlying processing of grids.

Results show that for easy then hard and random processing orders, increasing values of φ increases the averaged error rate across simulated CF-eliciting paired grid trials linearly, such that the inequality $P_{ic} > S_{ic}$ holds. The parameter values for the simulation results shown below were based on values chosen to best illustrate the possible plausibility of the model, see Figure 2.12.

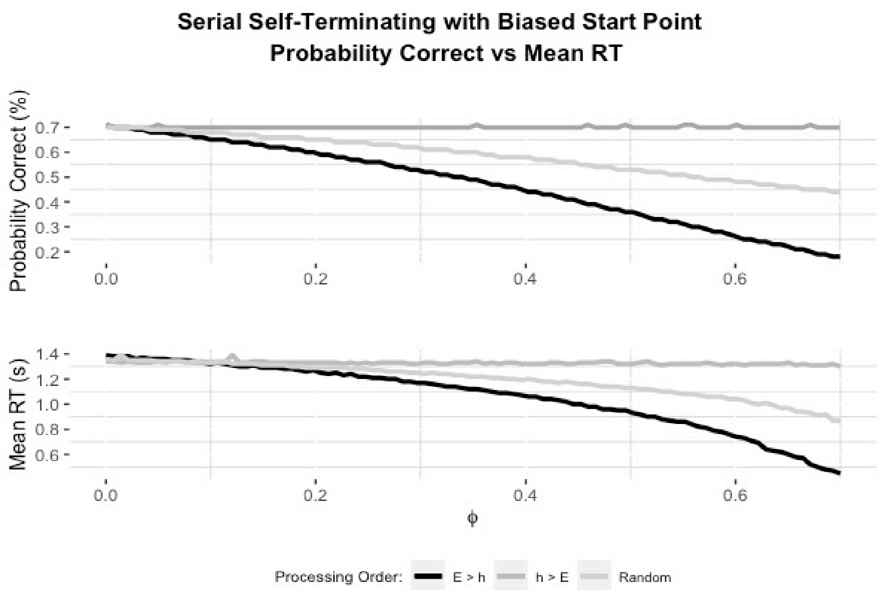

Figure 2.12. The top row shows accuracy rates as φ increases for the three processing orders: easy then hard grid, hard then easy grid and random. The bottom row shows accompanying mean RT rates for increases in φ. Each data point is based on averages over 100,000 simulated CF-eliciting trials.

Results show that for the processing orders of easy then hard grid and random, there are values of φ which cause overall accuracy rates to be such that they satisfy the inequality $P_{ic} > S_{ic}$ for both the simulated and experimental results. These model findings are thereby consistent with the occurrence of a CF. A bias start point appears to have little effect on the hard then easy grid processing order. This in itself is a testable model prediction and will be further explored later. The present focus is to determine if the introduction of a response bias in a SFT serial or parallel information processing architecture can produce error rates on simulated CF-eliciting trials similar to those observed in Experiment 1. For sufficiently large values of φ, the simulation results appear to be able to do so for the serial self-terminating architecture.

However, a biased start point cannot be applied to a parallel information processing architecture. A start point bias can only be applied to the start of the second grid being processed in a paired grid trial, after an initial first grid has been processed to pass on the bias. In a parallel architecture, it is assumed that both grids are processed simultaneously. This removes the possibility of passing on a bias to the start of the second grid being processed, only after processing the first grid. As such, we define the effect of order in a parallel processing architecture as being a result of a collapsing response threshold.

For a parallel processing order on a paired grid trial, if the first grid being processed does not trigger a response, the second processed grid triggers the response. In standard SFT architecture, this is a toggling between self-terminating and exhaustive processing orders. However, just like the similarities between serial self-terminating and serial exhaustive architectures, we argue that even in the parallel processing case, a self-terminating definition better captures the actual processing done by participants on paired grid trials. Nonetheless, participants can switch between self-terminating and exhaustive processing in the first experiment and are not confined to using only one process. In relation to an effect of order existing in this processing architecture, if a response cannot be determined after one of the two grids stops being processed, I assume that the remaining grid will be processed with a collapsing response threshold.

Similar to the modelling of the serial self-terminating simulations with a biased start point, the processing of each grid is represented by two accumulators each: one for the correct response and one for the incorrect response. If one of the two grids finishes being processed and a response is not yet determined, the time to threshold for the second grid and therefore the amount of time it takes to determine a response, collapses at rate γ. In the example where a response is not determined after processing the first grid, the amount of time taken to process the second grid is defined by

$$\frac{b-k+\gamma T_1}{d+\gamma} \qquad (5)$$

where T_1 refers to the time taken to process the first grid.

I classify both the biased start point and the collapsing threshold as an effect of order, because I view them as having a common underlying process. In a serial architecture, on a paired grid trial, the first grid to be processed biases the perception of the colour proportion in the second grid to be processed. Simulation results show that this bias can yield error rates indicative of a CF as observed in Experiment 1. In a parallel architecture, on paired grid trials, both grids are processed simultaneously. If after processing the first grid a response is not determined, the second grid continues being processed and yields the trial response. In the real experiment, the first grid to finish being processed is most likely going to be the easy grid. This is because the easy grid has a high level of visual discriminability and determining the colour proportion in this grid is easier than for the hard grid, which has a much lower level of visual discriminability. It is also known that processing the easy grid alone does not provide sufficient information to determine a response, as while the information provided by the easy grid can technically negate the trial question if it is perceived incorrectly, probabilistically, it is more likely to be correctly perceived by participants. The trial response is therefore determined after processing the hard grid. This gives a processing order similar to the processing order in the serial biased start point architecture: easy then hard grid. The effect of order is then modelled as a collapsing threshold, which in effect decreases the probability of a correct response with time. As a result, this increases the chance of an erroneous decision and therefore increases the error rate on paired grid trials to possibly yield error rates like those in Experiment 1.

Simulation results for a parallel processing architecture with a collapsing bound show that increasing γ decreases the overall accuracy exponentially across simulated CF-eliciting trials, where only one of the two grids is congruent with the trial question. Although, the inequality $P_{ic} > S_{ic}$ does hold for these simulation results, the proportion of incorrect responses in these trials in not greater than the proportion correct, as in Experiment 1. The parameter values for the simulation results shown were based on values chosen to best illustrate the possible plausibility of the model. Refer to Figure 2.13 below.

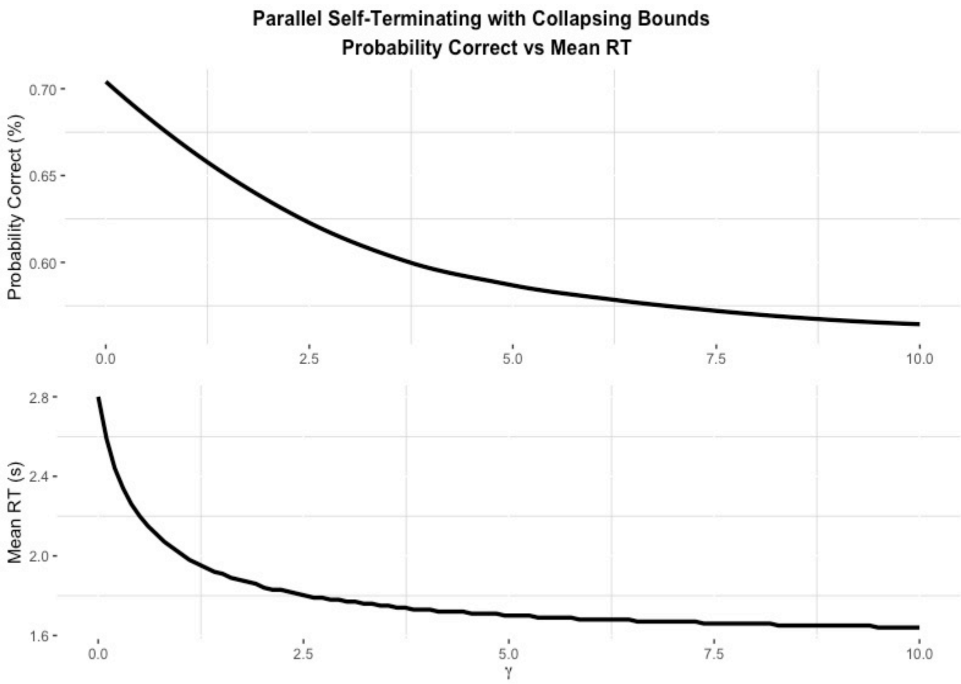

Figure 2.13. The top row shows accuracy rates as γ increases. The bottom row shows accompanying mean RT rates for increases in γ. Each data point is based on averages over 100,000 simulated CF-eliciting trials.

In Experiment 1, participants consistently responded incorrectly to the question on CF-eliciting trials beyond chance, such that there was a greater proportion of incorrect to correct responses on these trials. Comparing this quantitative feature with the simulation results, we see that this feature is only found in the serial architecture with a biased start point. Although a parallel processing architecture with a collapsing bound does cause a decrease in accuracies as γ increases, it does not yield accuracies systematically below chance. As such, an initial assessment of the difference between the experimental results and the augmented SFT simulation results show that a serial processing architecture with a bias start point is a more accurate representation of the information processing that gives rise to the behaviour observed in Experiment 1. This is because the extended serial processing architecture is the only simulation that captures the two key features of Experiment 1: $P_{ic} > S_{ic}$ and a greater proportion of incorrect to correct responses on CF-eliciting trials.

The scope for exploring the extended SFT simulations further is constrained by parameter recovery. The simulation results are based on simulations derived from fitting the LBA to the results of Experiment 1. Each participant only answered 48 single patch trials with a high level of visual discriminability (easy single grid) and 48 trials with a low level of visual discriminability (hard single grid). The LBA fitting procedure was applied to all participant responses on these trials. However, only 12 participants were recruited for Experiment 1, giving a total of 576 trials for the two grid types being fitted. Donkin et al. (2009) showed that approximately 1000-4000 data points per experimental condition are required to yield acceptable levels of bias for parameter recovery. Experiment 1 therefore provided an insufficient number of data points to perform parameter recover with an acceptable level of accuracy, even when compared to the lower end of the advised range. As such, an immediate task is the collection of more data, the result of which I subsequently discuss.

Experimental results. The mean RT for all participants across every condition in the experiment never exceeded 1.6s. Only on 0.4% of all trials were there RT outliers of 5s or more. Therefore, RTs over 5s were removed from the final analysis. The proportion of incorrect responses for all participants for the different trial types is shown in Table 2.7 below.

Table 2.7. Proportion of incorrect responses across all participants.

	Easy Single	Hard Single	CF-Eliciting	Non CF-Eliciting
Percent Incorrect	12%	19%	53%	15%

The conjoint hypothesis (that is, a conjunction) in this experiment is similar to that in Experiment 1 and is represented by CF-eliciting trials. Between paired grid trials and single grid trials, we see a clear difference in averaged responses. Participants, on average, largely performed well on single grid trials that represented conjunct probabilities. However, on trials representing conjunction probabilities, participants on average performed in the opposite direction and responded with a higher number of errors. Recall that error rates function as a measure of CF rates; therefore, participants displaying a greater proportion of errors on CF-eliciting trials compared to hard single grid trials indicate a CF ($P_{ic} > S_{ic}$). To check this behaviour, I performed a series of inferential analyses subsequently described.

A 2(condition: CF-eliciting trials vs non CF-eliciting trials) x 2(target colour: blue vs orange) x 2(distance: far vs close) x 2(side: easy proportion grid is on the left vs right) Friedman test was performed on participants' proportion of correct responses during paired grid trials. A significant main effect of condition was found, $X^2(1) = 61.273$, $p < .001$, with the CF-eliciting trials (*Mdn* = .46) having significantly lower accuracy rates than the non CF-eliciting trials (*Mdn* = 1). A non-significant main effect of distance was found. A non-significant main effect of side was also found. Additionally, a non-significant man effect of target colour was reported. Similar to Experiment 1, these results provide evidence in support of H1.

I then examined the difference in the proportion of incorrect responses between CF-eliciting trials and hard single grid trials, to determine if error rates for CF-eliciting paired grid trials were significantly different, consistent with H2. To accomplish this, a non-parametric Wilcoxon Signed Rank Test was performed on error rates for CF-eliciting trials and hard single grid trials. The results showed that participants in the CF-eliciting trials had a significantly higher proportion of incorrect responses (*Mdn* = 43.75) compared to single hard trials (*Mdn* = 21.35), $W_s = 229$, $p < .001$. Again, I interpret this as a CF, as expected given the design. Recall, this is H2, according to which error rates on CF-eliciting trials should be higher than error rats on hard single grid trials.

Consistent with Experiment 1 and with committing a CF, participants had significantly higher error rates when assessing CF-eliciting trials (conjunction probabilities) compared to hard single grid trials (conjunct probabilities). Furthermore, accuracy rates were again significantly lower on CF-eliciting trials compared to non CF-eliciting trials. These results initially show that the CF can indeed be successfully transposed to the psychophysical domain, consistent with results from Experiment 1. Additionally, these results again show a clear difficulty in combining conjunct probabilities into conjunction probabilities, consistent with the original CF phenomenon.

Modified SFT fits

The two LBAs within each of our two modified models were fitted to the data from Experiment 2 using an identical procedure to that in Experiment 1. Similarly, the main objective here was to determine whether a modified serial or parallel SFT information processing architecture could produce error rates indicative of a CF, as in Experiment 1. Specifically, if I simulate performance for easy and hard single grid trials in Experiment 1, and then combined performance using our modified SFT architectures with an effect of order to simulate paired grid trials, can we get an averaged error rate similar to Experiment 1 and indicative of a CF? Additionally, I again wanted to assess if such simulations can yield reasonably realistic RT data.

I first plotted the RTs for correct and incorrect responses during easy and hard single grid trials for Experiment 2 and those produced by standard LBA simulation models using parameters from the fitting procedure. Simulation results yielded satisfactory fits to the data. Refer to Figure 2.14 below.

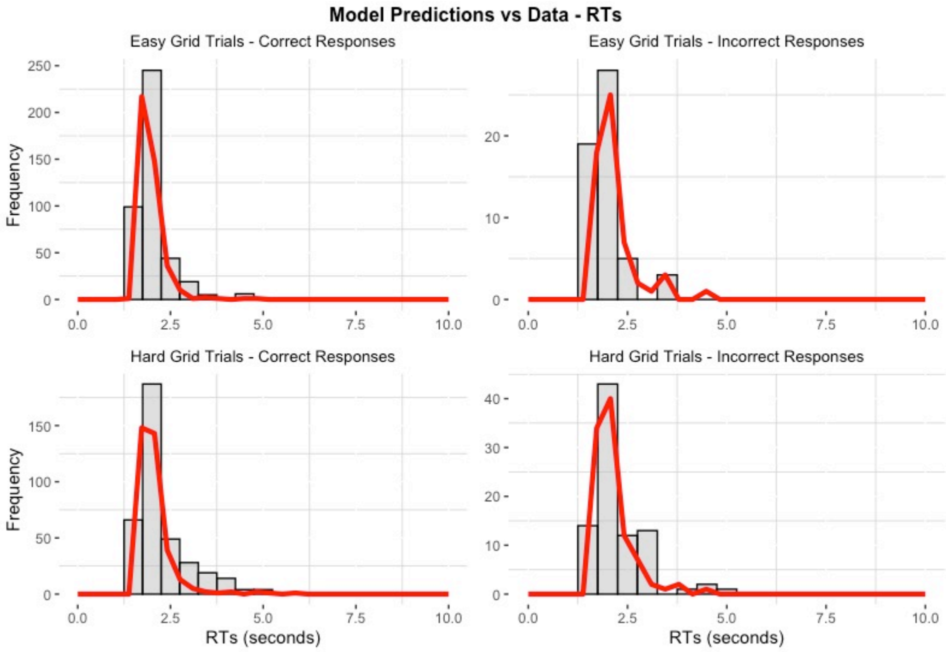

Figure 2.14. The RT fits between correct and incorrect responses during easy and hard single grid trials in Experiment 2 and the standard LBA simulation models, using parameters from the fitting procedure. The histograms represent the experimental data and the red lines represents the model results.

The quantile probability (QP) plots also show satisfactory combined accuracy and RT fits between the experimental data on easy and hard single grid trials, and standard LBA simulation models using parameters from the fitting procedure. Refer to Figure 2.15 below.

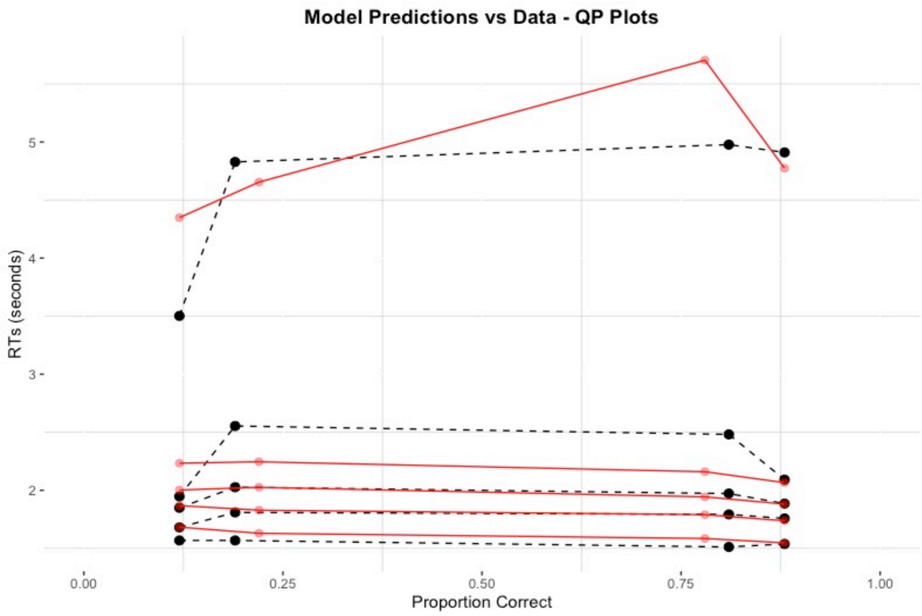

Figure 2.15. Quantile probability plot of the experimental data and the model results. The black points represent the experimental data and the red points represent the model results. The two points on the right-hand side of the x-axis indicate the proportion of correct responses for hard and easy single grid trials respectively, at the five different RT quantile values. The two points on the left-hand side of the x-axis indicate 1 minus the proportion of correct responses, for easy and hard single grid trials respectively, at the five different RT quantile values.

The LBA models used to simulate easy and hard single grid trials were then combined using a SFT serial self-terminating or parallel self-terminating information processing architecture, with an effect of order. Similar to the first set of simulations, this was done to simulate CF eliciting paired grid trials, where the colour proportion in only one grid is congruent with trial question. As these types of trials were explicitly designed to elicit CFs. More specifically, these paired grid trials are generally defined by one grid where it is visually easy to determine that the colour proportion in this grid is congruent with the trial question (easy grid) and another grid where it is visually hard to determinate that the colour proportion in that grid is not congruent with the trial question (hard grid).

Results for the serial self-terminating architecture with an effect of order (biased start point model) again show that for certain values of φ similar error rates to those found in Experiment 2 can be simulated. However, these similar error rates can only be identified in the specific processing order where the easy grid is processed first, refer to Figure 2.16 below.

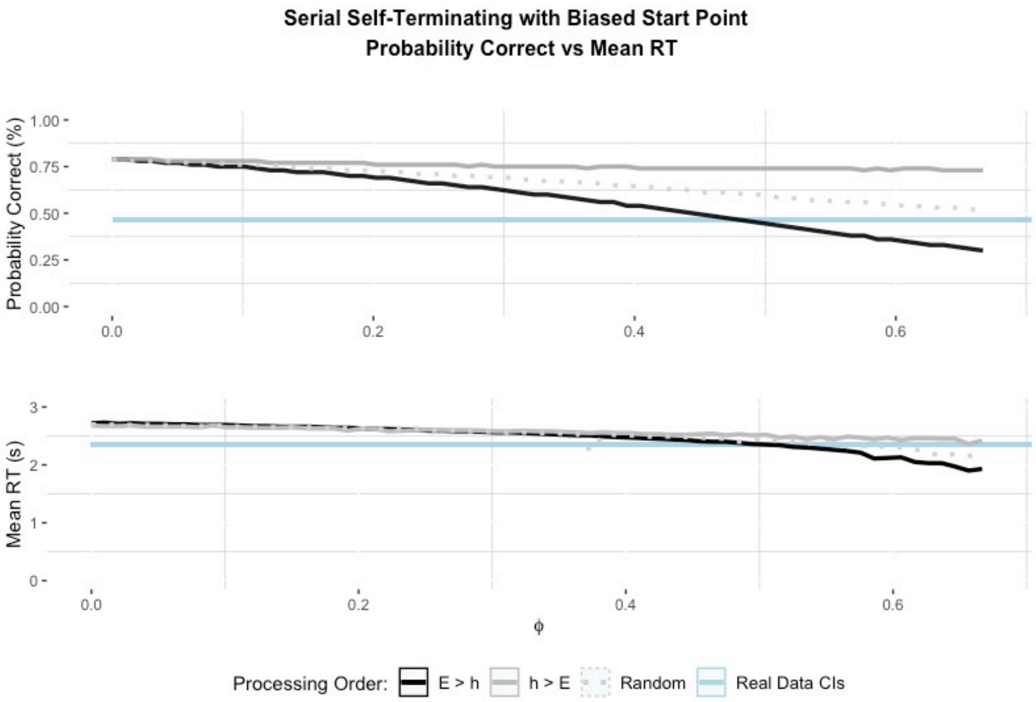

Figure 2.16. The top row shows accuracy rates as φ increases for the three processing orders: easy then hard grid (E > h), hard then easy grid (h > E) and random. The bottom row shows accompanying mean RT rates against increases in φ. Each data point is based on averages over 100,000 simulated CF-eliciting trials. The upper and lower edge of the light blue bar represent the upper and lower 95% confidence intervals for the data from Experiment 2.

Figure 2.16 shows that there exists a value of φ where the mean error rate and RT simultaneously yield results within the confidence intervals of the real data, somewhere within the φ range of 0.4 – 0.5. However, this is only true for the easy then hard processing order. Refer to Figure 2.17 below for a close up of the φ range that produces simulation results similar to the results observed in Experiment 2.

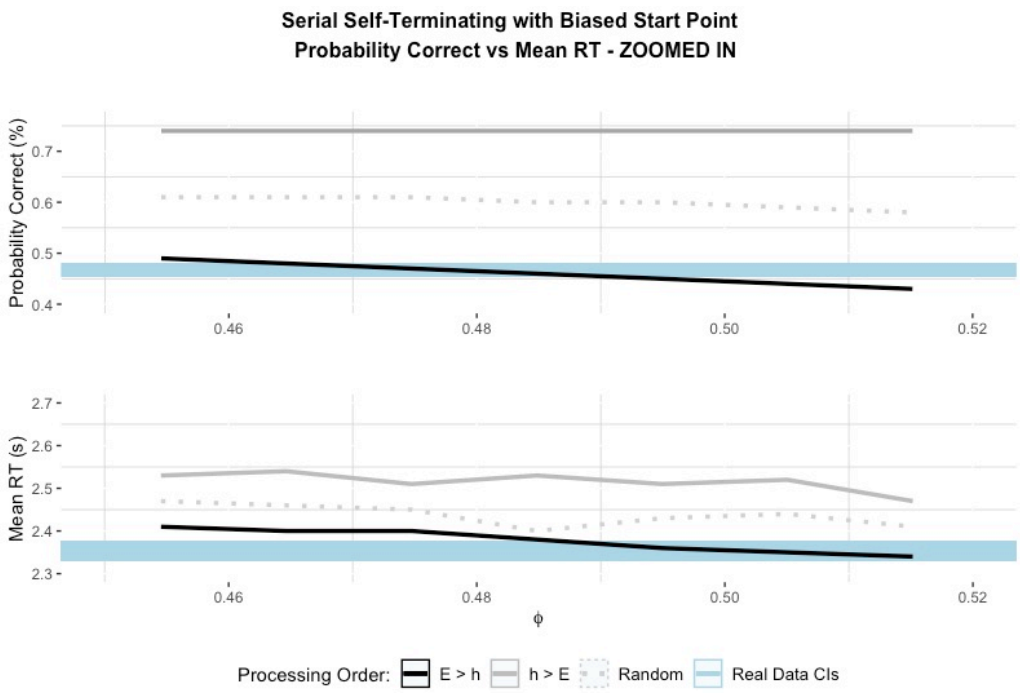

Figure 2.17. The top row shows accuracy rates as φ increases within a smaller range for the three processing orders: easy then hard grid (E > h), hard then easy (h > E) grid and random. The bottom row shows accompanying mean RT rates for increases in a smaller φ range. Each data point is based on averages over 100,000 simulated CF-eliciting trials. The upper and lower edge of the light blue bar represent the upper and lower 95% confidence intervals for the data from Experiment 2. This figure more clearly illustrates that there are certain values of φ only for the easy then hard processing order, which simultaneously lay in the confidence intervals for the experimental accuracy and RT data.

Results for the parallel self-terminating processing architecture show that there is no value of γ where mean error rates and RT simultaneously yield results within the confidence intervals for the experimental accuracy and RT data. Please refer to Figure 2.18 below.

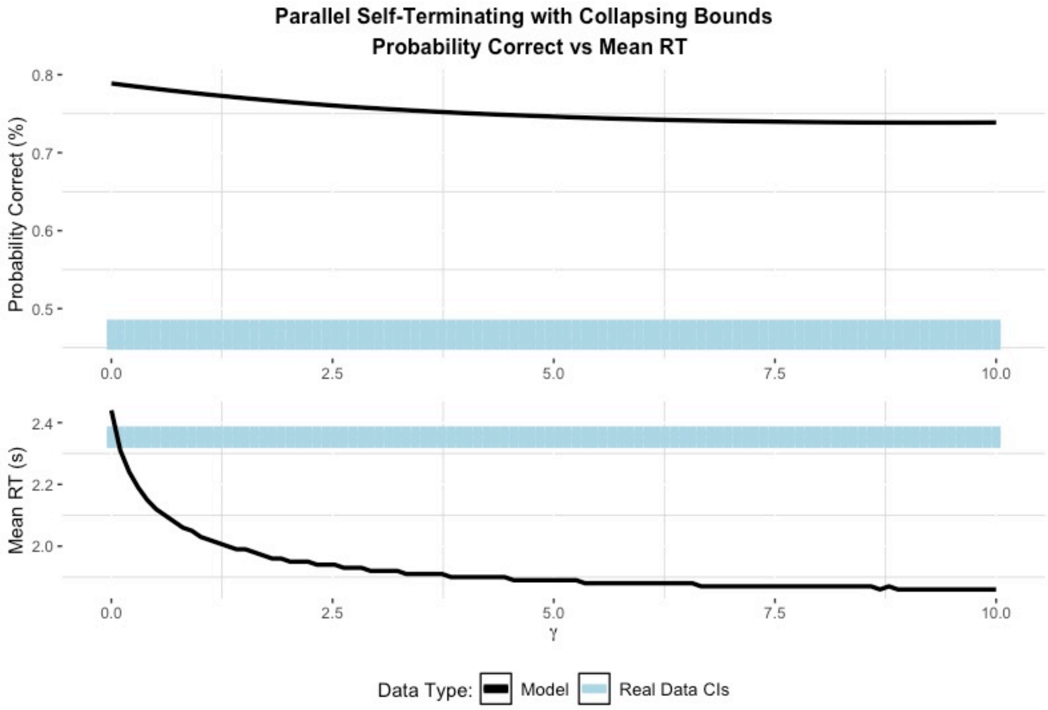

Figure 2.18. The top row shows accuracy rates as γ increases. The bottom row shows accompanying mean RT rates against increases in γ. Each data point is based on averages over 100,000 simulated CF-eliciting trials. The upper and lower edge of the light blue bar represent the upper and lower 95% confidence intervals for the data from Experiment 2.

2.2.3 Discussion

The preliminary modelling results from Experiment 1 showed that, within the paired grid trials, those that were designed to elicit a CF had a different rate of information processing relative to trials that were not designed to elicit a CF. The purpose of this second experiment was to further explore the underlying information processing associated with the CF task, to determine whether a particular processing order was eliciting the observed CF behaviour. I subsequently used SFT architectures to determine this ordering.

The first objective was to simulate paired grid trials designed to elicit a CF by combining separate models for processing single grid trials, which were fitted to the data from single grid trials in Experiment 1. I combined these models by creating an algorithm based on the various SFT information processing architectures. I then identified two main quantitative features that were unique to the present task and indicative of a CF. Firstly, the CF was observed if the proportion of incorrect responses on CF-eliciting trials was higher than the proportion of incorrect responses on hard single grid trials, $P_{ic} > S_{ic}$. Secondly, the proportion of incorrect responses on these CF-eliciting trials should be greater than the proportion of correct responses on these same trials. If a standard SFT architecture could plausibly replicate performance in Experiment 1, then these two quantitative features had to be observed in the simulation results. The simulation results showed that standard SFT architectures were not capable of consistently replicating the two main quantitative features in Experiment 1. However, this was unsurprising as the inherent logic of standard SFT architecture is unable to yield error rates indicates of a CF.

The assumption made for the modelling results of Experiment 2 is that committing a CF is dependent on completely processing the easy grid first, on CF-eliciting trials. This in turn suggests that participants are processing stimuli in this particular order for the majority of the CF-eliciting trials, but not always, as error rates would be 100% in such a case. How participants are able to identify whether a grid is "easy" or "hard" is represented by the non-decision time in EAMs. Non-decision times represent the non-decision aspects of the task, such as motor responses and stimulus encoding (Evans, 2019). As such, they occupy a period in time before and after standard stimulus processing. It is assumed that the most salient properties of a stimulus, such as general colour distribution, are identified and processed during this non-decision time. For the present task, it is therefore assumed that participants identify the easy grid during this non-decision time.

The only distinguishing feature between Experiment 1 and 2 was the introduction of a calibration stage at the start of Experiment 2. The purpose of this was to produce data with higher internal validity. I believe the calibration stage was successful in accomplishing this, as evidenced by the consistent descriptive results for Experiment 2. Experiment 2 also aimed to provide substantially more data points, for model fitting. This was as intended, as we recruited an additional 8 participants, which increased the total number of data points from Experiment 1 to 2 by 3,840, a 40% increase.

Two modified SFT serial and parallel self-terminating processing architectures were proposed based on an assumption that there exists some effect of order capable of inflating error rates to levels similar to those found in Experiment 1. In the serial self-terminating architecture with an effect of order (biased start point), I assumed that the decision for the first grid to be processed biased the decision for the second grid to be processed. For example, if on a paired grid trial the first grid is perceived as having more blue than orange squares in it, the second grid is also more likely to be perceived as having more blue than orange squares in it. In the parallel self-terminating architecture with an effect of order (collapsing bound), the first grid to finish being processed causes the response threshold for the last grid to finish processing to collapse exponentially with time. For example, if on a paired grid trial the left hand-sided grid finishes being processed first, but provides insufficient information to determine the trial response because the colour proportion within it is congruent with the trial question, the second grid continues being processed, but the response threshold begins to collapse exponentially with time. This causes the amount of time that can be spent processing the relevant stimulus information to collapse rapidly. Both these models are assumed to increase error rates in principle and represent the two main features of Experiment 1.

Instead of basing these modified models on a combination of LBA models fitted to the data on single grid trials in Experiment 1, I selected best fitting parameters through trial-and-error. This was done to initially identify if there existed a parameter set that could possibly be extracted from fitted data, that could allow the modified models to yield results closely matching the data from Experiment 1. The results showed that there were different parameter sets for both modified models that could yield error rates to replicate the two main features of Experiment 1, for some values of the bias parameter.

Experiment 2 was therefore focused on replicating the findings from Experiment 1 and providing better calibrated experimental data, to which the modified SFT models could be fitted. The descriptive and inferential results from Experiment 2 were similar to those in Experiment 1. Fitting the LBA models to the single grid data from Experiment 2 yielded different parameter sets for the processing of easy and hard single grids. Additionally, these parameter sets (one for the easy single grid and another for the hard single grid) were able to yield results that replicated the two main features of Experiment 1, for the serial self-terminating biased start point model only. The parallel self-terminating collapsing bound model was not able to replicate similar results with the same parameter sets. Furthermore, the biased start point model made a specific prediction as to the order of information processing, which can yield error rates indicative of a CF. That is, only if the easy (high visual discriminability) grid is processed before the hard (low visual discriminability) grid on CF eliciting paired grid trials can a CF be observed, according to the model.

Therefore, the objective of the subsequent experiment is to test this prediction. Specifically, the prediction that the easy-then-hard processing order is the only processing order in the serial biased start point model, that can replicate the two main quantitative features of the present CF task.

It is interesting to note that the finding that the CF is associated with a bias in the serial processing of stimuli is largely consistent with the literature on serial processing in the visual domain. For example, Whitney's (2012) findings on serial dependence in visual perception are consistent with my findings in this chapter. Specifically, Whitney (2012) argues that the visual system attempts to preserve visual continuity, such that our perception of a sequence of stimuli has a perceptual attraction towards the preceding stimuli. In other words, there is a perceptual bias towards the preceding stimuli. This is partly why the experiments in this chapter were not based on a more common random dot motion task, so motion perception bias could be minimised in the task. As Whitney (2012) states, by biasing our current perception towards what was previously observed, our cognitive system attempts to compensate for variability in visual input that might disrupt perceptual continuity (Manassi, Liberman, Kosovicheva, Zhang & Whitney, 2018).

The serial vs parallel debate in information processing is not a novel concept and predates the SFT work by Townsend and his colleagues, especially in the visual domain (Broadbent, 1958; Treisman, 1969). The emphasis on the seriality of information being processed in both perception and short-term memory search increased during the early 1960s, with the introduction of the information processing paradigm. However, much of this literature relates to the claim that accuracy rates and RT curves on cognitive tasks largely indicate a serial information processing order in general (Sagi & Julesz, 1987; Sperling, 1963, 1967), whereas later research has explored evidence for parallel (as well as serial) processing.

The argument that the serial processing order is partly responsible for producing CFs is further supported by findings from the Quantum Cognition framework (Pothos and Busemeyer, 2013; Pothos and Busemeyer, 2022). Quantum cognitive models generally assume a serial processing order. Consistently with this picture, the experimental and modelling results in this section show that a bias for the serial processing order in visual perception uniquely produces a CF. That is, conversely, a parallel processing order is not able to produce error rates in the present task indicative of a CF. Note, even though the different stimuli are presented simultaneously, their processing times are not equal. Therefore, if one stimulus is processed before a paired stimulus in a parallel system, the stimulus still has the 'opportunity' to influence the yet unprocessed stimulus, in a manner indicative of serial processing. However, results showed that such a visual biasing effect is not sufficient in a parallel processing system to produce a CF and that these CF-eliciting effects are unique to a serial processing system. This is a unique, interesting conclusion from this section.

Section 2.3 The Conjunction Fallacy and the Effect of Order

Introduction

The purpose of this experiment is to determine whether the occurrence of a CF in our psychophysical CF task is dependent on the stimuli being processed in a specific order. That is, can a CF only be observed if the easy (high visual discriminability) grid is processed before the hard (low visual discriminability) grid, on CF-eliciting trials?

In order to test this, I replicated Experiment 2, but removed the single grid trials and altered the stimulus presentation of paired grid trials. As stimulus processing order was examined only on paired grid trials, single grid trials beyond calibration trials were redundant and therefore removed from Experiment 3. The stimulus processing order was manipulated by presenting one of the two paired grids by itself for a few milliseconds, before simultaneously presenting the second paired grid.

As Experiment 2 predicts that the easy then hard processing order facilitates the CF, I additionally wanted to assess the extent to which error rates on CF-eliciting trials, indicative of a CF, are dependent on this specific processing order. The reasoning here is that if processing the easy grid first biases participant's perception of the second grid to be processed, then even if the second grid to be processed has a 50/50 visual discriminability ratio, participants should still have a biased perception of the colour proportion in that second grid. As such, on these trials we would still expect to observe error rates indicative of a CF or at least errors higher than on paired grid trials where this ordering does not occur. Conversely, according to the model prediction from Experiment 2, one would not expect to see error rates indicative of a CF when the first grid to be processed is the hard grid and the second grid to be processed has a 50/50 visual discriminability ratio. The assumption here is that the first grid to be processed biases perceptions of the second grid in a similar direction. For example, if the first grid to be processed is an easy grid with more blue than orange patches in it, the participant is not only more likely to perceive the grid as having more blue than oranges patches in it, but will have a bias towards perceiving the second paired grid in a similar way.

In order to assess these assumptions, I will introduce a series of pseudo-experimental trials. These trials will present participants with paired grids. On one set of trials participants will be presented with the easy visual discriminability grid first and then presented with a second grid with a 50/50 visual discriminability ratio. On another set of trials participants will be presented with the hard visual discriminability grid first and then presented with a second grid with a 50/50 visual discriminability ratio. The hypotheses for this experiment are as follows:

H1: Only when the easy visual discriminability grid is shown before the hard visual discriminability grid will participants display error rates indicating a CF, as defined in the previous experiments.

H2: Only when the easy visual discriminability grid is shown before the 50/50 visual discriminability grid will participants display error rates indicating a CF.

2.3.1 Method

Participants

20 participants were recruited via the online participant recruitment platform Prolific. Each participant was paid approximately £20 pounds for their participation. The sample size was exploratory, as there is no prior work with manipulations similar enough to the present ones.

Design and procedure

The experiment was based on a within participants design with three factors: target colour (V1), condition (V2) and processing order (V3). V1 had two levels: "*Are there more ORANGE than blue patches in both grids separately?"* and "*Are there more BLUE than orange patches in both grids separately?*" V2 had two levels: experimental vs control trials. V3 had three levels: the paired grid with hard visual discriminability was shown 800 milliseconds before the grid with easy visual discriminability (HE), the paired grid with easy visual discriminability is shown 800 milliseconds before the grid with hard visual discriminability (EH), the order in which the grids are pressed is randomised (RR). Refer to Figure 2.19 below for a flow chart of the presentation order in a trial.

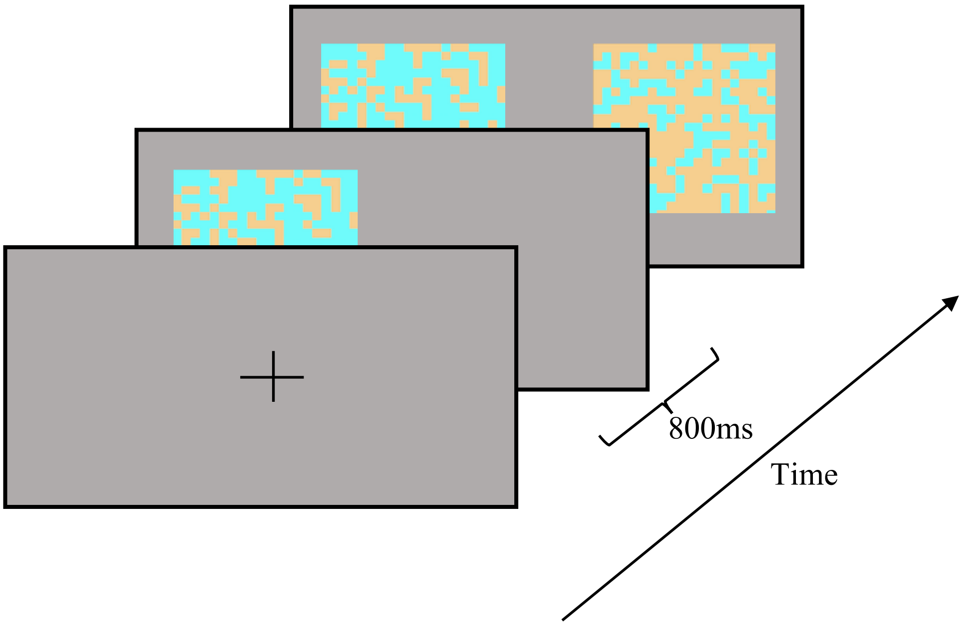

Figure 2.19. A flow diagram of the stimuli presentation order for on a CF eliciting paired grid trial.

The experiment was divided into three stages: calibration stage, block 1 and block 2. The calibration stage here was identical to that in the second experiment. Each of the two blocks contained three sets of 72 trials, totalling 432 trials. Each of the three sets in a block were setup to facilitate three distinct methods of processing the stimuli: EH, HE and RR. Trials were randomly presented within each block. Blocks 1 and 2 were identical, except that the target colour for each block changed.

The trial set EH contained 24 experimental paired grid trials. Half of these trials were presented with grids such that one corresponded to easy discriminability and the other to hard. The other half of these trials were presented with grids corresponding to the other level of easy vs hard discriminability proportions. The trial set HE was identical to the trial set EH, except for reversing the stimulus presentation order. The trial set RR contained 6 trials for each of the two trial types from the trial sets EH and HE. This gave a total of 24 experimental trials for the trial set RR.

Each of the three trial sets additionally contained 48 control trials. 24 trials were presented such that both grids were shown with 12 trials for each of the two trial sets EH and HE, so that there was a greater blue to orange proportion in each grid. An additional 24 trials were presented where both grids were shown with 6 trials for each of the two trial types from the trial sets EH and HE, with a greater orange to blue proportion in each grid.

Each block contained an additional 36 pseudo-experimental trials. These consisted of 18 paired grid trials where the grid with easy visual discriminability was presented before a paired grid, with a 50/50 blue-to-orange colour proportion (E50/50). Half of these trials were presented with the paired easy discriminability grid having one of two easy discriminability proportion levels and the other half presented at the other level. Another 18 paired grid trials were included, where the grid with hard visual discriminability is presented before a paired grid with a 50/50 blue-to-orange colour proportion (H50/50). Half of these trials were presented with the paired hard discriminability grid having one of the two hard discriminability proportion levels and the other half presented at the other level. As such, the entire experiment had a sum total of 504 trials.

On 50/50 trials, paired grid trials were presented, where one grid had either an easy or hard discriminability blue-to-orange colour proportion and the paired grid had 50/50 blue-to-orange colour proportion. Therefore, for any of the two trial questions (e.g. "are there more blue than orange patches in both grids"), the correct answer is always "no". The pseudo experimental trials were included to test the main prediction made by the S-BSP model: processing the easy grid first should bias perception of the second grid to be processed. Therefore, even if the second grid to be processed has a 50/50 visual discriminability ratio, participants should still have a biased perception of the colour proportion in that second grid. These pseudo-experimental trials check this assumption.

Hard discriminability grids were set for each participant individually, through a calibration stage at the start of each experiment, based on trials which would lead to error rates systematically below chance. Easy discriminability grids were set for each participant individually, through an analogous calibration stage at the start of each experiment, based as trials which would lead to error rates systematically above chance. 50/50 trials, as the name suggests, are supposed to produce performance at chance level. In either case, if one simply assumes that the perception of the second grid is biased by the perception of the preceding grid, then one ought to expect error rates on trials where the hard visual discriminability grid precedes the 50/50 grid to be substantially lower, than when the easy grid is processed first. The hard grid always has less of the target colour in the trial question. Therefore, in this case one would expect a participant to perceive the second grid as also negating the trial question. Now, because negating the trial question is always the correct response on these trials, error rates ought to be low. However, to anticipate our results, they were not and this result is surprising. These results further support the argument that a perception bias may be present, but has a stronger effect on the EH processing order.

2.3.2 Results

Experimental results

As the data was not normally distributed, a 2(target colour) x 2(condition) x 3(processing order) Friedman test on accuracy rates was performed to assess the effect of stimuli processing order on CF rates (represented as error rates in the task). A significant main effect of condition was found $X^2(1) = 72.602$, $p < .001$, with CF-eliciting trials having lower accuracy rates (*Mdn* = .58) than non CF-eliciting trials (*Mdn* = 0.97) A significant main effect of processing order was found $X^2(2) = 5.143$, $p < .05$. A non-significant main effect of target colour was found. Table 2.8 below shows the results of a post-hoc Bonferroni–Holm *t*-test performed on the significant main effect of processing order.

Table 1.8. Results of a post-hoc Bonferroni–Holm *t*-test performed on the significant main effect of processing order.

Process Order		MD	SE	t	d	p_{bonf}
HE	EH	0.17	0.03	6.777	1.52	< .001
	RR	0.06	0.03	2.233	0.50	0.095
EH	RR	-0.12	0.03	-4.544	-1.02	< .001

A 2(target colour) x 3(processing order) Friedman test was performed on accuracy rates to assess the effect of initially processing the easy or hard visual discriminability paired grid on CF rates, within the pseudo-experimental trials. A significant main effect of processing order was found $X^2(2) = 19.203$, $p < .001$. A non-significant main effect of target colour was found. Table 2.9 below shows the results of a post-hoc Bonferroni–Holm *t*-test performed on the significant main effect of processing order.

Table 2.9. Results of a post-hoc Bonferroni–Holm *t*-test performed on the significant main effect of processing order on pseudo-experimental trials. The RR (random) processing order accuracy rates were calculated for each participant as the average between the two other processing orders.

Process Order		MD	SE	t	d	p_{bonf}
E50/50	H50/50	-0.29	0.05	-5.390	-1.21	< .001
	RR	-0.07	0.05	-1.280	-0.29	0.625
H50/50	RR	0.22	0.05	4.110	0.92	< .001

Modified SFT fits

I fitted the serial biased start point model (S-BSP) to all 20 participants individually. No participants were removed from the final fitting procedure but RTs < 0.5s and RTs > 5s were removed from the procedure. The S-BSP model was fitted to each participant using a gradient descent optimisation algorithm. The S-BSP model consists of two LBA models, one for the easy grid and one for the hard grid, for paired grid trials. Each LBA has five parameters: the starting amount of evidence (A), the response threshold (b), the non-decision time ($t0$), the drift rate sampling noise (s), and the biasing parameter (φ).

There were an additional four drift rate parameters that were fitted for each LBA. For the easy grid these were: the drift rate for correct responses when the grid had more much more blue-than orange (*Bdc*), the drift rate for incorrect responses when the grid had more much more blue-than orange (*Bde*), the drift rate for correct responses when the grid had more much more orange-than-blue (*Odc*), and the drift rate for incorrect responses when the grid had more much more orange-than-blue (*Ode*).

For the hard grid these were: the drift rate for correct responses when the grid had slightly more blue-than-orange (*bdc*), the drift rate for incorrect responses when the grid had slightly more blue-than-orange (*bde*), the drift rate for correct responses when the grid had slightly more orange-than-blue (*odc*), and the drift rate for incorrect responses when the grid had slightly more orange-than-blue (*ode*).

Gradient descent is an optimization algorithm for finding a local minimum of a differentiable function. More simply, gradient descent finds the values of a model's parameters that minimize the discrepancy function that calculates the error between the actual data and the model predictions (Haji & Abdulazeez, 2021; Izzo, Zou, & Ying, 2021). The process starts by defining some initial arbitrary parameter values for the chosen model. From there on, the gradient descent algorithm iteratively adjusts the initial parameter values, so that they minimize the discrepancy function. The discrepancy function used in this fitting procedure was defined as

$$\sum_{g=1}^{8}[(A \cdot RT_C^D - RT_C^M)_g^2 + (1 - A \cdot RT_I^D - RT_I^M)_g^2 + (100 \cdot A^D - A^M)_g^2] \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

where g extends over the 8 different trial conditions in the experiment (four pairings for CF-eliciting and non CF-eliciting trials each) D = data, M = model, I = incorrect and A = accuracy. C was a constant defined to be 100. A variant of the gradient descent algorithm was used called the Limited Memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS). This variation was used as it could both perform the gradient descent function and allow for upper and lower boundaries to be placed on the estimated parameter values (Haji & Abdulazeez, 2021).

Equation 6 represents an arguably sensible discrepancy function, as it captures various constant features of the model, such as response threshold, non-decision time etc, and variable features, such as drift rates across different trial types. Additionally, given that the main aspects of performance on the present CF task related to accuracy, slightly more weight was given to accuracy data in the equation. Nonetheless, RT data was still included in the function. Furthermore, the discrepancy function included a comparison between the observed and model data, for all eight CF-eliciting and non CF-eliciting trial types. Although more refined discrepancy functions could be created through more in-depth simulation analysis, I believe this function represents a good starting point function for this model.

The fitting procedure yielded reasonably good accuracy model fits to the real data for the experimental CF eliciting paired grid trials, as seen in Figure 2.20 below.

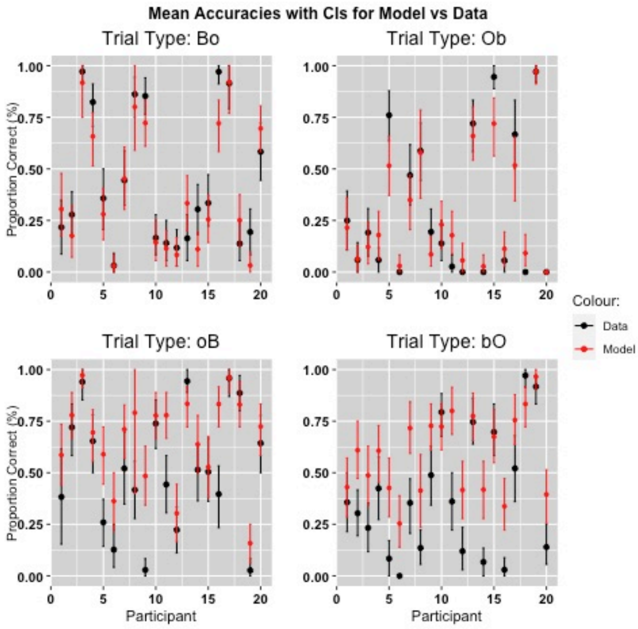

Figure 2.20. S-BSP model vs real data accuracy for the four CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

The different CF-eliciting trial types represent the four different stimuli presentation orders that were used to elicit a CF: there were much more blue than orange grids preceding the little more orange than blue grids (Bo), the reverse order (oB), the much more orange than blue grids preceding the little more blue than orange grids (Ob) and the reverse order (bO). The fitting procedure also yielded reasonably good RT model fits to the experimental data, for the CF-eliciting trials, refer to Figure 2.21 below.

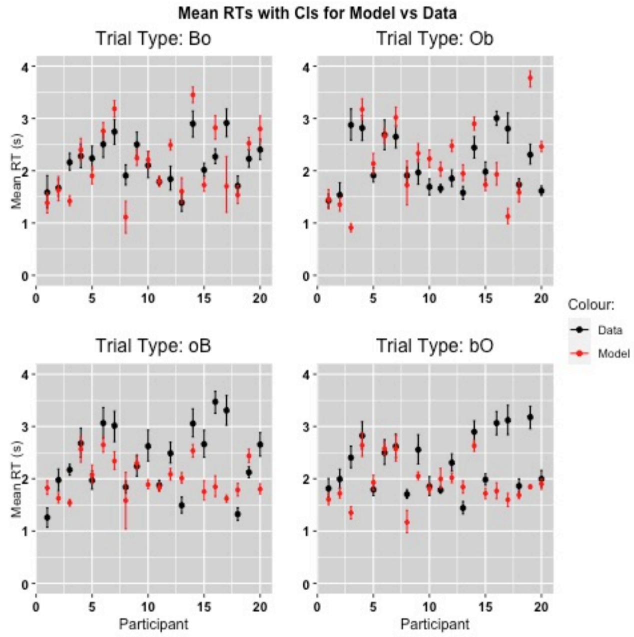

Figure 2.21. S-BSP model vs actual data mean RTs for the four different CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

In order to assess the effect of the biasing parameter, φ, I plotted φ against the mean RTs for the easy then hard (E > h) minus the hard then easy (h > E) processing orders for the model: Bo-oB and Ob-bO, refer to Figure 2.22 below. The purpose was to assess how increases in φ affected the different processing orders in relation to RTs.

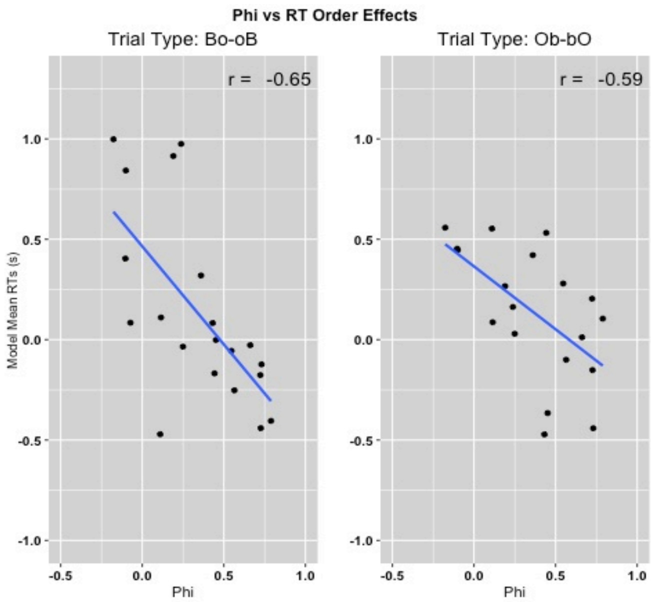

Figure 2.22. φ vs the mean RTs for the easy then hard (E > h) minus the hard then easy (h > E) processing orders (Bo-oB and Ob-bO), fitted with a regression line.

The results show that the difference in the mean RTs between CF-eliciting trials decreases as φ increases. These results show that there appears to be an interaction between φ and participant RTs. As previous descriptive and inferential analyses show that only the (E > h) processing order yielded a CF, results on the relationship between φ and mean RT may be indicative of the underlying cause of the CF. For example, for the (E > h) processing order, increases in φ may be responsible for incrementally increasing RTs by introducing uncertainty or noise in the processing of the stimuli due to an effect of order.

The fitting procedure yielded reasonably good accuracy model fits to the real data for non CF-eliciting trial types. Model fits performed better for the (E > h) processing order compared to the (h > E) processing order. refer to Figure 2.23 below.

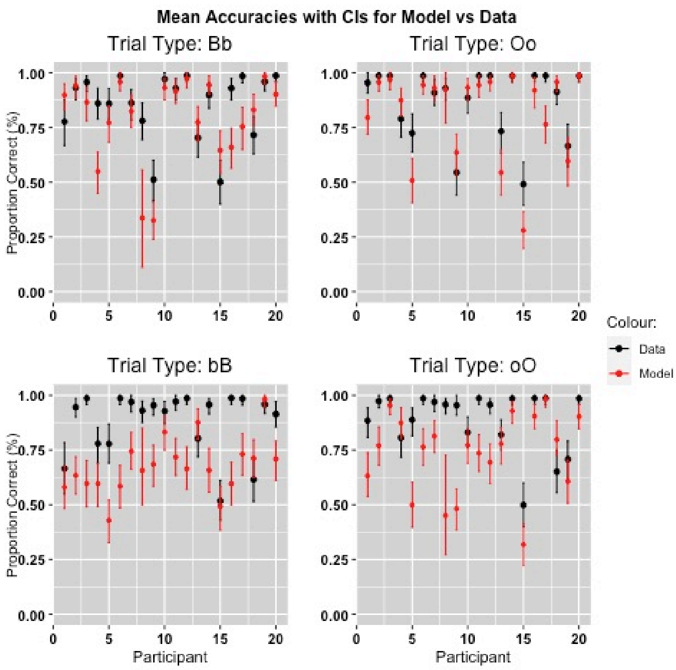

Figure 2.23. S-BSP model vs real data accuracy for the four different control paired grid trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

The different non CF-eliciting trial types represent the four different stimulus presentation orders that were used to not elicit a CF: a much more blue than orange grid preceding the little more blue than orange grid (Bb), the reverse order (bB), a much more orange than blue grid preceding the little more orange than blue grid (Oo) and the reverse order (oO). The fitting procedure also yielded reasonably good RT model fits to the real data for the non CF-eliciting trials, refer to Figure 2.24 below.

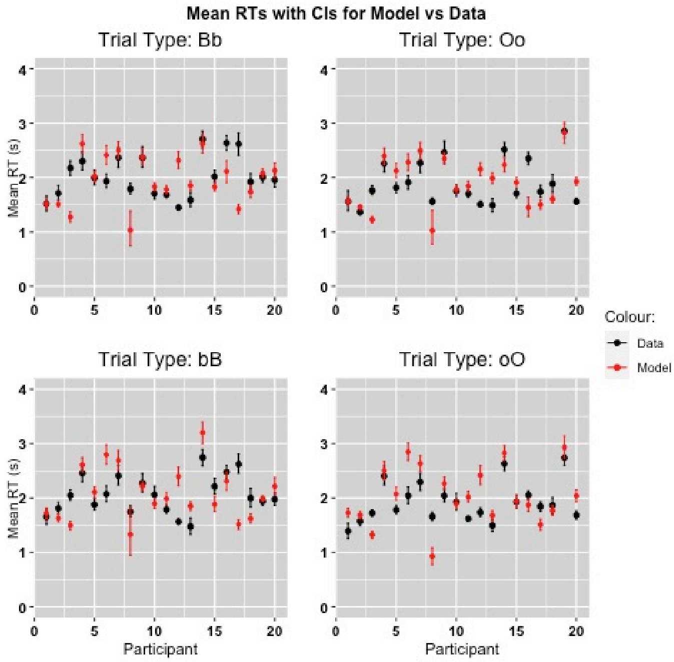

Figure 2.24. S-BSP model vs real data mean RTs for the four different control paired grid trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

For comparison, we fixed the biased start point parameter, φ, to be zero. This in effect removed the biasing effect in the model and reduced it to the standard SFT serial self-terminating architecture. The fitting procedure for the equivalent standard SFT architecture yielded significantly worse accuracy model fits to the real data for CF-eliciting trials, compared to the S-BSP model, refer to Figure 2.25 below.

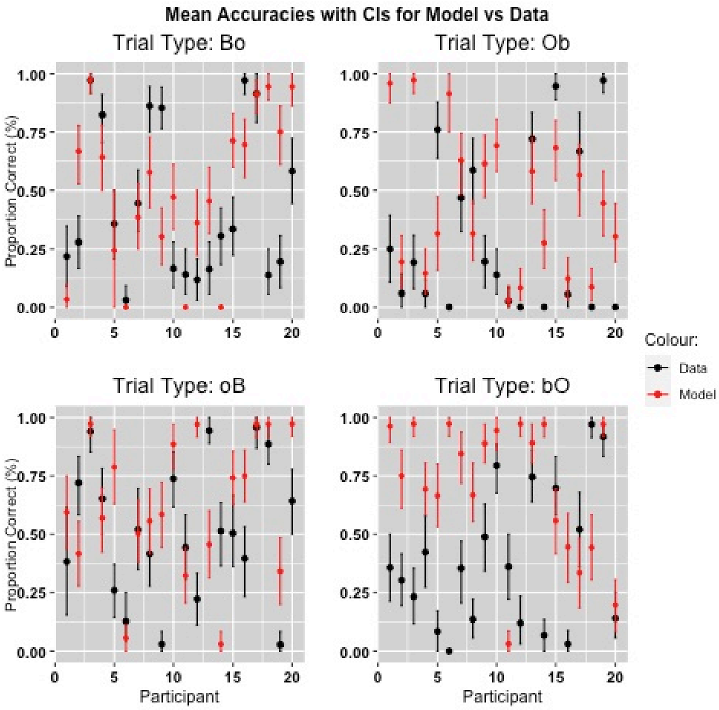

Figure 2.25. Standard serial self-terminating SFT model vs real data accuracy for the four different CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

The fitting procedure also yielded poorer RT model fits to the real data for the CF-eliciting trials compared to the S-BSP model, refer to Figure 2.26 below.

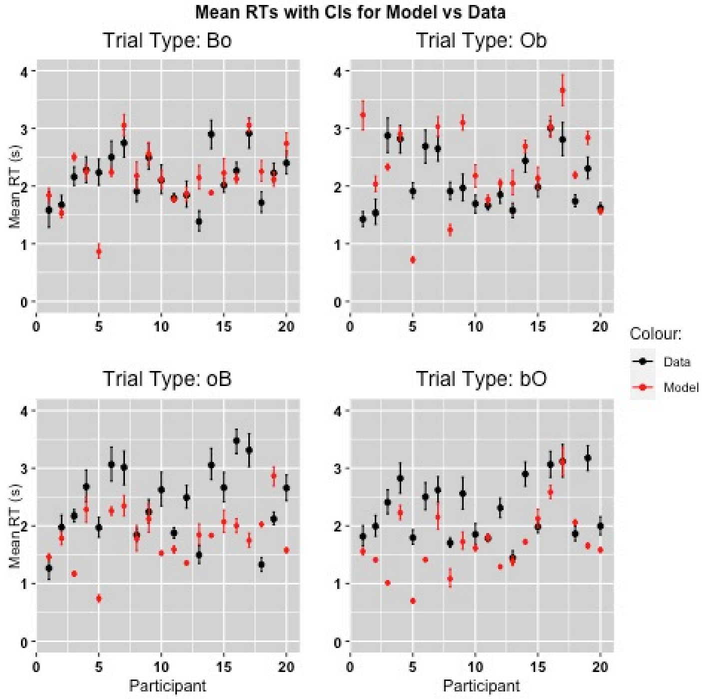

Figure 2.26. Standard serial self-terminating SFT model vs real data mean RTs for the four different CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

The fitting procedure yielded poorer accuracy model fits to the real data for non CF-eliciting trials compared to the S-BSP model, refer to Figure 2.27 below.

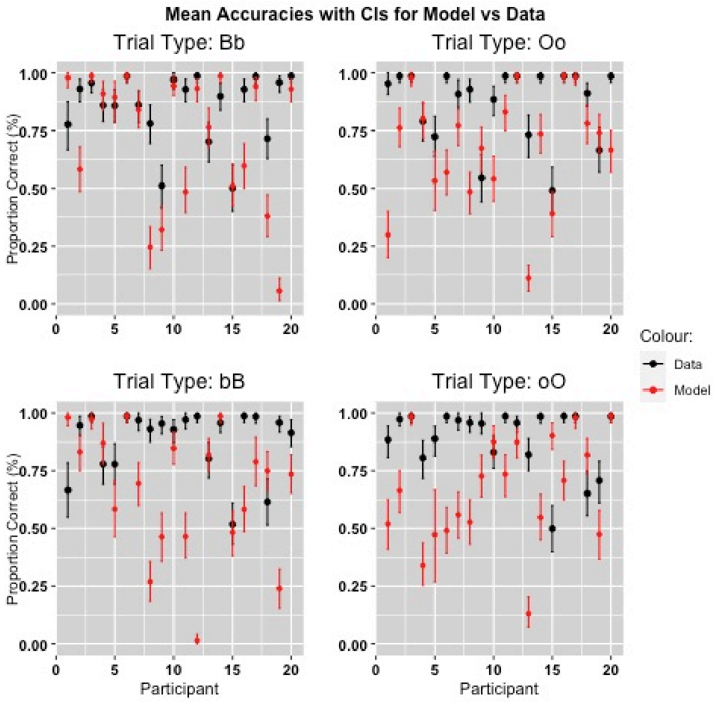

Figure 2.27. Standard serial self-terminating SFT model vs real data accuracy for the four different non CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

The fitting procedure also yielded poorer RT model fits to the real data for the CF-eliciting trials compared to the S-BSP model, refer to Figure 2.28 below.

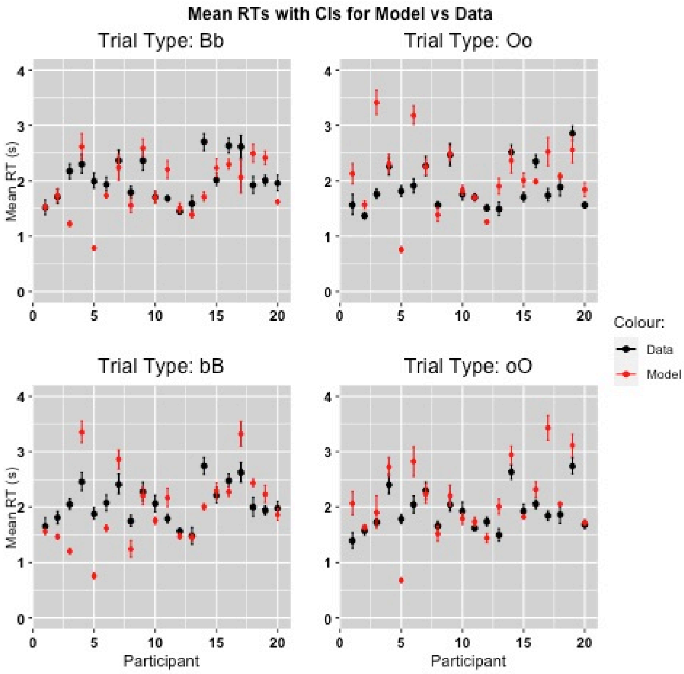

Figure 2.28. Standard serial self-terminating SFT model vs actual data mean RTs for the four different non CF-eliciting trial types for each participant. Results are shown with bootstrapped 95% confidence intervals (CIs).

To summarise, the experimental results show that the (E > h) processing order produced significantly higher incorrect responses, and therefore significantly higher CF rates, compared to the other two processing orders. These results also showed that for the pseudo-experimental trials, the (E50/50) processing order still produced significantly higher incorrect responses compared to the other processing orders.

In line with these findings, the (E > h) processing order in the S-BSP model was fitted to the experimental data, to assess if the model estimates could reproduce the experimental findings. Additionally, we examined whether the (E > h) processing order in the S-BSP model yielded better accuracy and RT fits compared to the standard SFT (E > h) processing algorithm. The results showed that the (E > h) processing order in the S-BSP model produced better accuracy and RT fits to the data than the standard SFT (E > h) processing order fits. Refer to Table 2.10 below for a comparison of the Bayesian Information Criterion (BIC) between the data and the model estimates for the two models.

Table 2.10. BIC scores for the S-BSP and standard SFT model.

S-BSP	SFT
-30092	-31670
-29525	-30795
-28556	-29959
-30600	-31099
-29939	-31428
-30591	-30997
-29205	-30853
-30267	-30647
-29596	-31439
-29461	-30742
-30009	-30967
-28981	-30962
-28839	-30829
-29461	-30947
-29585	-31948
-29269	-30409
-28742	-29689
-28684	-31104
-29712	-30531
-27837	-30666

BIC values were calculated for both models using RSS scores and are presented in Table 2.10. The results show that BIC scores were consistently lower for the S-BSP model, indicating a better model fit for the S-BSP model over the standard SFT model.

2.3.3 Discussion

With this experiment I wanted to assess if a CF can only be observed if the (E > h) processing order is used to process CF-eliciting paired grid trials. In order to do this, the paired grid trials were shown in such a way that one of the paired grids was briefly shown before being shown alongside the paired grid. Two stimulus presentation orders were presented to facilitate the (E > h), (h > E) and RR processing orders on experimental paired grid trials: either the easy (high visual discriminability) grid proceeded the hard (low visual discriminability) grid (Bo, Ob), or the reverse (bO, oB).

The experimental results support the assumption made in the conclusion of Experiment 2: the (E > h) processing order in the S-BSP model is responsible for producing sufficiently high error rates indicative of a CF. These findings are further supported by the results from the pseudo-experimental trials. These trials show that observing sufficiently high incorrect response rates (indicative of a CF) is contingent on the serial processing of the easy visual discriminability grid first. Moreover, this is irrespective of the degree of visual discriminability in the paired second grid.

Additionally, the pseudo-experimental trials revealed that when the second grid to be processed on CF-eliciting trials had an indistinguishable 50/50 colour ratio, if the easy visual discriminability grid was presented first, error rates were significantly higher than when the hard grid was presented first. Remember, that an effect of order bias is assumed to cause the second processed grid to be perceived in a similar way to the first processed grid. Furthermore, the strength of this bias is assumed to be greater for the (E > h) processing order. However, there remained the possibility that error rates on CF-eliciting trials, where the second grid to be processed is the hard visual discriminability grid, were largely the result of the low visual discriminability in the second processed grid. Results for the pseudo-experimental trials show that even when the second processed grid on these trials had a 50% chance of being perceived as orange or blue, the bias made participants perceive the colour proportion as largely similar to the first grid to be processed. Thereby indicating a bias in the processing order of the stimuli, that is not entirely dependent on the low visual discriminability in the second processed grid. In all, these results further support the modelling and general inferential findings of this experiment. That is, it is a potential bias that occurs during information process within a specific order that facilitates error rates indicative of a CF.

These findings also imply that the CF may not be a result of some bias or error that occurs when combining conjunct probabilities into conjunction probabilities in general. Rather, that a bias results when a specific information processing order is performed cognitively. This is evidenced by the modelling results in Experiment 2, and the inferential and modelling results from Experiment 3, where only the error rates for the (E > h) processing order was significantly higher than all other processing orders. The modelling findings also revealed that the model systematically underpredicts accuracy on non CF-eliciting trials. There are some possible reasons for this. Firstly, it may be the case that performance on non CF-eliciting trials fluctuated more than on CF-eliciting trials. As such, model fits to the data may be unable to adequately capture and replicate the nuanced behaviour for these control trials overall. There is always a possibility that an alternative model fitting procedure may be able to provide better fits to the data, such as a Maximum Likelihood Estimator. I highlighted this as a possible limitation of the present work, which employed a gradient descent fitting procedure.

These findings also build on the preliminary results of Experiment 1. In Experiment 1, modelling results showed that there was a significant difference in the rate of information processing (drift rates) between all single grid trials and all paired grid trials. Note, that the drift rate specifically represents the rate at which information is processed. Therefore, given that paired grid trials are made up of single grid trials, one would not assume that the rate of information processing for each single grid would significantly change only because they are presented in pairs. However, in the context of the results of the present experiment, if a bias in the information processing system did occur during certain paired grid trials, then one could expect this bias to affect the information processing rates too. As the modelling results in the present experiment and Experiment 2 suggest, if the second grid to be processed on certain CF-eliciting trials is perceived in a biased way, then one can assume that the way in which information is processed and combined for this grid would change. However, why this is represented as a slower rate of information processing (lower drift rate) in Experiment 1 is harder to determine. As a bias in colour perception would be expected to decrease assumed task difficulty and thereby increase the rate at which information in favour of one of the two binary responses in the task is processed. One possibility is that the bias increases the probability of perceiving the colour proportions in the second processed grid as being similar to the first processed grid and thereby increases the noise in the overall processing system. This increase in noise brought on by the introduction of a bias could function to independently decrease the rate at which information processing occurs.

In the present experiment, the easy and hard visual discriminability grids were created as psychophysical abstractions of the two main features of the original Linda problem: Linda is a feminist and Linda is a bank teller respectively. The colour discriminability of each grid represents their congruence with the trial question in the present experiment: e.g. are there more blue than orange patches in both grids? In other words, are the proportions in the easy and hard grids congruent with the trial question. In the Linda problem the different ranked statements about Linda represent their congruence to the description of Linda. In the present experiment, the easy visual discriminability grid is more congruent with the trial question. In the Linda problem, the statement about her being a feminist is more congruent with the general description of Linda. In effect, the present findings suggest that the congruence, or similarity, of processed features directly manipulate our perception of them if they are processed in a specific order. That is, if the features most congruent with a main comparison feature are processed first, they can bias our judgements about them such that if the order in which they were processed was reversed, our judgements would be significantly different. Reducing the Linda problem from a high to low level cognitive task has revealed the CF to be a product of an effect of order bias on feature congruence.

Chapter Three: Evaluating Non-Normative Decision-Making

Section 3.1 Evaluating Rationality with Positive Interference

Introduction

Understanding the descriptive and rational foundations of human decision making has been a key objective for scientists and philosophers essentially since antiquity. It is mostly uncontroversial that Bayesian probability theory is the correct approach (Oaksford & Chater, 2007) and indeed a substantial body of evidence has accumulated that non-human animals approximate Bayesian inference, e.g., in foraging or predation risks (Ramirez & Marshal, 2017; Valone, 2006). Unfortunately, humans, unlike non-human animals, are faced with such a staggering range of questions that (baseline) Bayesian inference quickly becomes intractable. This problem has been recognized by behavioural scientists very early, leading Simon to propose that humans can, at best, be bounded-rational (Simon, 1955) and Tversky, Kahneman, and their colleagues to initiate an influential research programme into apparent Bayesian paradoxes into human behaviour (e.g., Kahneman, 2001; Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1983). The ensuing debate has been recognized by no less than three Nobel prizes (for Simon in 1978, Kahneman in 2002, and Thaler in 2017).

The tension between Bayesian prescription and human intuition is well illustrated by the famous conjunction fallacy (Tversky & Kahneman, 1983). When naïve participants are presented by a hypothetical person described as a feminist (F), but not as a bank teller (BT), they generally conclude that $Prob(F\&BT) > Prob(BT)$. That such an inference appears nonsensical is immediately obvious if we recast the problem in a set-theoretical form (Tentori, Bonini, & Osherson, 2004): how can participants decide that it is more likely that a Scandinavian person would have blond hair and blue eyes, than just blond hair? Yet, as Gould (1988) noted, “I know that the conjunction is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—‘but she can’t be just a bank teller; read the description”’(p. 469). There are several analogous results, referred to as fallacies.

One way to explain fallacies is to recast Bayesian theory in more psychological terms, notably by acknowledging the limited capacity of human minds (Griffiths et al., 2010; Lieder & Griffiths, 2019; Tenenbaum et al., 2011). An influential approach has been to posit that probabilities are inferred from internal sampling processes. But such processes are noisy, therefore the produced probabilities are only approximately Bayesian (Costello & Watts, 2014; Zhu, Sanborn, & Chater, 2020). For example, consider the law of total probability, that is relevant in the conjunction fallacy: Bayesian theory requires that $Prob(BT) = Prob(F\&BT) + Prob(\sim F\&BT)$, from which it is obvious that we can never have $Prob(F\&BT) > Prob(BT)$. However, if we rewrite the law of total probability as $Prob(BT) = Prob(F\&BT) + Prob(\sim F\&BT) + noise$, then clearly there is a possibility that the noise term can balance out other terms, so that a conjunction fallacy emerges. Costello and Watts' (2014) proposal is along these lines and can readily capture both the conjunction fallacy and related results.

Another way to explain fallacies is to consider alternative systems for probabilities. The pioneering physicists who developed quantum mechanics realized Bayesian probabilities were inapplicable to microscopic physical systems – so, they developed an alternative system for probabilistic inference, which we can call quantum theory. In principle, quantum theory is relevant whenever there is a need to formalize uncertainty. In cognition, quantum models have been successfully applied in cases which appear problematic from a classical perspective (Busemeyer & Bruza, 2011; Haven & Khrennikov, 2013; Pothos & Busemeyer, 2013).

A characteristic feature of quantum models is the emergence of *interference* effects, when two or more questions are considered together. Consider the two-slit experiment in physics, one of the key early discoveries in quantum physics. Identical particles are emitted towards a plate with two slits. Further away from the plate, there is a detector screen which counts the number of particles arriving at a particular position. Classically, we would expect that each particle goes through one slit or the other, so that the detector screen simply records 'shadows' of the two slits. In practice, the detection probability when both slits are open is not a simple sum of the probabilities when only a single slit is open, resulting in the famous interference pattern, which signals a failure of the classical idea that these events are disjoint (Feynman et al., 2005). The key observation is that something new is introduced when we have two possible ways an event can happen (a particle at a position on the detector screen), which was not present when there was only a single possible way for the event to occur.

Concerning the conjunction fallacy, using a quantum approach, we can write the law of total probability as $Prob(BT) = Prob(F\&BT) + Prob(\sim F\&BT) + \mathcal{I}$, where $\mathcal{I}$ is an interference term. Depending on the value of $\mathcal{J}$, a conjunction fallacy is allowed (Busemeyer et al., 2011) and, generally, it is possible to cover several results problematic from a Bayesian perspective (Pothos & Busemeyer, 2022).

It might appear that sometimes a quantum approach works better than a Bayesian one. However, sharp conclusions are complicated by several factors, including the fact that it is possible to fairly smoothly relate Bayesian and quantum models (Trueblood, Yearsley, & Pothos, 2017) and that the latter can be seen as a local version of the former (Pothos et al., 2021). The most compelling way to make the case for a distinct role of quantum theory in cognition is with a priori, parameter-free empirical tests. So far, only Wang, Solloway, Shiffrin, and Busemeyer (2014) have provided such a test, based on the so-called quantum question (QQ) equality. Considering two binary questions presented in two possible orders, the QQ equality concerns the probabilities for various possibilities for responding to the questions. Quantum theory requires that $\left[P\left(A_{yes}\ \&then\ B_{no}\right) + P\left(\ A_{no}\ \ \&then\ B_{yes}\right)\right] - \left[P\left(B_{yes}\ \&then\ A_{no}\right) + P\left(B_{no}\ \&then\ A_{yes}\right)\right] = 0$ and, fairly impressively, this seems to be true in several settings.

The purpose of the present work is to report a novel, a priori, parameter-free empirical test for quantum theory. To explain the general idea, consider again the conjunction fallacy: classically, one could utilize noise to explain a reported conjunction fallacy, concerning two questions A, B. However, we expect $noise = constant$, that is, we do not expect noise to systematically vary, depending on the questions. If we consider three pairs of exhaustive and mutually exclusive events (i.e., $p(A \cup B \cup C) = 1$; note, we use the symbol $\cup$ to indicate disjoint disjunctions, that is disjunctions between mutually exclusive events), there is nothing we could say about the relation between $noise_{AB}$, $noise_{BC}$, $noise_{CB}$. By contrast, quantum interference terms systematically vary, depending on the questions. A surprising mathematical result shows that, as long as the three-way disjunction is fixed, the interference terms for the three question pairs *always* combine in a specific way, $\mathcal{J}_{AB} + \mathcal{J}_{BC} + \mathcal{J}_{BC}$, to exactly balance the other probabilities. Below, we develop a corresponding empirical test. Importantly, the test is not limited to a comparison between Bayesian and quantum theory, but encompasses a class of related theories, augmented by different kinds of bias and/or noise.

Linear versus bilinear probability theories

There are infinite systems for how to assign probabilities to events, by which we mean not that there are multiple values that could be assigned to a given event, but rather that there are multiple formal systems for doing so. One way to classify and order different systems is in terms of the complexity of the interference terms, which correspond to the complexity of *interaction* between different questions (Sorkin, 1994).

The first, trivial member of this hierarchy is a probability 'model' which sets all probabilities to 0. The next member is the general linear model, which assumes that probabilities are computed as $p(A) = f(A) + \varepsilon$, where ε is a constant and could reflect noise in judgment. The linearity property is expressed as:

$f(\cup_i A_i) = \sum_i f(A_i)$……………………………………………………………….(1)

The crucial characteristic of Equation (1) is that probability assignment is only a function of the event in question (e.g., A). Therefore, the probability assigned to a disjoint disjunction of events is a function of the probability of each event separately. Equation (1) is an expression of linearity in Bayesian theory (more formally σ-additivity, Kolmogorov, 1933/1950) and, without noise, leads to the usual disjunction rule in Bayesian theory $p(\cup_i A_i) = \sum_i p(A_i)$. The general linear model encompasses most Bayesian theory variants, for example, noisy Bayesian probabilities (Costello & Watts, 2014; Zhu et al., 2020).

The next member in the hierarchy is the general bilinear model, for which $p(A) = g(A, A) + f(A) + \varepsilon$, where $f(\cdot)$ is as above and $g(\cdot,\cdot)$ is linear in both its arguments (bilinearity):

$g(\cup_i A_i, B) = \sum_i g(A_i, B)$, $g(A, \cup_i B_i) = \sum_i g(A, B_i)$ ………………………………..…….(2)

For a disjoint disjunction, we now have $p(A \cup B) = g(A \cup B, A \cup B) + f(A \cup B) + \varepsilon = g(A, A) + g(B, B) + g(A, B) + g(B, A) + f(A) + f(B) + \varepsilon = p(A) + p(B) + g(A, B) + g(B, A) - \varepsilon$. Here, probability depends both on the probability of each event separately and on the interaction/ interdependence between the two evens, expressed by the terms $g(A, B)$ and $g(B, A)$ – this is a property that was absent in the general linear model. It can be proved that QPT is a bilinear probability theory (Sorkin, 1994) and the most famous example of bilinearity is the two two-slit experiment in quantum mechanics, whereby when both slits are open we do not just observe a pattern which is a simple sum of the patterns when individual slits are open.

This hierarchy can continue indefinitely. Next, we would have a general trilinear model. If we imagine a screen with three slits, we can ask whether the detection probability when all three slits are open can be expressed in terms of the probabilities when only each slit is open individually and when only pairs of slits are open. If the answer is no, then something fundamentally new is introduced by the possibility of three alternatives. If the answer is yes, we might conclude that we can study probability theory by just using two events and nothing is gained by adding more alternatives. Note, trilinear probability models have been developed (Dakic et al, 2014; Lee & Selby, 2017), but overall there has been limited effort to go beyond bilinear models (Hardy, 2001).

The relation between general linear, general bilinear, and more complex probability models can be captured by the extent of interaction between events. We can formalize this idea by defining interference terms in the computation of disjoint disjunctions (note, it is the $\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3$ quantities that we refer to as interference terms). The first three such terms are given by:

$$\mathcal{I}_1(A) \equiv p(A)$$
$$\mathcal{I}_2(A, B) \equiv p(A \cup B) - p(A) - p(B)$$
$$\mathcal{I}_3(A, B, C) \equiv p(A \cup B \cup C) - p(A \cup B) - p(B \cup C) - p(A \cup C) + p(A) + p(B) + p(C) \quad (3)$$

For Bayesian probabilities, $\mathcal{I}_2(A, B) = -\varepsilon$, so that interference does not reflect any interaction between the events (it is not a function of the events), but is rather a constant, which might correspond to noise or a response bias. For basic Bayesian theory $\varepsilon = 0$. Having $\varepsilon \neq 0$ allows probabilities to deviate from Bayesian prediction (Costello & Watts, 2014). For quantum theory probabilities, *even though* $\mathcal{I}_2(A, B)$ *varies depending on the events A, B*, we have $\mathcal{I}_3(A, B, C) = \varepsilon$. That is, the different pairwise interference terms conspire so that their sum always balances the sum of the other probabilities (specifically the marginals, Appendix 1) and does not depend on any interaction between the alternatives, over and above two-way interference effects. This is a surprising and powerful a priori prediction, which can distinguish general linear models (such as Bayesian theory) from general bilinear models (such as quantum theory). For basic quantum theory, $\varepsilon = 0$ (Appendix 1) and $\varepsilon \neq 0$ can be used to capture noise/ response biases, as for Bayesian theory. More generally, Sorkin (1994) showed that for a level n theory (i.e., a probability model based on an n-linear function), the quantity $\mathcal{I}_n(A, B, C \dots)$ depends on the events A, B …, but the quantity $\mathcal{I}_{n+1}(A, B, C \dots)$ is constant.

We can ask how these different interference terms translate to intuitions about psychological process. General linear models assume that each event A is treated independently. In general bilinear models, considering two events together changes their meaning, compared to having them individually. In QPT, a disjunction $p(A\ or\ B)$ has to be understood as $p(A\ or\ then\ B)$, so that the presence of the earlier event reveals perspectives or thoughts which alter our understanding of the subsequent one (Busemeyer et al., 2011). So, bilinear inference requires the mental flexibility to nuance the meaning of an event, depending on a preceding event – this is a kind of contextuality. In general trilinear models, having events A, B, C together changes their meaning compared to both having them in pairs and individually. It may appear that the nuancing and contextual perception required for trilinear inference goes beyond what the human mind is capable of, but this is an empirical issue.

It might seem simpler to have constant interference terms, as in Bayesian theory. Is there an adaptive perspective to the complexity/ contextuality from non-trivial interference terms in quantum? Actually, quantum probabilities can be shown to be algorithmically simpler than Bayesian ones, using fairly basic assumptions (Pothos et al., 2021). Additionally, quantum probabilities can offer better models of environmental statistics, when measurements disturb the relevant system, e.g., when just asking a question changes the mental state of the responder (Pothos et al., 2017). But there is a further, subtler adaptive consideration, which we explore next.

In quantum theory the disjunction rule is $p(A \cup B) = p(A) + p(B) - \mathcal{I}_2(A, B)$, so that a disjunction can be stronger or weaker (depending on the sign of the interference term), compared to Bayesian theory. The strength of joint probabilities is a measure of the strength of the causal connection between the two events. In Bayesian theory, causal connections are modeled with Bayesian networks (Pearl, 1988) and the strength of such connections impacts on how probabilities propagate through the network. In quantum theory, interference terms allow causal connections between events so strong or weak that the Bayesian sum rules are violated (e.g., Busemeyer et al., 2011). There is some prior suggestive work that this is psychologically relevant. Kareev and colleagues (2000; Kareev, Lieberman, & Lev, 1997) argued that, because of working memory limitations, correlations between events are computed across likewise limited samples, which means that correlations are typically overestimated (because the sampling distribution for correlation is skewed). This has been argued to be adaptive, because it can help with the early detection of associations in nature (Alloy & Tabachnik, 1984; Lopes, 1982).

Overview of empirical tests

Using Equation (3), we can derive expectations for the two-way and three-way interference term for disjoint disjunctions, for Bayesian theory, Bayesian theory with a simple form of response bias/ noise (which is independent of the events A, B), and Bayesian theory with an elaborate (which allows some dependence on A, B, but is still more constrained compared to bilinear probability models; Table 1, Appendix 2). Thus, we can offer a test for a wide class of general linear models. Furthermore, we can derive corresponding expectations for basic quantum theory and quantum theory with a simple response bias. Specifically, quantum theory requires that $\mathcal{I}_3(A, B, C) = constant$, regardless of the events A, B, C. Thus, we obtain a general, a priori test of quantum theory in cognition, which extends the work of Wang et al. (2014). Note, the use of disjoint disjunctions does not reduce generality, since there is a close connection between disjunctions and other probabilities, e.g., classically, $p(A \cup B) = 1 - p(\sim A \cap \sim B)$.

We asked participants to consider the probability of ailments A, B, C in the accident and emergency (A&E) ward in fictional places. The probability of $Prob(A \cup B \cup C)$ was fixed to 1, that is, a patient would suffer from at least one of A, B, or C. Hypothetical common causes were presented for pairs of ailments, {A,B}, {B,C}, {A,C} (Experiments 1, 2). A common cause should affect $Prob(A \cup B)$, but would it do so beyond what is allowed by general linear theories? Common causes should also affect the two-way interference terms, but would they do so beyond what is allowed by general bilinear theories (which require that two-way interference terms combine in a certain way)? Specifically, the experimental design involved three different scenarios (Fig 3.2), each drawing attention to common causes between different pairs of events. Bayesian theory with a simple response bias requires that $\mathcal{I}_2(A, B)$ should be the same, regardless of the events; Bayesian theory with an elaborate response bias allows $\mathcal{I}_2(A, B)$ to vary with different events but requires interference terms to be consistently negative. Both quantum theory variants allow $\mathcal{I}_2(A, B)$ to vary with events, but require a constant value for $\mathcal{I}_3(A, B, C)$ (which depends on $p(A \cup B \cup C)$). To simplify presentation, when $p(A \cup B \cup C) = 1$, we will invariably refer to $\mathcal{I}_3(A, B, C)$ as $\Delta(A, B, C)$.

This design is analogous to slit experiments in physics. Having slits A and B open means we are computing $p(A \cup B)$ and different common causes are analogous to different measurement positions on the screen, for which quantum theory makes varying predictions for constructive vs. destructive interference (Feynman et al., 2005). The condition $p(A \cup B \cup C) = 1$ is that a particle has to go through one of the slits. We can open or close slits so that different experimental runs correspond to different two-way disjunctions, allowing us to compute $\mathcal{I}_2$. In physics, general linear models are readily falsified, because $\mathcal{I}_2$ depends on which two slits are open, that is, the interference patterns are not the same across slit pairs. However, there is no evidence against level 2 theories: $\mathcal{I}_2(X, Y)$ for different pairs of slits apparently combine in such a way that their sum exactly balances the other probabilities (as long as the three-way disjunction is fixed), across different experimental setups (Sinha et al., 2010). It would be a curious and surprising finding if the same can be said for psychological processes.

A general null hypothesis is that neither linear nor bilinear variants can capture human behavior, that is, three-way (and two way) interference varies across events. For example, it may be that the whole approach of trying to understand inference with quantitative approaches is wrong, which might be the case if the mind has no way of representing probabilities, even if approximately so (Shiffrin, 2021). Even though two-way interference effects from quantum theory have generally been offered as explanations for some observed fallacies (as in Busemeyer et al., 2011), there have been few demonstrations that these interference effects are consistent with quantum theory (an exception is Yearsley & Trueblood, 2018). Therefore, the present approach offers a major test of the psychological relevance of quantum theory.

Table 3.1. Interference terms for level 2 and level 3 probability models, under different assumptions for response bias (or noise).

Model	$\mathcal{I}_2(A,B)$	$\mathcal{I}_3(A,B,C)$; when $p(A \cup B \cup C) = 1$, $\mathcal{I}_3(A,B,C) \equiv \Delta(A,B,C)$	Comments
BT	0	0	Special case of below.
BT, simple response bias	$-d$ Constant within, between scenarios.	$2d$ Constant within, between scenarios.	Special case of below.
BT, elaborate response bias	$-d-\delta[2p(A \cup B)-1] < 0$ Varies within, between scenarios.	$2d+\delta>0$ Constant between scenarios	
QT	$I_2(A,B) \neq 0$ Varies within, between scenarios.	0	Special case of below.
QT, simple response bias	$(1-2d)I_2(A,B)-d \neq 0$ Varies within, between scenarios.	$2d > 0$ Constant between scenarios.	

NB. BT=Bayesian theory, QT=Quantum theory. 'Within a scenario' refers to whether an interference term is expected to vary as the event pair varies for the same scenario, e.g., is $\mathcal{I}_2(A,B) \neq \mathcal{I}_2(B,C)$? Between scenarios refers to whether an interference term is expected to vary as we switch from one triplet of events {A,B,C} to another {X,Y,Z}.

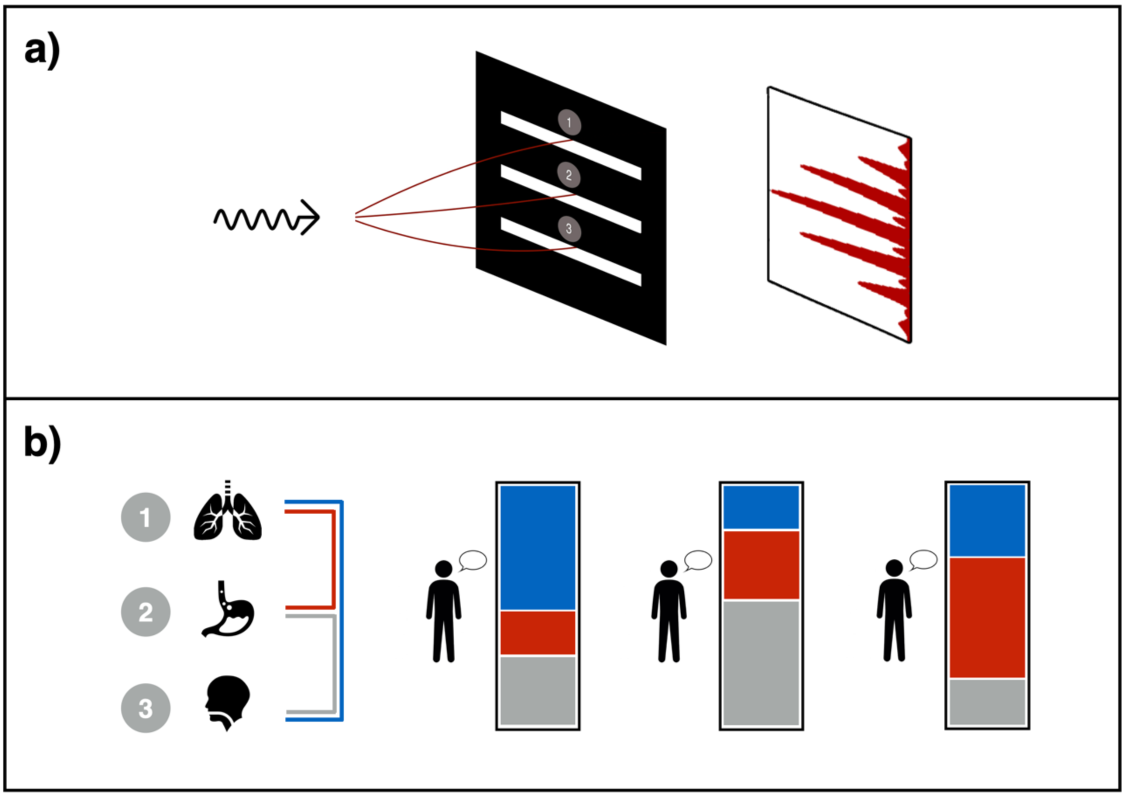

Figure 3.1. (a) and (b): The analogy between three-slit interference experiments in physics and the present experimental paradigm. In both cases, the surprising prediction is that pairwise interference terms ‘conspire’ with other probabilities, to produce a constant (set by the three-way disjunction), regardless of the events *A*, *B*, *C* in question.

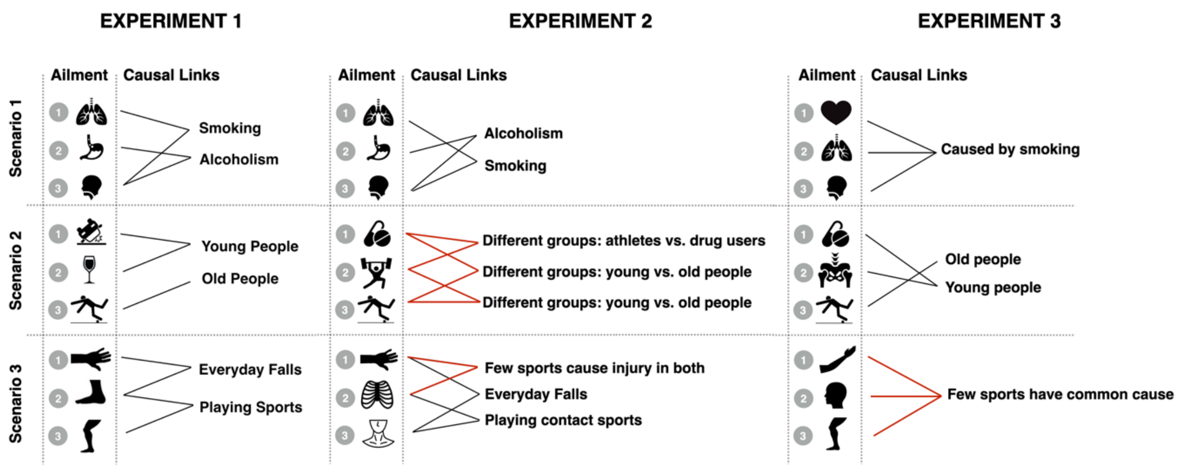

Figure 3.2. Ailments and causal relations (positive and negative links shown in black and red, respectively) in the three experiments. The ailments in Experiment 1 were lung, stomach and

throat cancer (Scenario 1), auto accidents, alcohol poisoning, and falls (Scenario 2), and fractures to wrists, ankles, or lower legs (Scenario 3). In Experiment 2, these were lung, stomach and throat cancer (Scenario 1), drug overdose, sports injury, and falls (Scenario 2), and wrist, rib, and collar fractures (Scenario 3). In Experiment 3, these were heart, lung, and throat cancer (Scenario 1), drug overdose, hip injury, and falls (Scenario 2), arm, head, and lower leg injury (Scenario 3).

To recap the hypotheses then, the test that linear models (such as CPT or CPT with noise) are inadequate is that $I_2(A, B)$ varies with scenarios. It may seem that as causal relations vary, so should $I_2(A, B)$, but recall this key point. In general, causal relations are naturally incorporated in CPT, e.g., with the use of Bayesian Networks. The emergence of variable two-way interference terms could only occur from causal relations strong enough to break classical bounds. The test that bilinear models (such as QPT or QPT with noise) are inadequate is that $I_3(A, B, C)$ varies with scenario. Again, it might seem that since $Prob(A \sqcup B \sqcup C) = 1$, $I_3(A, B, C)$ should be likewise fixed; or that the presented causal relations were not memorable enough or strong enough for variable three-way interference. The condition I test is that the sum of individual probabilities and two-way disjoint disjunctions sum to a constant across scenarios, despite variations in the events and their causal relations. This is a surprising, stringent constraint. Essentially, both CPT and QPT (and the corresponding more general classes of probability models) require probabilities to combine in a certain way – otherwise, we leave the realm of formal probabilistic inference and are forced to entertain hypotheses for decision making based on heuristics (or higher order probabilistic frameworks). Note, I adopt Bayesian statistics throughout, since in part we are testing that certain probability terms are not different from zero.

3.1.1 Method

Participants

I recruited 400 participants, equally split into two between participants counterbalancing conditions (referred to as just ‘conditions’ below), which just varied the scenarios I employed. Recruitment was through Amazon Mechanical Turk, restricting geographical location to North America. Participants required approximately 20 minutes to complete the task and they were compensated $1 for their time. The sample size for this experiment was, also, exploratory; there is no prior related work at all.

Design and procedure

Each of the two conditions involved three scenarios and each scenario involved three diseases. Each scenario described a hospital ward in a fictional town, specializing in a particular type of ailment. For example, for Scenario 1, participants were told of a cancer ward, treating only patients of three types, those with lung cancer, stomach cancer, or throat cancer; for Scenario 2, the three ailments were auto accidents, alcohol poisoning, and falls; for Scenario 3, they were fractures to wrists, ankles, or lower legs.

In a training phase, participants went through each scenario in a blocked format presentation, so that, for example, no information about a subsequent scenario would be presented prior to finishing all questions relevant to the current scenario (scenario order was randomized). The block for each scenario had analogous format. Participants were first reminded about the information about the hospital ward, the ailments treated there, and the causal relations. Subsequently participants went through four or five multiple choice questions testing knowledge of the causal relations. The questions were meant to be straightforward and answerable on the basis of a simple understanding of the presented information. Participants received corrective feedback, specifically if they responded incorrectly, they were told so and asked to try again until they answered correctly (there were more than two alternatives for each question).

In the test phase, once the training part was over, participants were told that they would be asked to make judgments about the proportion of various categories of patients at the fictional hospital. With each question, the text describing the hospital ward and the causal dependencies was included so that participants did not have to memorize anything, just understand the information provided. Participants first saw the three questions from each scenario corresponding to the three-way disjunction. The three-way disjunction questions corresponded to catch questions, since the total number of patients was fixed at 100. Without these three questions, there were six questions per set, for a total of 18 questions. Participants saw these questions in a randomized order, but such that for each set of three consecutive questions there was one question from each scenario (the idea was to reduce response biases which might arise from participants overtly attempting to be consistent in their answers for questions within the same scenario). Each of the questions was prompted with the statement that each patient was brought to the hospital ward for only a single type of ailment (e.g., a single cancer type or a single fracture, depending on the scenario). Then, participants were asked to indicate on a 0 (None of them), to 100 (All of them) slider the proportion of patients likely to be admitted for ailment A in some questions, A or B in other questions, and A or B or C in another question; note, each combination of possibilities was shown only once. Finally, an additional three catch questions were included, where participants were just told to select a particular response, as a check that they were paying attention.

3.1.1 Results

I consider whether two-way interference terms, $I_2(A,B)$, vary between pair of diseases within the same scenario and between scenarios. Recall, the CPT, simple response bias model assumes no effect of both pair and scenario on $I_2(A,B)$. The CPT, elaborate response bias model assumes an effect of pair and scenario, but that the value of $I_2(A,B)$ will be consistently negative. Regarding $\Delta(A,B,C)$, both CPT models require no effect of either pair or scenario. For the QPT models, in both versions I predict an effect of pair (with inconsistent sign for $I_2(A,B)$) and scenario. Importantly, both models also predict no effect of pair or scenario for $\Delta(A,B,C)$.

I report Bayes factors for inclusion ($BF_{Inclusion}$) for the alternative versus the null hypothesis, so that values greater than 1 indicate evidence for the alternative hypothesis, that is, that the statistical model with the corresponding term is better compared to the null (intercept only) model. The statistical tests were Bayesian RMANOVAs (repeated measures multivariate ANOVAs), with $I_2(A,B)$ and $\Delta(A,B,C)$ as dependent variables and pair, scenario, and condition as independent variables (the latter variable is a counterbalancing variable).

Table 3.2. Analysis of effects for Bayesian Repeated Measures ANOVA of two-way interference terms.

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
Pair	0.737	1.000	∞
Scenario	0.737	1.000	22291.949
Condition	0.737	0.998	233.410
Pair ✻ Scenario	0.316	1.000	19960.810
Pair ✻ Condition	0.316	0.998	1277.105
Scenario ✻ Condition	0.316	0.998	1290.978
Pair ✻ Scenario ✻ Condition	0.053	0.998	10584.689

Concerning the two-way interaction terms, in Table 3.2, the high $BF_{Inclusion}$ for the pair independent variable disconfirms the CPT, simple response bias model. The large Bayes factors for both independent variables of pair and scenario, as well as the interaction terms, are unsurprising given the experimental design - different combinations of ailments in different scenarios were designed to elicit different expectations regarding combined probabilities. Note, in these tests the alternative hypothesis was that the mean is greater than 0. Therefore, very high Bayes Factors (BFs) indicate means for which we have a lot of confidence that they are positive and very low BFs indicate high confidence that the means are negative.

I next consider the three-way interaction term, $\Delta(A, B, C)$, see Table 3.3. Note, the pair variable that concerns the particular ailments paired together – while this is clearly relevant when discussing $I_2(A, B)$ (Table 3.2), it does not apply for $\Delta(A, B, C)$ (Table 3.3), as there is only one combination of ailments.

Table 3.3. Analysis of effects for Bayesian Repeated Measures ANOVA of three-way interference terms.

Effects	P(incl)	P(incl|data)	$BF_{inclusion}$
Scenario	0.600	0.482	0.620
Condition	0.600	0.538	0.775
Scenario ✻ Condition	0.200	0.472	3.581

The results are striking; none of the main effects have Bayes factors for inclusion greater than 1, indicating that no model containing any combination of these effects can be preferred over a null model. I therefore conclude that $\Delta(A, B, C)$ terms do not vary when I manipulate scenario, indicating that the QPT, simple response bias model is sufficient to account for participants' behaviour.

The main effect of CRT (Cognitive Reflection Test) and its interactions were found to be consistently non-significant across all scenarios and conditions, and are subsequently removed from the reported results. The lack of a significant effect of the CRT is actually reassuring; recall that the terms I3(A,B,C) should be constant regardless of whether a decision maker is using a linear (classical) or bilinear (quantum) model. The CRT has previously been associated with the strength of various measures of non-normativity (Yearsley et al, 2015) and the fact that it is not predictive here suggests that these effects behave very differently from other measures such as the size of conjunction fallacies.

Table 3.4. The results regarding all two-way interference terms, for the different scenarios. S1, S2 etc. indicate the scenario and int12 etc. indicate the particular pair of ailments for each scenario (there were three kinds for each scenario). Recall that I employed three scenarios and that the causal relations between ailments varied within each scenario.

Interference Term	BF_{10}	Mean
S1 int12	41539.25	-6.50
S1 int23	18048.69	7.15
S1 int13	1.639e +18	14.70
S2 int12	1.932e +11	-9.75
S2 int23	1.599e +8	8.70
S2 int13	2.174e +12	11.27
S3 int12	4.516e +8	-10.08
S3 int23	2.666e +6	8.67
S3 int13	23.59	4.40

Note. For all tests, the alternative hypothesis specifies that the mean is greater than 0.

Recall that the CPT, elaborate response bias model assumes an effect of pair and scenario, but that the value of $I_2(A, B)$ will be consistently negative. Further analysis of the value of interference terms across the event pairs in each scenario are not consistently negative, providing further support for the QPT, simple response bias model being sufficient to account for participants' behaviour. Refer to Table 3.4 above.

3.1.3 Discussion

The present results show that an effect of interference can be introduced between scenario pairings that produce significant differences in behaviour. These results additionally show the limitations of CPT in predicting certain forms of human behaviours that persist even in extensions of the approach. Overall, the results support the quantum account of participants' behaviour over the classical account and its extensions. Nonetheless, a key limitation is whether having only three events correctly addresses the issue of how the estimation of the three-way disjunction constrains the other probabilities (two-way disjunctions and marginals). This is because in the current experiments, three-way disjunctions are set by the paradigm. Including an additional condition, where the three-way disjunction is estimated directly by participants, would require four separate events. While this might be a plausible avenue for future work, I decided against it for the present time, because of the increasing complexity of the materials (with four events, the paradigm/ scenarios might be too complex for participants to adequately follow).

Another interesting result was the finding that interference terms were not consistently negative between scenario pairs. Although these results do support the quantum account of behaviour, they are not completely surprising given the form of interference introduced between scenario pairings. Present scenario pairings are linked through a common cause and this commonality functioned as interference for the various pairings. From a quantum perspective, interference can be both positive and negative. However, just as an effect of interference can exist between scenario pairings due to some commonality between the pairs, and effect of interference should also exist between scenario pairings due to a distinct lack of commonality between pairs. I assume that a commonality or a distinct lack of it between scenario pairings can function as positive and negative interference respectively.

As the present experiment possibly only induced a positive effect of interference, it is undetermined if a similar pattern of interference would appear if both positive and negative interference were introduced into the paradigm. That is, the introduction of negative interference may yield consistently negative interference terms congruent with the CPT account. Therefore, the aim of the second experiment in this chapter is to introduce negative interference into the experimental paradigm to assess its effects on evaluating the CPT and QPT frameworks.

Section 3.2 Evaluating Rationality with Positive and Negative Interference

Introduction

In Experiment one, the interference between pairs of events in each scenario was operationalised through a common cause between pairs of events. As mentioned in the introduction, the definition of interference can be either positive or negative. In the first experiment, our operationalisation of interference between pairs of events focused on a common cause between such pairs, which functioned as positive interference.

I define positive interference between pairs of events as a link between pairs of events derived from a common cause. For example, consider a scenario in Experiment one where the accident and emergency ward in the hospital of fictional Eastville treats patients from auto accidents (*A*), alcohol poisoning (*B*), and falls (*C*). Events A and B are linked by providing participants with a description of a common cause, such that participants were told that *A* and *B* typically involve young people. However, as previously mentioned, interference can be either positive or negative. Given that the first experiment only used positive interference, I was unable to determine if the results are dependent specifically on positive interference, or if they vary across positive and negative interference. As such, for completeness a second experiment was conducted to assess if negative interference between pair of events in each scenario yielded results consistent with the first experiment. Therefore, for this experiment, the hypothesis is that the introduction of negative causal links will not alter the conclusions from Experiment 1.

3.2.1 Method

Participants

I recruited 300 participants, equally split into two between participants counterbalancing conditions (referred to as just ‘conditions’ below), which just varied the scenarios I employed. Recruitment was through Amazon Mechanical Turk, restricting geographical location to North America. Participants required approximately 20 minutes to complete the task and they were compensated $1 for their time. The sample size for this experiment was, also, exploratory; there is no prior related work at all.

Design and procedure

The experimental design was identical to the first, expect pairs of events were linked through a lack of a common cause. I define negative interference between pairs of events as a link between pairs of events derived from there being no common cause between those events. For example, consider a scenario in Experiment one where the accident and emergency ward in the hospital of the fictional town of Southville treats patients with wrist factures (A), rib fractures (B), and collar bone fractures (C). Events A and B are linked by providing participants with a description of a lack of common cause between these events, such that participants were told that there are few activities which can cause injuries in A and B. In some other cases, the lack of a common cause was presented for both pairwise sets, (A, B) and (B, C), but again was not expanded to for the all-inclusive set, (A, B, C). The rest of the experimental structure remained identical to the first experiment.

3.2.2 Results

An identical set of analyses to Experiment one are reported below, where the aim is to determine if two-way, $I_2(A, B)$, and three-way, $I_3(A, B, C)$, interference terms vary between pair of diseases within the same scenario and between scenarios.

If the best description of this situation is via a linear model, i.e. if non-normative effects are either absent, or due only to response error, then I expect to see no effect of scenario, condition or pair for two-way interference. However, if there exists more elaborate noise, then I expect to see an effect of scenario and pair, with all two-way interference terms being negative. Additionally, I expect to see no effects for three-way interference. For bilinear models, I would expect to see an effect of scenario and pair in two-way interference, but not in three-way interference. Contrary to results from experiment 1, although large Bayes factors are still found for scenario and condition, a very large effect of pair is present. Refer to table 3.5 below.

Table 3.5. Analysis of effects for Bayesian Repeated Measures ANOVA of two-way interference terms.

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
Scenario	0.737	1.000	3.165e +11
Pair	0.737	1.000	23173.991
Condition	0.737	1.000	6.087e +8
Scenario ✻ Pair	0.316	1.000	80991.409
Scenario ✻ Condition	0.316	1.000	2.937e +9
Pair ✻ Condition	0.316	0.972	73.894
Scenario ✻ Pair ✻ Condition	0.053	0.971	597.261

Results for two-way interaction terms yield high $BF_{Inclusion}$ for the main effects of scenario, pair and condition, which provides evidence against the CPT, simple response bias model. The large Bayes factors for the interaction terms are again unsurprising given the experimental design, where different combinations of ailments were designed to elicit different expectations regarding combined probabilities. I next consider the three-way interaction term, $\Delta(A, B, C)$, see Table 3.6 below.

Table 3.6. Analysis of effects for Bayesian Repeated Measures ANOVA of three-way interference terms.

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
Scenario	0.600	0.179	0.145
Condition	0.600	0.278	0.257
Condition ✻ Scenario	0.200	0.162	0.775

Similar to the first experiment, none of the Bayes factors for inclusion are greater than 1, indicating that no model containing any combination of these effects can be preferred over a null model. I therefore again conclude that $\Delta(A, B, C)$ terms do not vary when I manipulate scenario, indicating that the QPT, simple response bias model is sufficient to account for participants' behaviour, as found in the results of the results of the first experiment.

For these results there is strong evidence for the effect of CRT, indicating that no model containing any combination of these effects is preferred over a null model. The conclusion then is that the terms I3(A,B,C) are not constant. The terms do vary when I manipulate common causes implied for the events. This does not imply that a bilinear model is sufficient to explain these effects.

Table 3.7. The results regarding all two-way interference terms, for the different scenarios. S1, S2 etc. indicate the scenario and int12 etc. indicate the particular pair of ailments for each scenario (there were three kinds for each scenario).

Interference Term	BF_{10}	Mean
s1_12	9.556e +18	-18.41
s1_23	3.148e +20	-19.93
s1_13	1.357e +35	-25.76
s2_12	3.400e +28	-21.86
s2_23	5.356e +28	-22.64
s2_13	7.242e +23	-20.44
s3_12	2.447e +22	-20.70
s3_23	3.231e +15	-16.08
s3_13	1.544e +23	-20.49

Note. For all tests, the alternativities that the mean is greater than 0.

Recall that the CPT, elaborate response bias model assumes an effect of pair and scenario, but that the value of $I_2(A,B)$ will be consistently negative. A further analysis of the value of interference terms across the event pairs in each scenario are consistently negative. Refer to Table 3.7 above.

3.2.3 Discussion

The results of the present experiment are largely in line with Experiment 1. Two and three-way interference effects were observed as predicted by the QTP framework and not the CPT framework. However, contrary to Experiment 1, the results for the interference terms found that they were consistently negative and therefore in support of the CPT extension model. As the present experiment sought to introduce both positive and negative interference into the paradigm, these results may imply that they are conditional on the specific introduction of negative interference.

I believe these findings ultimately support the QPT account. This is because the CPT elaborate plus noise model predicts that observed interference terms should be consistently negative. However, it cannot be assumed that these findings are dependent on the specific introduction of negative interference into the paradigm. As Experiment 1 showed, if it is assumed that an effect of interference can be isolated as either positive or negative, the results in the first experiment revealed that inducing solely positive interference can nonetheless yield negative interference terms. In which case, it holds that across both experiments interference terms should be consistently negative, if they are to support the CPT account. This was not the result. On the other hand, predictions of two and three-way interference remained consistent throughout both experiments and again supported the QPT account of participants' behaviour.

The differences observed in the interference terms between Experiment 1 and 2 are hard to reconcile with the consistency in two and three-way interference found between experiments. As if the introduction of negative interference is assumed to have an effect on the overall behaviour of participants, just as positive interference had, it is hard to understand why this effect would only be observed in the interference terms and not in the two and three-way interference results. It is presently difficult to determine, however, whether trying to manipulate the type of interference introduced into the experiment resulted in undefined and unwanted experimental manipulations instead.

Nonetheless, it remains to be determined if differences in the type of interference introduced into the paradigm have an effect on three-way interference specifically. For completeness, a third experiment is proposed. In this experiment, both positive and negative three-way interference will be introduced to assess their effects on evaluating CPT and QPT frameworks.

Section 3.3 Evaluating Rationality with Three-Way Interference

Introduction

In the first and second experiments I tested whether there were variations in two-way and three-way interference terms across pairs of events, scenarios and conditions. I found that for both positive and negative interference, two-way interference terms do not remain constant across event pair and scenario, however, three-way interference does. Therefore, the present results are in support of the QPT, simple response bias model's account for participants' behaviour.

However, I have not explicitly tested whether introducing a common cause, or a lack of one, for the all-inclusive set (A,B,C) for the three presented events in each scenario still yields results consist with the first and second experiment. That is, I have yet to empirically determine whether direct variations in the links between three events in a scenario can cause three-way interference effects. Presently, the results from the first two experiments do not imply this. However, this manipulation has not been empirically tested and as so is the purpose of the third experiment. For the final experiment of this chapter, the hypothesis is that introducing a common cause, or a lack of one, for the all-inclusive set of events (A,B,C) in each scenario still yields results consistent with the first and second experiment.

3.3.1 Method

Participants

I recruited 300 participants, equally split into two between participants counterbalancing conditions, which just varied the scenarios I employed. Recruitment was through Prolific, restricting geographical location to the United Kingdom. Participants required approximately 15 minutes to complete the task and they were compensated £2 for their time. Following from previous experiments, the sample size for this experiment was exploratory.

Design and procedure

The experimental design was largely similar in structure to the first and second experiment. As the focus of this experiment was to assess primarily three-way interference, there was no need to account for the differences that arise in the presentation of joint causes or lack thereof in this experiment. Therefore, I reduced the number of blocks to two.

Each participant was presented with three scenarios in each block, containing three events. One scenario linked all three events to a common cause (positive interference). For example, consider a scenario in Experiment one where the accident and emergency ward in the hospital of the fictional town of Northvile treats patients with lung cancer (A), stomach cancer (B), and throat cancer (C). Events A, B and C are linked by providing participants with a description of a common cause between these events, such that participants were told that the deposit of tar from smoking can cause A, B and C (positive interference). A similar scenario was presented for another fictional hospital treating patients with different conditions, except that a lack of a common cause was stated among the three events (negative interference). Each block also contained one scenario from the first or second experiment, testing positive and negative two-way interference respectively. The rest of the experimental structure remained identical to the first and second experiment.

3.3.2 Results

An identical set of analyses to Experiment one are reported below, where the aim is to determine if two-way, $I_2(A,B)$, and three-way, $I_3(A,B,C)$, interference terms vary between pair of diseases within the same scenario and between scenarios. CRT results were removed from the final analyses presented below because they did not yield any significant results. Largely constant with the results from Experiment one and two, there are strong effects of scenario, pair and this time condition.

Table 3.8. Analysis of effects for Bayesian Repeated Measures ANOVA of two-way interference terms.

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
SCENARIO	0.737	1.000	∞
Pair	0.737	1.000	∞
Condition	0.737	1.000	∞
SCENARIO ∗ Pair	0.316	1.000	4.539e +14
SCENARIO ∗ Condition	0.316	1.000	4.539e +14
Pair ∗ Condition	0.316	1.000	∞
SCENARIO ∗ Pair ∗ Condition	0.053	1.000	2.533e +15

Referring to Table 3.8, results for two-way interaction terms yield high $BF_{Inclusion}$ for the main effects of scenario, pair and condition, largely similar to the results found in the first and second experiment. Thereby providing evidence against the CPT, simple response bias model. The large Bayes factors for the interaction terms are again unsurprising given the experimental design, where different combinations of ailments in different scenarios were designed to elicit different expectations regarding combined probabilities. I now consider the three-way interaction term, $\Delta(A, B, C)$, see Table 3.9 below.

Table 3.9. Analysis of effects for Bayesian Repeated Measures ANOVA of three-way interference terms.

Effects	P(incl)	P(incl|data)	BF $_{inclusion}$
SCENARIO	0.600	0.240	0.211
Condition	0.600	0.177	0.143
SCENARIO ✻ Condition	0.200	0.009	0.036

Results show that again, none of the Bayes factors for inclusion are greater than 1, showing that no model containing any combination of these effects can be preferred over a null model. I can therefore reaffirm the conclusion that $\Delta(A, B, C)$ terms do not vary when I directly manipulate scenario, again indicating that the QPT, simple response bias model is sufficient to account for participants' behaviour, as found in the results of the first and second experiment.

Table 3.10. The results regarding all two-way interference terms, for the different scenarios. S1, S2 etc. indicate the scenario and int12 etc. indicate the particular pair of ailments for each scenario (there were three kinds for each scenario).

Interference Term	BF_{10}	Mean
s1_12	563217.674	8.445
s1_23	69388.408	10.170
s1_13	4.007	5.417
s2_12	443.283	5.138
s2_23	1688.476	6.951
s2_13	177691.477	8.741
s3_12	3.980	3.457
s3_23	0.176	1.891
s3_13	2.484	3.745

As found in Experiment one, analyses of the two-way interference terms provide evidence against the CPT, elaborate response bias model. As this linear model assumes that the presence of two-way interference terms are constantly negative. Refer to Table 3.10.

3.3.3 Discussion

I created a generic probabilistic estimation task, where participants had to make probability estimates for simple events, two-way disjunctions or three-way disjunctions. Present findings of two-way disjunctions are not surprising. However, the finding that the two-way interference terms varied with scenario is a novel empirical finding, as is the lack of a three-way interference term. The former is important because it precludes a particular CPT plus response bias account of (CPT) probabilistic fallacies. According to Costello and Watts' (2014) influential ideas, errors in probabilistic judgments (which could lead to fallacies) depend on probabilities and probabilities were fixed across all scenarios. Additionally, memory or simulation errors are meant to depend on experience or memory recency of the corresponding events. Even though I cannot preclude some participants being more familiar with some ailments than others, across the large sample I employed I would expect such individual differences to average out. So, the current specification of Costello and Watts' (2014) model is hard to reconcile with variability in the interference terms across scenarios, though of course a more elaborate version of this model might be able to capture the necessary dependencies. Their model is also inconsistent with the lack of evidence for a three-way interference term, which is predicted from the observation of a two-way one. The lack of a three-way interference term provides support for QPT as the appropriate framework for describing human probabilistic judgments.

The empirical results for so-called probabilistic fallacies especially in decision-making has led to an intense debate of how much (if at all) of human cognition should be understood in terms of the principles of CPT (Tversky & Kahneman, 1974, 1983). The advent of QPT cognitive models raised the possibility that all (or most) of human cognition could be understood in formal probabilistic terms, but the appropriate approach is not CPT, but instead QPT (Pothos & Busemeyer, 2013). This work questions the focus on just CPT and QPT and illustrates how there is a whole hierarchy of probabilistic frameworks, which differ in the dimensionality of representational space, the ease of specifying states, and (importantly) the complexity of interference terms.

The function of interference is crucial and generalizes the idea of incompatibility, which is central in QPT cognitive (and otherwise) models. Two-way incompatibility, as embodied in QPT, means that for pairs of incompatible questions, asking one question provides a context or perspective which alters the meaning of the subsequent question and subsequently functions as an order effect. In higher level probability theories, interference as understood from a QPT perspective is defined as a contextual effect, which generalises to any given number of presented events. It is worth noting that so far there have been extremely few systematic attempts to explore what higher order probability theories would even look like (Hardy, 2002). Another key point is that this discussion need not be restricted to CPT vs. QPT, but rather concerns any probability theory where the probability measure is in a simpler linear form (such as CPT) vs. any probability theory where the measure is a more elaborate bi-linear form (such as QPT). As such, I can test the plausibility of simple vs. more elaborate probability frameworks in cognition, by looking for the presence of simple vs. more elaborate interference terms. Both probability frameworks make testable predictions about the presence and interaction of interference terms. These predictions thereby constrain the probability framework and function as a method for testing the plausibility of various probability frameworks in underlying human cognition. Thus, the presence of two-way interference terms found in all of our experiments is hard to reconcile with CPT. Similarly, a three-way interference term cannot be reconciled with QPT, unless one postulates a response bias term of a particular kind. An important aspect of the manipulation was that the probabilities for the simple events and the two-way disjunctions were intended to be fixed across the different scenarios that I employed.

An important question is how much I would expect the present demonstration to be a foundation for general conclusions regarding linear (such as CPT) vs. bilinear (such as QPT) probability frameworks. Clearly, I employed only a single decision-making task and so the empirical approach is best seen as one seeking an existence proof that higher order interference terms may be present in probabilistic reasoning. I easily replicated the finding regarding two-way interference terms and added the result that such interference terms can vary even though the corresponding objective probabilities (for both single events and disjunctions) were constant. Could the lack of evidence for three-way interference just be taken as indicating that our paradigm was not sufficiently sensitive? While I cannot preclude this possibility, I think it is unlikely. The experimental design involved all the implied causal links which would be thought as conducive to higher order interference terms. However, all three experiments clearly showed very little support for its presence. Ultimately, I think the three-way incompatibility required for three-way interference terms is implausible from a cognitive perspective, on the basis of the complexity of nuanced contextual effects which will be required. As such, the results can in one way be seen as the first attempt at simultaneously ruling out linear and trilinear or higher probability frameworks in human cognition.

As previously stated, not only will capturing the nuanced contextual effects to facilitate three-way interference be difficult, but I believe it may function as a synthetic overcomplication. There is no reason to believe that human cognition is underpinned by higher order probability frameworks beyond bilinear ones. Trilinear and ever more complex higher order probability frameworks largely represent overcomplex and unrealistic probability structures. What the present results may be indicating through the lack of three-way interference, is a breakdown in the cognitive support for ever more complex probability structures. The presence of two-way and three-way interference may be indicative of a constrain on the ability for human cognition to correctly understand higher order probability structures, in addition to constraining linear or bilinear probability models. This is not surprising, given the literature on human decision-making errors in numerous contexts. The CF is a perfect example of this. The current literature does not support the idea that human cognition processes events beyond a pairwise fashion, as suggested by the bilinear (QPT) models, to support more complex probability structures.

Nonetheless, at least one objection to using quantum probability theory (there are many) is that it is unclear how exactly this expands the space of possible models. Most accounts of the relationship between quantum and classical models tend to focus on the issue of incompatibility, but this is notoriously hard to make precise. In addition, it is far from clear that quantum probability theory is the only way to generalise classical probability to include incompatible events.

In closing, I hope that this work provided a novel perspective on the debate for the probabilistic principles possibly relevant in cognition, generalizing both the CPT vs. CPT plus response bias vs. QPT debate (by recognizing the essential difference between CPT vs. QPT as involving linear vs. bi-linear probability measures) and placing these theories in a broad hierarchy, of several additional entries, such that entries higher up would involve more complex forms of incompatibility.

Chapter Four: Quantum Constraints of EAMs

Section 4.1 Quantum Characteristics of EAMs

Introduction

From the research in chapter two I have created a model for the CF, or more broadly for interference effects in the psychophysical domain, where the evidence accumulation process for one stimulus is directly impacted by the evaluation of a previous stimulus. This is the S-BSP model. This model could also be extended to a number of different effects, but an obvious one is to look at two consecutive evaluations of the same stimuli, which has a similar serial processing structure to the S-BSP model. That is, the effect of an initial judgment on a later one, in the same manner as Kvam, Pleskac, Yu, and Busemeyer (2015) did. Intuitively, the S-BSP model should be able to capture these non-normative effects in different serial processes, as they function within the same serial processing architecture. This can thereby help in determining if these other effects also possibly originate from an explicit effect of interference between evidence accumulation processes. As such, this chapter will explore the origin of quantum interference effects, by focusing on the interference effect in the choice-confidence judgement task by Kvam et al. (2015).

The S-BSP model says that interference happens at the level of evidence accumulation and it also has relatively standard model features, with an internal stopping rule (participants respond at their own discretion) and a binary choice. In contrast, Kvam et al. (2015) had a different experimental set up with an initial binary choice with an external stopping rule (participants responded when prompted) and a final 100-point scale confidence choice. Therefore, an important question, not just for the application of the S-BSP model but also more generally, is where exactly the interference effect in this task arises from. If it comes from the basic evidence accumulation process it ought to be visible in a simpler set up with only binary choices and an internal stopping rule. If it somehow relies on an external stopping rule it will only show up then. Furthermore, if it only shows up when we have confidence judgments it is not obvious that it arises from the evidence accumulation process at all.

The aim of this chapter was to look at a putative interference effect in a psychophysical domain and to try break down the relevant experimental paradigm, in a way that makes it is easier to identify where the effect comes from. In particular, the key question is whether the effect can be said to be based on the evidence accumulation process, as identified by the S-BSP model, or whether it arises from other processes, perhaps relating to the confidence judgments. In the latter case, we would say that the apparent interference effect might be the result of a more complex EAM-like process, that is not captured by standard EAMs. My overall belief is that this is the best way to approach the interpretation of these results, questioning the conclusions from Kvam et al. (2015). It should be clear that more work is needed to fully resolve these issues. Nevertheless, my analyses below show the kind of care that is needed before conclusions, such as those from Kvam et al. (2015), are drawn.

In the original experiment, 9 participants each completed 2,688 experimental trials of a random dot motion task, where they judged whether a cloud of white dots (within a circle on a black background) were moving towards either the left or the right of the screen. Each trial had either 2%, 4%, 8%, or 16% of the dots coherently moving in one of the directions and all others moving randomly. After 500ms of each trial, participants were prompted by a beep for a response: in half of the 24 blocks participants had to make a binary choice of the direction that the dots were moving (intermediate choice trials), and in the other half participants were required to only acknowledge the beep by clicking the mouse (standard no choice trials). Either 50ms, 750ms, or 1,500ms after the initial response, participants were prompted for a confidence judgment, ranging from 0 (certain left) to 100 (certain right). Their findings showed that on trials where participants had to make an intermediate decision, participants subsequently reported less confidence then when there was no intermediate judgement.

Kvam et al. (2015) claim that the results provide strong evidence against classical models within a purely classical framework. Instead, they argue that the findings show that judgements and decisions create, rather than reveal preferences and beliefs. The researchers state that the standard EAM process account of decision-making does not allow for the interference result found due to the “read-out” assumption: essentially, that reporting a belief or preference does not change the associated mental state. However, within the framework of QPT judgments and decisions are treated as a measurement process that constructs a definite state from an indefinite state (Pothos & Busemeyer, 2013, 2022). When a decision is made, the indefinite state collapses onto a set of evidence levels that correspond to the observed choice, producing a definite choice state. Choice-confidence judgements work in a similar way. An initial intermediate choice acts as the indefinite state that collapses onto a more specific set of evidence levels corresponding to a later confidence judgement. This process of changing states represents a change in the mental state of the decision maker and it is through this change in states that interference effects are assumed to occur and lead to the observed results. As a result of this assumed interference, the QPT framework allows for a situation where participants can make an intermediate choice judgement followed by a final confidence judgement and results differ from an identical situation where no intermediate choice judgement is made.

For example, take the instance where a decision maker has to make a choice between deciding if statement A or B is true and then later has to rate their confidence that the chosen statement is true. According to the read-out assumption by Kvam et al. (2015), a decision is made on the basis of existing evidence for either choice. After the choice decision is made, evidence accumulation continues and the later confidence judgement is made by determining existing evidence for either choice again. It is important to note that the choice-confidence task presented by Kvam et al. (2015), assumes that evidence accumulation continues after the initial choice decision, because the subsequent confidence judgement is assumed to derive from the information processing that occurred for the choice decision. However, this assumption is not elaborated on in the paper. Nonetheless, this process does not change the system state (associated mental state) or create a new one. From the perspective of CPT, the resulting distribution of confidence ratings should therefore be identical to conditions in which the person makes no choice at all, as both states are unconnected. However, from a QPT perspective, when a choice is made the indefinite (superposition) state collapses onto a set of states each represented by a different level of accumulated evidence and creates a definite state. This change in system state from indefinite to definite creates an entirely new state. Consequently, after the initial choice decision a new state is created, different from the one if no initial choice was made. This difference in system states subsequently results in different confidence judgements being made. As such, it is the creation of this new system state that produces an interference effect and consequently, different confidence judgements.

From the perspective of an EAM account, the task by Kvam et al. (2015) has two features that are not captured in standard models. Firstly, participants were prompted to make decisions at specific points in time. The standard EAM framework assumes that participants respond at their own discretion, or in other words, when the amount of accumulated evidence for either binary response reaches the response threshold for one of the binary responses and not before. This discretion represents an internal stopping rule, whereas the task by Kvam et al. (2015) represents an external stopping rule. Secondly, in the condition where participants were required to make an intermediate binary judgement before a final one, the trial does not end. Instead from this point through to the end of the trial, the stimulus remains on screen. This stop-start of evidence accumulation that occurs during and directly after the intermediate judgement is also not represented in standard EAMs. From this perspective, the process of accumulating evidence occurs for only one judgement at a time. One process of evidence accumulation is responsible for one judgement and does not extend to multiple judgements. This is different from the case of multiple discretionary responses in a single trial.

If an equal comparison is to be made between standard EAMs and the quantum EAM proposed by Kvam et al. (2015), then the original experimental paradigm must be adjusted to accommodate both frameworks. This is the purpose of the first experiment in this chapter. The aim is to create an experimental paradigm where at least one of the two main features of the Kvam et al. (2015) task that are not represented in standard EAM accounts are exchanged for more appropriate features.

If an appropriate comparison is to be made between standard EAMs and the quantum EAM proposed by Kvam et al. (2015) or more broadly, QPT vs CTP, then the original experimental paradigm must be adjusted to accommodate both frameworks. This is the purpose of the first experiment in this chapter. The aim was to create an experimental paradigm where at least one of the two main features of the Kvam et al. (2015) task, that are not represented in standard EAM accounts, are exchanged for more standard features. The present experiment kept the external stopping rule for the initial choice decision on trials, with and without an intermediate judgement. However, the final confidence judgement was replaced with a final binary judgement on the overall direction of the moving cloud of dots. Doing this bypasses the need to rescale and remap the confidence rating judgements in the original experiment and satisfies the binary response requirement of EAMs. Additionally, the difference between the final and intermediate judgement still allows us to capture overall confidence. The idea here is that confidence is determined, to some extent, by the degree to which a participant sticks with their initial answer, if asked the same question twice at different time points. Here we are assuming that inferred confidence is binary: participants either have high or low confidence in the initial answer given. For example, if on the first judgement participants indicate that the cloud of dots it moving to the left and then indicate on a subsequent judgement, based on the same trial question, that they still believe the cloud of moving dots is still moving to the left, this indicates a reasonable degree of confidence in their initial judgement. If participants indicate on the second judgement that the cloud of dots is now moving to the right, this should indicate a low level of confidence in their initial judgement. This is an approximate picture since it discounts the possibility of participants simply offering erroneous responses, still, to a first approximation it seems to offer a reasonable intuition.

The rationale for the above approach is that redesigning the original task by Kvam et al. (2015) into a task better suited for analysis by standard EAMs, while still capturing the main features of the original quantum task, would allow for a more accurate assessment of EAMs and therefore a CPT account of the observed phenomena. I believed that this new experimental paradigm would show that the results originally observed reflect the unique conditions of the original experiment and not the limitations of EAMs (as it turned out to be the case). Instead, if an appropriate experimental paradigm is designed to capture the main features of EAMs, it will reveal the ability of these models (and subsequently CPT) to capture the observed interference effect. As such, it was expected that results would not significantly differ from the original results observed by Kvam et al. (2015). To clarify, for Experiment 1 in this chapter, the hypothesis is that removing the external-stopping rule and replacing the final confidence rating decision with a binary decision will produce results indicative of an interference effect, as in the original experiment conducted by Kvam et al. (2015).

4.1.1 Method

Participants

We recruited 26 participants through City, University of London's internal participant recruitment platform. All participants had normal or corrected to normal vision. Participants were all compensated £15 for their participation. The sample size was guided by sampling considerations in the original Kvam et al. (2015) study.

The first experiment in this chapter was split into two stages of data collection. The first stage was collected as a stand-alone experiment and the second stage was collected during a two-part experiment. This two-part experiment was the final experiment in this chapter and involved the experiment presented here and an alternate version (presented later). This separation in data collection was done for any later modelling purposes. 10 participants were recruited for the first stage of data collection and 16 participants for the second stage.

Design and procedure

The experiment was a within-participants design with three variables: the number of judgements that had to be made on a trial (V1), stimuli coherence level (V2) and the moving direction of the stimuli (V3). V1 had two levels: either a single final direction judgement was made on a trial or there was an additional intermediate direction judgement. V2 had four levels: 0%, 5%,10 and 20% coherence. V3 had two levels: right or left.

The experiment consisted of a cloud of dynamic grey dots on a black background, with a proportion of the dots (determined by V2) moving either left or right (determined by V3). The remaining dots moved randomly. The stimuli moved at an average of 12 frames per second, with the motion speed fixed at 3 deg/s. Participants were initially presented with 1 practice block containing 16 trials. All trials had a coherence level of 40%, with half the trials having 40% of the cloud of dots moving to the right and the other half moving to the left. Each of these halves were further split in half. On one half of these trials, after 500ms, the stimuli became static and participants were prompted for a motor response, by pressing the space bar on the keypad (no intermediate judgement). After this initial response, the stimulus resumed its motion and then participants had to provide a second response on the overall direction of the moving dots, when they were confident of their response by pressing "z" on the keypad for left or "m" on the keypad for right (final judgement). On the other half of these trials, after 500ms, the stimulus became static and participants were prompted to make a judgement on the overall direction of the moving dots by pressing "z" on the keypad for left or "m" on the keypad for right (intermediate judgement). The second stage remained the same (final judgement). Only during the practice block were participants given feedback on their accuracy after each trial. For the 0% coherence trials, the correct response for half of these trials was "left" and for the other half was "right". These 0% percent coherence trials were presented for completeness.

Participants were then presented with 10 experimental blocks of trials each containing 40 trials. Each block consisted of 10 trials for each of the four coherence levels, with each set of 10 trials being equally split to have the dots moving either coherently to the right or left. Half of the total 10 blocks of trials contained only one binary direction judgement on a trial. The other half contained an additional intermediate binary direction judgement along with a final judgement on the trial.

4.1.2 Results

Descriptive results of mean accuracy rates across the 4 coherence levels show that rates increased with coherence level throughout the two conditions. Additionally, looking at the difference in accuracy between coherence levels on the two conditions yields positive and negative differences, refer to Figure 4.1 below. Note, that in the intermediate judgement condition a trial was defined as correct if the final direction decision matched the actual direction of the moving cloud of dots.

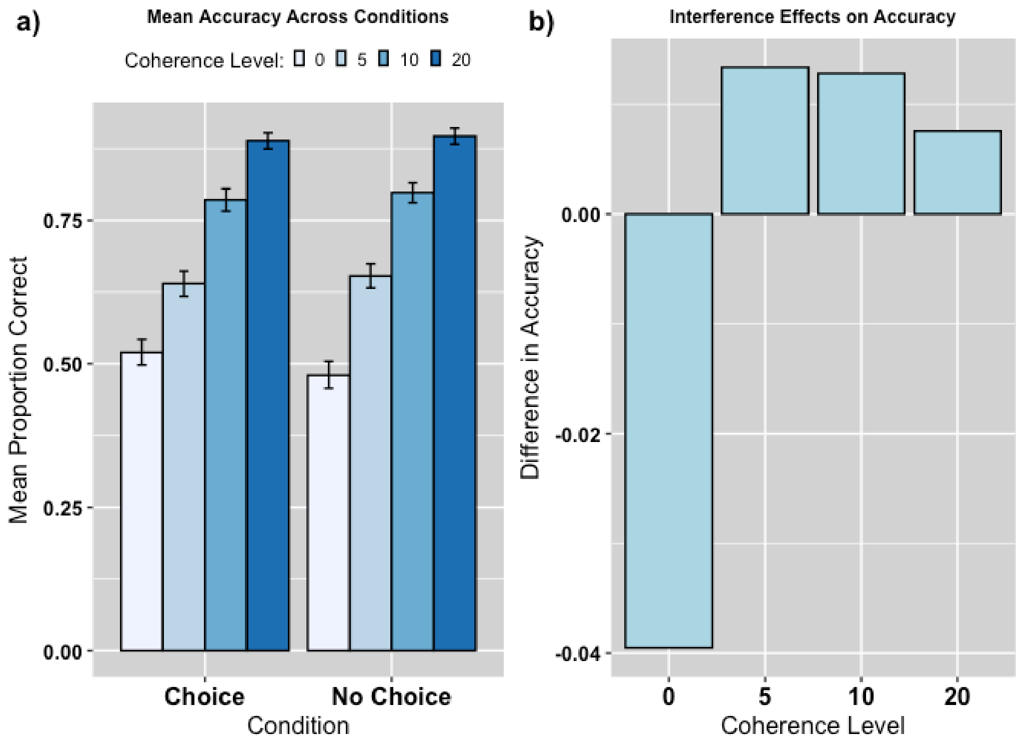

Figure 4.1. a) Mean accuracies across conditions and coherence levels. b) The difference between accuracies between no choice and choice conditions across coherence levels, illustrating differences in accuracies caused by an effect of interference. Error bars represent bootstrapped 95% confidence intervals.

Descriptive results of mean RTs across the 4 coherence levels show that RTs decreased with coherence level throughout the two conditions. Additionally, looking at the difference in RTs between coherence levels on the two conditions yields positive differences, refer to Figure 4.2 below.

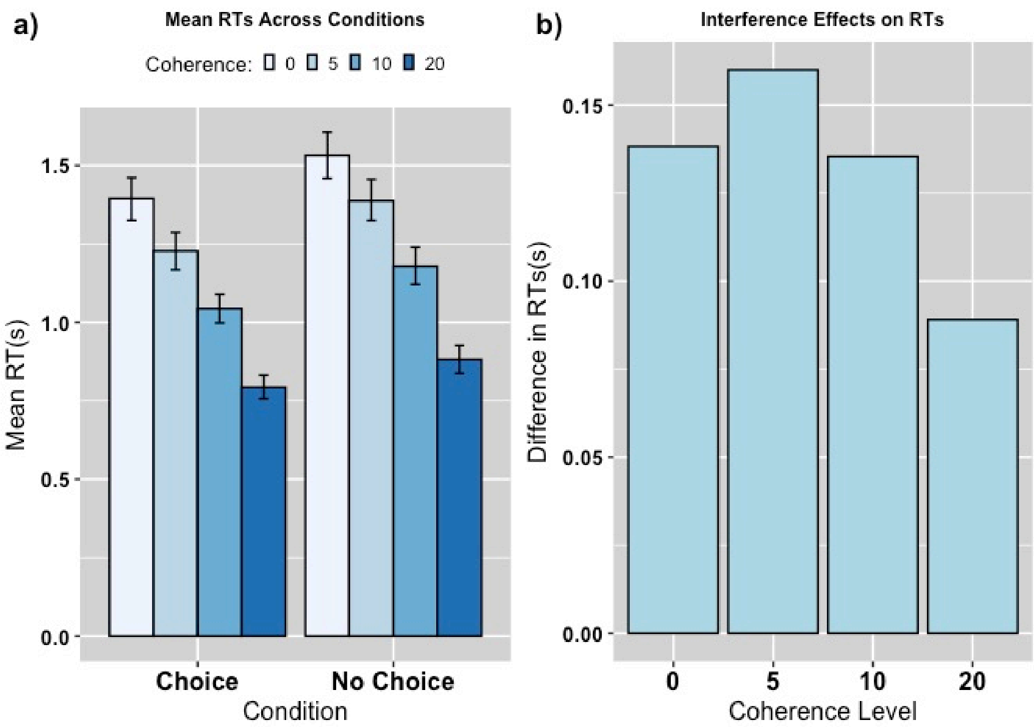

Figure 4.2. a) Mean RTs across conditions and coherence levels. b) The difference between RTs between no choice and choice conditions across coherence levels, illustrating any difference in RTs caused by an effect of interference.

Descriptively, these results do illustrate an interference effect, whereby an intermediate decision appears to produce differences in the accuracy and RT of a final decision, compared to if no intermediate decision was made.

To assess whether the difference between RTs in choice and no choice decision trials brought on by interference effects were statistically significant, inferential analyses were conducted. A 2(condition: choice vs no choice trials) x 4(coherence level: 0%, 5%, 10%, 20%) RM ANOVA was conducted on participant's RTs. Results found a significant main effect of condition: $F(1, 1293) = 35.242, p < .001, \eta_p^2 = 0.03$, with the choice condition having faster RTs ($M = 1.11$, $SD = 1.21$) than the no choice condition ($M = 1.24$, $SD = 1.34$) There was a significant main effect of coherence: $F(3, 3879) = 168.403, p < .001, \eta_p^2 = 0.12$. There was a non-significant interaction between the main effects of condition and coherence. Bonferroni post-hoc *t*-tests were conducted to assess how the RTs in the four coherence levels differed, refer to Table 4.1 below.

Table 4.1. The significant mean difference between RTs during trials with one of the four different coherence levels.

		Mean Difference	SE	*t*	p
0	5	0.148	0.029	5.076	< .001
	10	0.347	0.029	11.906	< .001
	20	0.617	0.029	21.205	< .001
5	10	0.199	0.029	6.830	< .001
	20	0.469	0.029	16.129	< .001
10	20	0.271	0.029	9.299	< .001

These interferential results support the quantum view that differences in RTs ought to be found between conditions, because of an effect of interference in the choice condition. This poses problems when determining whether the observed effect is itself a result of quantum-like interference or some form of random noise in the general processing system or response priming. Specifically, within the choice condition, participants had to provide an initial left or right keyboard response, with the subsequent judgement being a left or right keyboard confidence response. However, in the no choice condition participants initially only had to provide an arbitrary motor response by pressing the space key. It may be that the significantly lower RTs in the choice condition trial is a result of response priming on these trials. The final decision is very similar to the initial decision, in the choice condition trial, involving no finger rearrangement. Contrary to the no choice condition, finger/hand rearrangement occurs between the two responses on these trials. Given that the significant difference in RTs is approximately 130ms, this shift in response types between conditions could be largely responsible for this difference. As such, it is difficult to assess the independent effect of interference in this case, as any response priming effects were not controlled for.

Differences in accuracy between the two main conditions were also analysed for any significant differences. A 2(condition: choice vs no choice trials) x 4(coherence level: 0%, 5%, 10%, 20%) Friedman test was performed on participants' accuracy rates between choice and no decision trials, and the various coherence levels. Results showed that for the main effect of condition (choice vs no choice trials), there was no significant difference. There was a significant main effect of coherence level, $X^2(1) = 961.054$, $p < .001$. As the most pertinent main effect of condition was found to be non-significant, no post-hoc analyses were performed on the significant main effect of coherence level. This is because the interference effect itself is dependent on a significant difference between conditions and not solely coherence levels. As such, further analyses are made redundant by the absence of a significant difference between conditions.

It remains difficult to appreciate how a general effect of interference can be restricted to RTs only. As previously mentioned, priming effects remain a very plausible and unintended effect in the original results. Furthermore, from the perspective that the observed effect of interference is a response priming effect, it is understandable why a significant difference between conditions would only be found in RTs and not accuracy.

Additionally, the present experiment attempted to create a more standard EAM paradigm to determine if this effect of interference can be observed in an experimental paradigm suited for a classical, rather than a quantum approach. The findings reported here are largely similar to those reported in the original experiment. In other words, a largely classical probability framework appears to be capturing the apparent quantum effect in the Kvam et al. (2015) study.

4.1.3 Discussion

This experiment aimed at creating an experimental paradigm that allowed for the original paradigm designed by Kvam et al. (2015) to be more representative of the EAM framework. This was done by replacing the final confidence rating decision with a binary decision and removing the external-stopping rule for the final confidence decision. The descriptive results do show a difference in RT and accuracy behaviour between the conditions with and without an intermediate judgement, indicative of an effect of interference. Although this difference is consistent throughout the RT and accuracy descriptive data, inferential results are not congruent with this finding. Nonetheless, the inferential results are congruent with the original experiment and support my hypothesis.

Inferential results show that while there is a significant difference in RTs between the two conditions, accuracy rates do not show this difference. However, it can be expected that the assumed effect of interference is not limited to RTs alone, but extends to all relevant behaviours including accuracy. Nonetheless, this was not found in the present experiment and is consistent with the original experiment.

The most pertinent finding is that while the present experiment moved the experimental paradigm closer to a standard EAM paradigm, the experimental results between the two experiments are largely the same. One possible conclusion of this, is that the observed interference effect is itself not a uniquely quantum effect that requires an explicit quantum EAM model to be detected. That is, a quantum probability framework does not appear to have an advantage over a more classical probability framework in identifying the observed interference effect. Rather, the results presented here suggest that more standard EAMs paradigms based on CPT are able to replicate and capture this supposedly quantum specific effect. This then poses the question of what may be responsible for the observed effect, if it is not an explicitly quantum effect? I believe that uncontrolled factors are responsible, specifically, response priming. The original and present experiment had a response pattern that remained static between judgments in the choice condition, which facilitated the perfect conditions for response priming. Coincidentally, it is this condition which displayed the observed interference effect.

Although the present experiment aimed to bring the original experiment in line with an EAM experimental paradigm, the external stopping rule from the original experiment remained. As such, it remains undetermined if the reintroduction of experimental features from the original experiment will yield different results. Realigning the present experimental paradigm wih the original one may yield results that explicitly differ to those found in the present paradigm based on more standard CPT EAM principles. To determine this, the experiment presented here must be brought back in line with the original. This requires replicating the present experiment with an external stopping rule for the final confidence judgement, similar to the original experiment. As such, we will introduce a second external stopping rule for second stage processing in an identical second experiment mentioned below.

The argument presented here is explicitly aimed at the EAM model proposed by Kvam et al. (2015). Specifically, mainstream EAMs have to abide by the constraints from classical probability theory and other core assumptions (e.g., the “read-out” assumption). By contrast, the EAM proposed by Kvam et al. (2015) involved the rules from quantum theory and so was intended as a kind of EAM different from standard models. The key question I have tried to address in this chapter is: are the alternative probability rules from quantum theory essential in explaining the empirical results observed by Kvam et al. (2015)? There is a completely valid point that with general frameworks it is always possible to tweak models to accommodate some empirical finding. However, I argue that the classical probability EAM that I proposed is a fairly natural extension of standard EAM models. Moreover, I argued that if, as Kvam et al. (2015) proposed, the interference effect can be located in the process of evidence accumulation, if the quantum EAM is tweaked to represent more standard EAMs, the interference effect should still be observed. If this is the case, than this calls into questions the argument proposed by the authors of the original experiment, that the observed effect is uniquely quantum and can only be captured by a quantum specific approach. If the interference effect is no longer observable, then we have to question whether the interference effect is the result of quantum-like processes (the theoretically significant part of Kvam et al.'s proposal) vs. more incidental architectural issues with their model (which would be less theoretically interesting).

The final subtlety is that, instead of exploring a more standard EAM for the data, I thought it more expedient to modify the experimental paradigm presented by the researchers, with a view to remove complications which made the paradigm inapplicable with standard EAMs. A natural next step for future work on this project is to then fit a more standard EAM to the such data, though I would hope that the current results already offer a degree of confidence regarding my conclusions and question the interpretation from Kvam et al. (2015).

Section 4.2 Time Constraint on the Interference Effect

Introduction

The present experiment attempts to bring the paradigm of the first experiment more in line with the original. The aim is to see whether the observed interference effects change to include a time-bound second stage processing window of approximately 1 second, as in the original experiment. As such, the present experiment will be largely identical to the first experiment, except that participants were prompted for a final binary decision exactly 1 second after their initial decision. This is because results from the original experiment show that the interference effect was strongest when participants had to provide a final decision 0.75s and 1.5s after making an initial binary or motor decision. Therefore, the hypothesis is that reintroducing an external stopping rule into the experimental paradigm will, again, produce results indicative of an interference effect, as in the original experiment.

4.2.1 Method

Participants

We recruited 10 participants through City, University of London's internal participant recruitment platform. All participants had normal or corrected to normal vision. Participants were all compensated £10 for their participation. The sample size was exploratory, but was broadly guided by sampling considerations in the original Kvam et al. (2015) study.

Design and procedure

The design and procedure of this second experiment remained largely the same as the first. However, exactly 1 second after the initial binary direction decision was made on trials with a binary intermediate judgement, the dynamic stimuli became static and participants were prompted to make a second binary direction decision similar to the initial decision. Additionally, participants were presented with 10 experimental blocks of trials each containing 28 trials. Each block consisted of 7 trials for each of the four coherence levels, with half of the trials in each block being equally split between trials where the correct stimulus direction was right and left. Half of the total 10 blocks of trials contained only one binary direction judgement on a trial. The other half contain an additional intermediate binary direction judgement along with a final judgement on the trial. Participants were presented with fewer trials due to a reduction in the funds available to compensate them for completing more trials, as in the first experiment.

4.2.2 Results

Descriptive results of mean accuracy rates across the 4 coherence levels show that rates increased with coherence level throughout the two conditions. Additionally, looking at the difference in accuracy between coherence levels on the two conditions yields positive and negative differences, refer to Figure 4.3 below. Note, that in the intermediate judgement condition a trial was defined as correct if the final direction decision matched the actual direction of the moving cloud of dots.

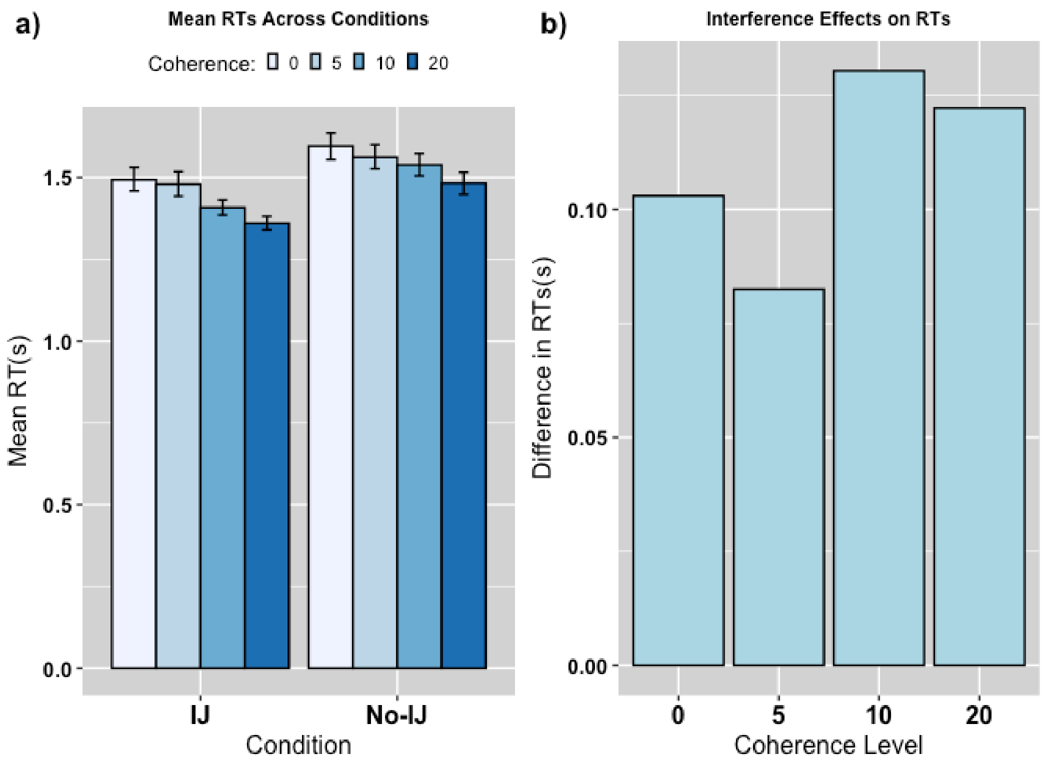

Figure 4.3. a) Mean accuracies across conditions and coherence levels. b) The difference between accuracies between no choice and choice conditions across coherence levels, illustrating any difference in accuracies caused by an effect of interference.

Descriptive results of mean RTs across the four coherence levels show that RTs decreased with coherence level throughout the two conditions. Additionally, looking at the difference in RTs between coherence levels on the two conditions yields positive differences, refer to Figure 4.4 below.

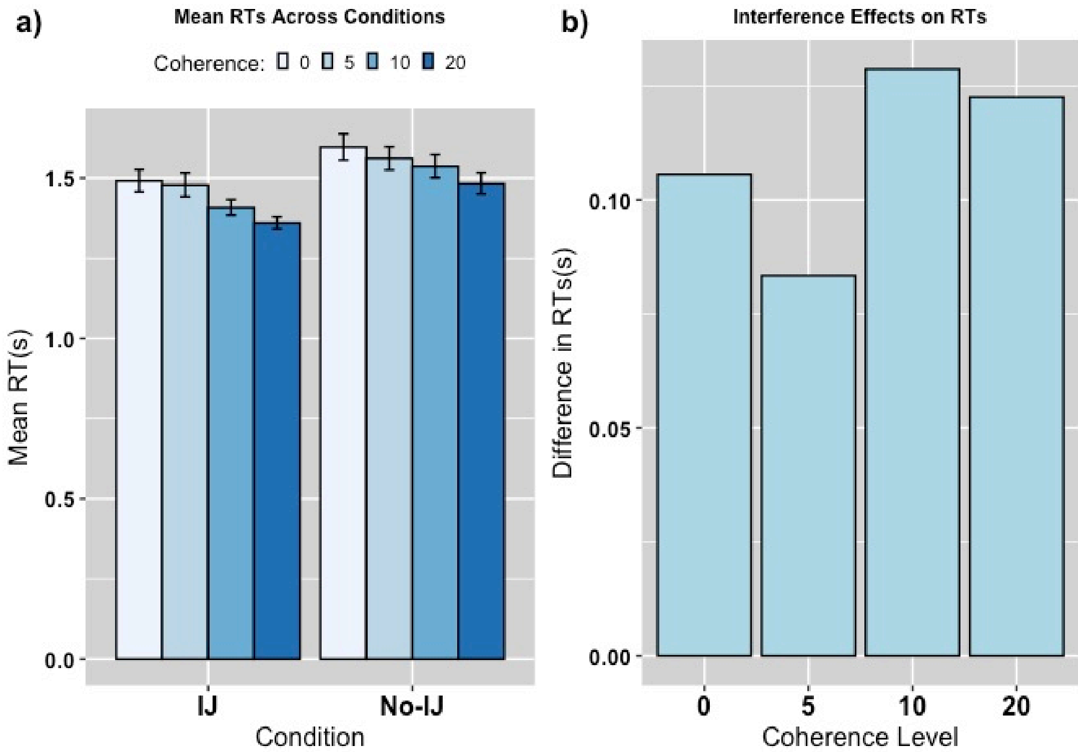

Figure 4.4. a) Mean RTs across conditions and coherence levels. b) The difference between RTs between no choice and choice conditions across coherence levels, illustrating any difference in RTs caused by an effect of interference.

Descriptively, these results do illustrate an interference effect, whereby an intermediate decision appears to produce differences in the accuracy and RT of a final decision, compared to if no intermediate decision was made.

To assess whether the difference between RTs in intermediate and non-intermediate decision trials brought on by interference effects were statistically significant, inferential analyses were conducted. A 2(condition: choice vs no choice trials) x 4(coherence level: 0%, 5%, 10%, 20%) RM ANOVA was conducted on participant's RTs. Results found a significant main effect of condition: $F(1, 349) = 60.388$, $p < .001$, $\eta_p^2 = 0.15$, with RTs being faster in the choice condition ($M = 1.43$, $SD = 0.35$), compared to the no choice condition ($M = 1.5$, $SD = 0.42$). There was a significant main effect of coherence: $F(3, 1047) = 18.775$, $p < .001$, $\eta_p^2 = 0.05$. There was a non-significant interaction between the main effects of condition and coherence. Bonferroni post-hoc *t*-tests were conducted to assess how the RTs in the four coherence levels differed, refer to Table 4.2 below.

Table 4.2. The significant mean difference between RTs during trials with one of the four different coherence levels.

		Mean Difference	SE	*t*	p
0	5	0.024	0.018	1.374	1.000
	10	0.071	0.018	4.024	< .001
	20	0.123	0.018	6.937	< .001
5	10	0.047	0.018	2.650	0.049
	20	0.099	0.018	5.563	< .001
10	20	0.052	0.018	2.913	0.022

These interferential results support the quantum view that differences in RTs ought to be found between conditions, because of an effect of interference in the intermediate judgement condition. This poses problems when determining whether the observed effect is itself a result of quantum-like interference or some form of random noise in the general processing system or response priming.

A 2(condition: choice vs no choice trials) x 4(coherence level: 0%, 5%, 10%, 20%) Friedman test was performed on participants' accuracy rates between choice and no decision trials, and the various coherence levels. Results showed that for the main effect of condition (choice vs no choice trials), there was no significant difference. There was a significant main effect of coherence level, $X^2(1) = 252.925$, $p < .001$. As was the case above, since the effect of condition was found to be non-significant, no post-hoc analyses were performed on the significant main effect of coherence level.

4.2.3 Discussion

The purpose of this experiment was to bring the first experiment in this chapter closer to the original experiment conducted by Kvam et al. (2015). This would make it easier to assess if the reintroduction of the features within the quantum experimental paradigm adopted in the original experiment, would yield different results indicative of a quantum specific effect.

However, for completeness, a final experiment is needed, which replicates the original experiment. This will allow for a complete comparison between the results of the various experimental paradigms in this chapter and allows for an analysis of how results differ as a more quantum or classical approach is taken to investigate the observed interference effect. Experiment 1 and 2 presented in this chapter had a final binary decision on each trial. In the original experiment, this final decision was made on a 100-point confidence scale. However, the two experiments presented in this chapter do not contain a final confidence rating decision. As such, a final experiment identical to Experiment 2 will be conducted, except that the final decision will be a collapsed 6-point confidence decision. This will bring the present experimental paradigm closer in line with the original experiment and will help determine if the complete reintroduction of quantum like features into the experimental paradigm will yield results explicitly different to previous experimental variations.

Section 4.3 Interference Effects on Confidence

Introduction

The previous two experiments in this chapter have altered the original experimental paradigm by introducing more CPT features compatible with standard EAMs. However, both these experiments have not replicated the final choice-confidence decision that is made at the end of each trial in the original experiment. For completeness, a final replication of the original experiment is conducted, to allow for a comparison between the results of the various experimental manipulations in this chapter and how results differ as a more quantum or classical approach is taken. Therefore, the hypothesis is that reintroducing a final confidence rating judgement into the experimental paradigm will, again, produce results indicative of an interference effect.

4.3.1 Method

Participants

We recruited 16 participants through City, University of London's internal participant recruitment platform. All participants had normal or corrected to normal vision. Participants were all compensated £10 for their participation. This sample size is consistent with previous experiments in this series and Kvam et al. (2015).

Design and procedure

The design and procedure of this third experiment was largely identical to the second experiment, except that participants had to make a final choice-confidence decision on each trial. During this final decision if participants believed the stimuli was coherently moving towards the top-left, they had to press "A" on the keypad if they were very confident, "S" if they were moderately confident and "D" if they were slightly confident. If participants believed the stimuli was coherently moving towards the top-right, they had to press "H" on the keypad if they were very confident, "J" if they were moderately confident and "K" if they were slightly confident.

4.3.2 Results

Descriptive results of mean accuracy rates across the 4 coherence levels show that rates increased with coherence level throughout the two conditions. Additionally, looking at the difference in accuracy between coherence levels on the two conditions yields positive and negative differences, refer to Figure 4.5 below. Note, that final choice-confidence decisions were collapsed across left and right direction responses to produce binary responses.

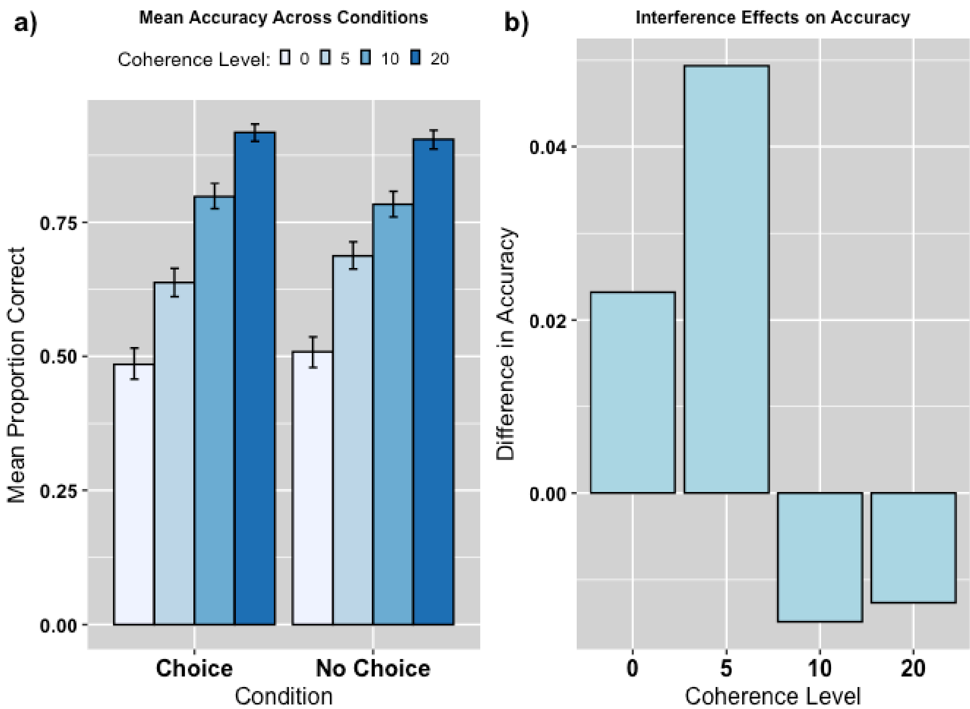

Figure 4.5. a) Mean accuracies across conditions and coherence levels. b) The difference between accuracies between no choice and choice conditions across coherence levels, illustrating any difference in accuracies caused by an effect of interference.

Descriptive results of mean RTs across the 4 coherence levels show that RTs decreased with coherence level throughout the two conditions. Additionally, looking at the difference in RTs between coherence levels on the two conditions yields positive and negative differences, refer to Figure 4.6 below.

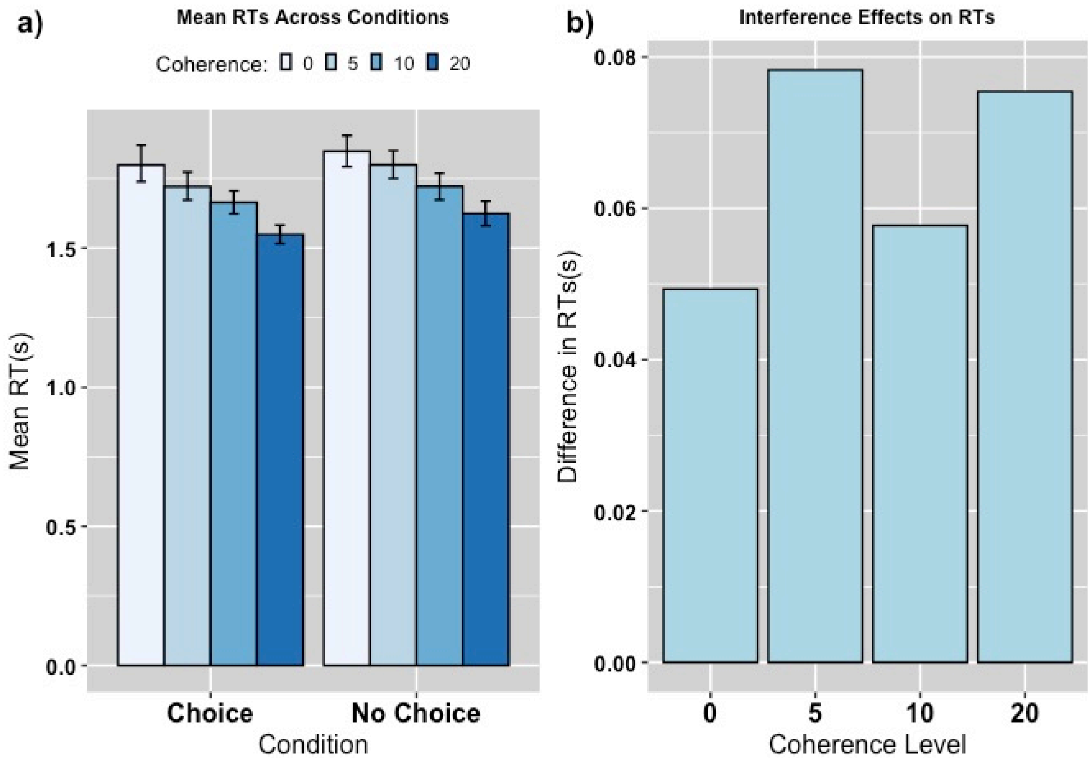

Figure 4.6. a) Mean RTs across conditions and coherence levels. b) The difference between RTs between no choice and choice conditions across coherence levels, illustrating any difference in RTs caused by an effect of interference.

Descriptively, these results do illustrate an interference effect, whereby an intermediate decision appears to produce differences in the accuracy and RT of a final decision, compared to if no intermediate decision was made.

Descriptive analyses of confidence ratings across conditions show results inconsistent with the those found by Kvam et al. (2015). Contrary to the original results, participants in the choice condition had higher confidence ratings (high confidence), relative to those in the no choice condition. Additionally, unlike in the original experiment, overconfidence was not the biggest difference between confidence ratings in the two conditions. Participants in the choice condition were also more underconfident compared to those in the no choice condition and this difference was comparably greater, compared to the degree of overconfidence between conditions. For medium confidence ratings, this pattern was reversed. Participants in the choice conditions had lower medium confidence, with the magnitude of this difference being greater than for over and under confidence. Overall, results showed a shift towards more extreme ratings (high vs low ratings), as opposed to medium ratings.

In the present experiment, confidence was not recorded as in the original experiment. Although the present experiment involves a more simplified way of measuring confidence, the empirical approach helps highlight the ambiguity associated with measuring and interpreting confidence. I do not believe that a clear argument can be made as to which confidence approach ought to be preferred in this instance. I believe that the present approach represents a more streamlined version of what was used in the original experiment. As such, the difference between the two scales should be considered negligible and findings ought to be consistent with the original experiment. Please refer to Figure 4.7 below.

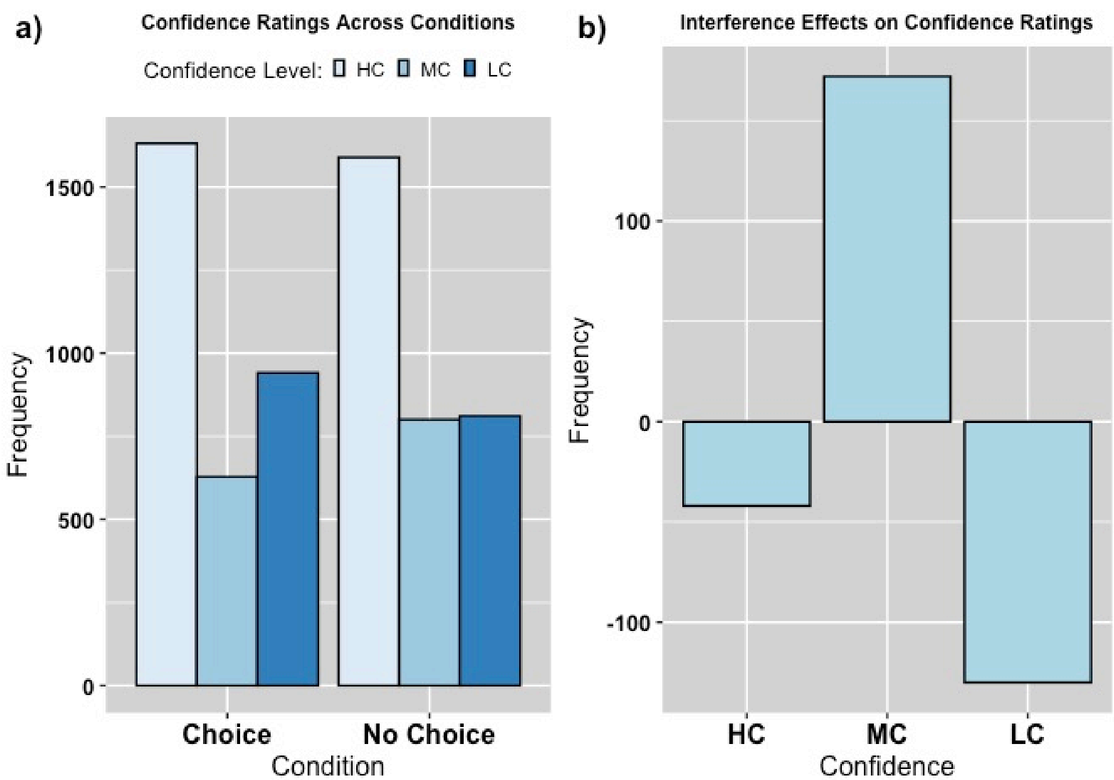

Figure 4.7. a) Confidence ratings for choice and no choice conditions. b) The difference in confidence ratings between conditions, calculated as no choice – choice for each rating. HC defines high confidence ratings, MC defines medium confidance ratings and LC defines low confidance ratings.

To assess whether the difference between RTs in choice and no-choice decision trials brought on by interference effects were statistically significant, inferential analyses were conducted. A

2(condition: choice vs no decision trials) x 4(coherence level: 0%, 5%, 10%, 20%) RM ANOVA was conducted on participant's RTs. Results found a significant main effect of condition: $F(1, 799) = 12.214, p <.001, \eta_p^2 = 0.02$, with RTs in the choice condition being faster (M = 1.68, SD = 0.84) than in the no choice condition (M = 1.75, SD = 0.85) There was a significant main effect of coherence: $F(3, 2397) = 35.488, p <.001, \eta_p^2 = 0.04$. There was a non-significant interaction between the main effects of condition and coherence. Bonferroni post-hoc *t*-tests were conducted to assess how the RTs in the four coherence levels differed, refer to Tab. 4.3 below.

Table 4.3. The significant mean difference between RTs during trials with one of the four different coherence levels.

		Mean Difference	SE	*t*	p
0	5	0.062	0.024	2.585	0.059
	10	0.128	0.024	5.364	< .001
	20	0.235	0.024	9.846	< .001
5	10	0.066	0.024	2.778	0.033
	20	0.173	0.024	7.261	< .001
10	20	0.107	0.024	4.482	< .001

A 2(condition: choice vs no choice trials) x 4(coherence level: 0%, 5%, 10%, 20%) Friedman test was performed on participants' accuracy rates between choice and no decision trials, and the various coherence levels. Results showed that for the main effect of condition (choice vs no choice trials), there was no significant difference. There was a significant main effect of coherence level, $X^2(1) = 627.156$, $p < .001$. Again, the main effect of condition was non-significant, so no post-hoc analyses were performed on coherence level.

These interferential results again support the quantum view that differences in RTs ought to be found between conditions, because of an effect of interference in the intermediate judgement condition, consistent with Experiment 1 and 2 in this chapter

4.3.3 Discussion

The aim of this chapter was to determine whether the interference effect observed by Kvam et al. (2015) originated from more explicit interference in the evidence accumulation processes or in another process. In order to do this, the original experimental paradigm was redesigned to align with an EAM framework to better assess if the observed effect could still be detected. If the original results could be replicated as the experimental paradigm shifted towards a more standard EAM framework, and consequently a more CPT approach, this would call into questions the purpose of a specific quantum EAM approach.

I also believe that these present results, differences in RTs between conditions in the present and original experiments, do not support the idea that a clear quantum effect of interference is responsible for a difference in behaviour between experimental conditions. The last experiment in this chapter represented a near replication of the original experiment. However, the results were largely similar to those found in Experiment 1. I argue on the basis of the results of Experiment 1, that a standard EAM framework based on CPT is still able to capture the supposedly quantum interference effect. The original experiment and Experiment 2 and 3 presented in this chapter represent an experimental paradigm that fails to capture the main qualitative feature of the EAMs they are investigating. That is, that participants are assumed to make decisions at their own discretion (internal stopping rule), as opposed to when prompted (external stopping rule). This is an experimental feature that the original paradigm must capture, if generalisations are to be made to the entire class of EAMs that do not presently represent this feature. However, in Experiment 1 of this chapter, where this qualitative feature was partially captured, the results were still largely consistent with the original experiment. Furthermore, whether the results attributed to a quantum effect of interference are actually distinguishable from other plausible explanations such as response priming, is unconclusive.

If the interference effect observed in the original Kvam et al. (2015) paper was not the result of some explicit interference in the evidence accumulation process, as would otherwise be suggested by the S-BSP model, then the first experiment in this chapter ought to have results that clearly do not show this effect. This is because the first experiment presented in this chapter captured the most EAM features, relative to the subsequent two. However, the results of this experiment still found an effect of interference among RTs.

I find it hard to reconcile how a general effect of interference can be restricted to RTs. It seems reasonable that if an effect of interference was responsible for the observed difference in RTs between conditions, then there should be some noticeable effect on accuracy. This leads me to conclude that this pattern of results, instead, represents alternative causes. Specifically, a plausible interpretation is that results are due to response priming brought on by the finger arrangements and response keys used by participants throughout the original experiment and the experiments presented here.

Assuming that results are due to response priming, the pattern of results observed in the original experiment and the present experiments make more sense. RTs in the choice condition, where participants had to provide an initial keyboard response similar to a final keyboard response, were consistently and significantly lower than RTs on the no choice condition. Furthermore, the similarity in keyboard responses represent good conditions for response priming, where the final response is primed by the initial preceding response in this condition.

Specifically, within the choice condition, participants had to provide an initial left or right keyboard response, with the subsequent response being a left or right keyboard confidence response. However, in the no choice condition participants initially only had to provide an arbitrary motor response by pressing the spacebar key. It may be that the significantly lower RTs in the choice condition is a result of response priming on these trials. The final decision is very similar to the initial decision in the choice condition, involving no finger rearrangement. Contrary to the no choice condition, finger/hand rearrangement occurs between the two responses in this condition. Given that the significant difference in RTs is approximately 130ms, this shift in response types between conditions could be largely responsible. As such, it is difficult to assess the independent effect of interference in this case, as any response priming effects were not controlled for.

The present chapter showed that quantum constraints on EAMs appear to be limited in their applicability to EAMs in general. This approach also seems limited in its ability to initially construct an experimental paradigm that both represents the main features of EAMs and captures an effect of interference. Although this chapter has attempted to show how this problem can be solved, it has highlighted how standard CPT approaches can be extended (experimentally in this case), to capture decision-making behaviour thought to be beyond its framework. The results presented in this chapter also demonstrate how the extended EAM in chapter two can be generalised to other decision-making effects to reveal their underlying cause. In the end, the results in this chapter call into question the domain-general nature of interference effects found in the original experiment and whether they are reliably distinguishable from response priming.

Chapter Five: Speed-Accuracy Trade-off

Section 5.1 Manipulating the Speed-Accuracy Trade-Off with EAM Response Thresholds

Introduction

In everyday life we are often presented with scenarios where we must perform a certain visual search. This can be something as simple as finding misplaced keys. Let us take the hypothetical scenario where an individual has lost their car keys and is certain that they are misplaced within their house. There are a multitude of ways in which the individual can search for their lost keys. For example, the individual can begin their search in the living room, then the bedroom and end in the hallway. Alternatively, they can search in the reverse order. In any case, the precise way in which the individual decides to search for their lost car keys is defined as their search strategy. The more places there are to search, the more complicated the search strategy becomes. In all these types of scenarios with multiple locations to search, there is a point in time at which the individual will stop searching if they have not found what they are looking for in their present location and go search another location. In our present example, this represents the time at which the individual stops searching one room of the house to go search another.

In the search scenario presented there are two main forces driving performance: 1) the search strategy used and 2) the point in time at which the individual stops searching one location and begins searching another. However, two questions remain: which search strategy should the individual use and at what exact point in time should the individual go and search a different location? Both these questions refer to optimality and how best to maximise performance. Expanding the second point we can find the existence of a speed-accuracy trade-off. In the present scenario we know that the longer the individual spends searching in one of the rooms in the house, the higher the probability that they will correctly assess if the car keys are or are not in that room. However, we also know that the individual will not indefinitely search that one room. They may decide to search another room at some point in time, as per their search strategy or they may decide to stop searching altogether. In either case, a certain amount of time spent searching corresponds with a certain amount of accuracy. The individual must decide if they want to trade time for accuracy (search for a short period of time but with little accuracy), or trade accuracy for time (search with higher accuracy but for a longer period of time). It is by exploring this speed-accuracy trade-off in visual searches that EAMs can provide insights into what is the main driving force behind performance in these tasks. As EAMs can successfully and robustly predict performance driven by speed-accuracy trade-offs (Brown & Heathcote, 2008; Donkin et al., 2009). Nevertheless, let us begin by exploring the question of what search strategy ought to be used to maximise the chances of a forager attaining what they are searching for.

Search strategies

Boot, Becic, & Kramer (2009) observed the oculomotor differences in a series of visual search tasks: dynamic dot detection task, an efficient search task (a tilted line among vertical lines), an inefficient search task (a T among Ls) and a change blindness task in which participants searched for changes in driving scenes. On these basic visual search tasks, the researchers found that in the absence of any response feedback, participants' visual search strategies converged across the different tasks. Boot et al., (2009) propose that these findings indicate that in the absence of response feedback providing information on maladaptive search strategies, a default strategy is employed. Boot et al., (2009) subsequently replicated this experiment with the same participants but provided response feedback at the end of each trial. In this variation, findings showed that feedback did cause a divergence in employed search strategies in participants across tasks. In the context of optimising the speed-accuracy trade-off, when participants were made aware of the maladaptive search strategies they were using through feedback, they attempted to identify strategies that better optimised performance. In essence, participants began to train their search strategy to yield optimal speed-accuracy trade-offs and thereby improve performance. However, this process of training largely applies to inter-task and not intra-task differences. Although inter-task search strategies were found to eventually converge in the presence of feedback, intra-task search strategies remained spread across a range of overt and covert strategies. Therefore, during intra-task searches there was no singular identifiable search strategy used by all participants.

For instance, Boot, Kramer, Becic, Wiegmann, & Kubose (2006) presented participants with a simple visual search paradigm. In this search task participants were shown a display containing 24 continuously moving dots. During some trials a new dot appeared in the display and participants had to press a button when this occurred. The results found a surprisingly large variation in accuracy, with some participants almost always detecting the new dot and others missing 50% of the time or more. As the researchers went on to argue, large variations in accuracies were determined by a large range of employed search strategies by participants. However, from an EAM perspective, this would be expected. EAMs define response time tasks like a visual search task as having two broad features: inherent properties of the task (the amount of information that can be derived from the stimuli) and participant-controlled properties (response conservativeness). The participant-controlled properties of the task are assumed to directly impact the response threshold value (response conservativeness) in EAMs and relate to a participant's attempt at adjusting their speed-accuracy trade-off. In a visual search task, search strategies are set to optimise the underlying speed-accuracy trade-off. However, this does not imply that only a single search strategy is capable of optimising the underlying trade-off. Instead, a set of different search strategies may be capable of optimising the underlying trade-off. Consequently, equal weight is given to the usage of any of the search strategies in this set. Therefore, a wide range of covert and overt search strategies being used in a visual search task would be expected, as found in Boot et al. (2006). The only relevant aspect is the amount of evidence required before deciding on a particular response. This is explicitly different from search strategies that are primarily concerned with searching patterns

From an EAM perspective the precise search strategy used during a visual search is dependent on and reflects the underlying speed-accuracy trade-off. In the context of the results by Boot et al. (2006), this would explain why even though there was a varied set of search strategies between participants, their individual performances were largely similar. That is, although search strategies varied between participants, the underlying speed-accuracy trade-offs may have been similar and it was this that ultimately drove performance. For example Nowakowska, Clarke, and Hunt (2017) presented participants with a compound search array divided into two sections and made up of line segments. The left side of the visual array was heterogenous, such that the line segments were randomly orientated. The right half of the visual array was homogeneous, such that line segments were largely orientated in the same direction. The target stimulus was a line orientated 45 degrees to the right and was located on both sides of the entire visual display. Using eye tracking to monitor saccade patterns, the researchers found that even though participants could identify the target stimuli on the homogeneous side with peripheral vision (as found in the pilot results), direct fixations on the stimuli were still being made. The researchers argue that this sub-optimal search strategy resulted from participants trading speed for a perceived (but not actual) gain in response certainty. In effect, participants' speed-accuracy trade-off setting determined their search strategy.

Results by Nowakowska et al. (2017) further highlight the EAM perspective that the most significant indicator and driver of performance on a foraging task is how the speed-accuracy trade-off is set. The researchers found that the dominant search strategy in the experiment involved more saccades within a region of the visual display that corresponded with participants trading accuracy for speed. In other words, participants' response conservativeness for the homogeneous target remained sufficiently high that more information regarding the target had to be gathered beyond that gathered through peripheral vision (the optimal strategy). This resulted in a search strategy that largely resembled the search strategy for the heterogenous target. However, questions remain as to whether or not participants are attempting to optimise their speed-accuracy trade-off, if optimisation is even possible and how an optimal trade-off can be identified.

Giving-Up Times and the Marginal Value Theorem

Solutions to such problems have their background in empirical work relating to giving-up times (GUTs). During field observations of carrion crows by Croze (1970), it was found that crows spent a constant amount of time foraging for and consuming foods in any given region. Only after a fixed amount of time did crows leave the region to begin foraging in another region. Furthermore, the observed crows moved onto another foraging region after some fixed amount of time, even if more edible food sources were present in the region. Croze's (1970) observations led to the notion of GUTs, that are defined as the time from finding a food item to abandoning the food item. The forager's aim is to maximise the average amount of food consumed over a series of N patches, assuming that each patch has diminishing returns. Food consumption here is the reward index for the foraging carrion crow and functions as a unit of measurement for the task's reward. A variety of reward indexes can be selected and primarily depend on what the reward for performing a particular task is (Starns & Ratcliff, 2010). Generalising from this, GUTs can be seen as the total amount of time that should be spent actively performing a relevant task in order to maximise performance and subsequently the reward (Lima, 1977; Pyke, Pulliam, & Charnov, 1977). In relation to the example at the start of this chapter, GUTs represent the total amount of time the individual should spend searching one of the rooms in their house for the lost keys before searching another room. The probability of finding the lost car keys is the reward index.

GUTs were first formalised through the marginal value theorem (MVT) by Charnov (1976). The MVT states that there exists an optimal amount of time that an animal should spend foraging for a resource in a region that diminishes over time. This predefined time is optimal because *giving-up* before or after this point does not maximise average returns across a series of similar tasks with diminishing returns. Therefore, the MVT provides a solution to the question: how long should one spend on each individual task with diminishing returns, in order to maximise average returns over a series of similar tasks? The MVT states that the answer is given by the amount of time that corresponds with when the instantaneous rate of reward equals the average rate of reward for any given patch searched, t^*. Generalising from the MVT, when the average reward over time spent acquiring the reward is equal to the rate of change of this average, an optimal decision strategy is identified for the relevant task. In other words, this strategy is the GUT. Before this point, the average return over a series of tasks per unit of time spent acquiring the return is sub-maximal. Therefore, to maximise average returns while minimising costs associated with doing so, one must stop (*give-up*) the task they are performing at time = t^*. The MVT thereby provides a precise answer to the question of at what point in time a different location should be searched. (Green, 1984). Furthermore, GUTs and the MVT constitute a larger body of literature on optimal foraging theory (OFT), that asserts that foraging behaviour is primarily driving by the optimisation or sub-optimisation of GUTs explicitly and not search strategies.

Optimal Foraging Theory

OFT is an extension of the initial empirical field work conducted by Croze (1970); Smith and Darwkins (1971); Charnov (1976) and others. Overall, OFT is a theoretical framework that aims to show how animals forage in a way that optimises performance (Kamil, 1983; Smith, & Dawkins, 1971). Animals seek to maximise their reward index while minimising the costs associated with doing so, in order to maximise overall returns (Engen & Stenseth, 1984). This amounts to optimising the associated speed-accuracy trade-off. Animals seek to maximise their accuracy on the task, while minimising the amount of time spent performing the task. This is similar to the previously discussed MVT, where the derived optimal decision strategy, t^*, maximises returns while minimising costs associated with doing so. The identified optimal search strategy for a task is defined as the optimal decision rule in OFT and can be derived through the MVT, for example (Krebs, Kacelnik & Taylor, 1978; Pyke, 1984). Essentially, all optimal decision rules are optimal decision strategies, in that they maximise some desired return while minimising costs associated with doing so (Engen, & Stenseth, 1984; McNamara & Houston, 1985). However, more recent experimental results have emerged demonstrating that humans also seem to abide by such optimal decision rules. Moreover, these findings are most apparent in the domain of perceptuo-motor and perceptual decision-making.

Optimal Human Decision-Making

Cain, Vul, Clark and Mitroff, (2011) assessed whether the direct principles of OFT are applicable to human participants in visual search tasks. Cain et al. (2011) presented three groups of participants with a visual search task in which they had to identify a geometric T-shape among several distractor stimuli. Within each trial, 40 items were shown, made up of a mixture of both target and distractor stimuli, with 0-12 of the stimuli being targets on each trial. Participants could choose when to stop searching for targets and move onto the next trial. Each group contained different proportions of trials within which targets could be found. On each of the three groups, 25%, 50% and 75% of the trials contained targets respectively. However, as the proportion of trials with targets increased, the number of displayed targets to be found decreased on those trials. Participants were assumed to maximise the number of targets found, Γ. A variant of the MVT was used to derive participants' GUTs, t*. The MVT states that once the rate of return for the current location is equal to the rate of return for the environment, the animal (or individual) should abandon the task at hand in the current location. After such a point, they should then move onto the next location in the same environment. Given that participants could choose to end a trial when they saw fit, they could therefore choose when to stop searching for targets at their discretion. As such, given the MVT and assuming participants use optimal decision strategies, participants should stop searching for targets on a given trial once Γ for that trial is maximal.

Cain et al. (2011) found that within the 50% condition participants performed near optimally, in that they were able to nearly maximise Γ. That is, participants stopped searching for targets and moved onto the next trial when the rate of return for the current trial was equal, or near equal to the rate of return for the experiment. However, participants in the 25% and 75% condition had relatively suboptimal performance, due to them not adjusting their responses according to target distributions. These results demonstrate mixed support for applying the MVT to the domain of human visual searches. However, even though this study does lend some support to the feasibility of applying the MVT to human visual searches, there is little further literature to support these findings. Not much literature exists on the use of the MVT by human decision-makers. As previously discussed, GUTs and the MVT relate to speed-accuracy trade-off optimisation. Consequently, it is not clear if the performance of human decision-makers is explicitly driven by the optimisation of their speed-accuracy trade-off in visual searches. However, other researchers have found evidence for a more general use of trade-off optimisation in human decision-makers.

For example, Navalpakkam et al. (2010) presented participants with a maximisation problem. Participants were presented with a series of stimuli with several distractors and two targets whose value and saliency were systematically varied. Participants were then instructed to identify the target stimuli as rapidly as possibly on each trial. Navalpakkam et al. (2010) wanted to assess how participants combined stimulus information to maximise the expected reward on each trial. Results showed that participants behave according to an ideal Bayesian observer who combines both (value and saliency) factors of the stimuli, to maximize the expected reward on each trial. As such, participants' RTs for stimulus recognition were optimal or near optimal for maximising expected rewards. In conjunction with OFT, this demonstrates the successful application of the optimal decision rule in the human-perceptual domain and therefore the successful optimisation of the speed-accuracy trade-off.

Similarly, Navalpakkam, Koch and Perona (2009) tasked participants with searching and identifying as rapidly as possible the presence of a target object in a cluttered scene containing several distractor stimuli. Participants were rewarded according to task performance. Results showed that a significant proportion of participants displayed optimal or near optimal decision-making strategies. Specifically, participants' search patterns and RTs per trial were near optimal. Consequently, participants' expected rewards per trial were at or near their theoretical optimum. Similar findings were observed in a variant of the experiment involving perceptual and perceptuo-motor decision-making tasks (Trommershauser, Maloney & Landy, 2004; Whiteley & Sahani, 2008).

These findings support the argument that human-decision makers are capable of finding theoretically optimal speed-accuracy trade-offs in visual search tasks. It appears as though human decision-makers are able to optimise their speed-accuracy trade-off in visual search tasks. However, it is currently unclear as to what precise features of the stimuli or the participant are controlling the setting of the speed-accuracy trade-off. Results by Nowakowska et al. (2017) suggest that trade-off manipulation was under the direction of participants seeking getter accuracy by increasing response conservativeness. However, research by Bogacz, Brown, Moehlis, Holmes and Cohen (2006) provided an analysis of optimal decision-making behaviour in two-alternative forced-choice (TAFC) tasks using an EAM, to determine if the rate of information processing was critical to maximising a reward index. Bogacz et al. (2006) elaborated on the Pure Drift Diffusion EAM (pDDM) and its relation to decision-making in TAFC tasks. Like all EAMs, the pDDM assumes that individuals accumulate evidence in support of choosing one of two choice alternatives over a period of time. The choice that is the first to receive enough evidence to satisfy the response threshold is the response that will be chosen.

Bogacz et al. (2006) further investigated how the reward index for a given TAFC task trial would vary as a function of A: the rate at which information is gathered about the stimuli, called drift rate. Results showed that reward rates depended on drift rate non-linearly. The stimulus presented on each trial for the TAFC task gives no information if $A = 0$. In this case, the optimal decision strategy is to make a response immediately, as prolonging response times will yield no greater information from the stimulus. As the drift rate increases and more information becomes available, it becomes more advantageous to integrate this information to decide when to respond. In this instance, longer response times would be of greater benefit, as they would provide more relevant information to increase the probability of making a correct response. However, as the drift rate increases it becomes more probable that the information gathered is contaminated. This is because higher drift rates result in more information about the stimuli being gathered and processed. Although this gathered information is in favour of one of the two task responses, the increase in the total volume of information gathered can overload the processing system and lead to errors in understanding all this information and relating it together. Therefore, RTs at this stage or beyond decrease the probability of a correct response and therefore overall returns. Refer to Figure 5.1 below for an illustration of how the reward index varies according to drift rate.

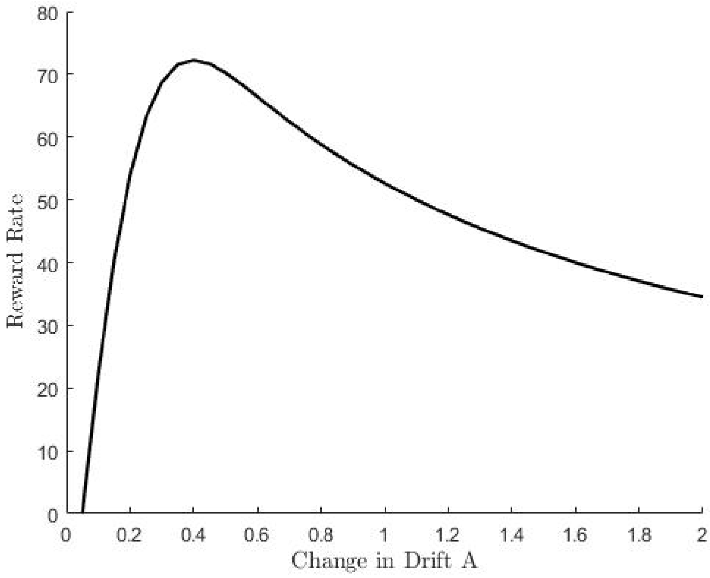

Figure 5.1. The curve shows the non-linear relationship between the arbitrarily scaled variables of reward rate and changes in drift A.

This non-linear relationship has also been observed between mean reward and participants' RTs on a task. For instance, Jarvstad, Rushton and Warren (2012) tested an experimental paradigm that captured individuals' ability to behave optimally in simple decision-making tasks. Jarvstad et al. (2012) investigated whether a perception-cognition gap could be observed in timing decisions. That is, whether the optimal decision strategies used by participants mainly in lower level perceptual-motor tasks and not higher-level cognitive tasks applies to timed decisions. The researchers used a two-stage experimental paradigm that allowed them to assess how the relationship between accuracy and time affects overall task reward. To do this, two consecutive experimental stages were used: the assessment and decision-making stage. On each trial of the assessment stage participants were presented with a descriptive scenario such as: a hunter is tracking prey through a forest. In his path lies a pond. Would it take him longer to pass the pond by going to the left or to the right? All presented scenarios were altered versions of this original scenario and all had two outcomes: success or failure in choosing the right path. During this stage participants were imposed six different RT windows in which they could give a response. Varying sequences of tones were used to indicate to participants the length of the time window in which a response has to be given. Results showed that during this stage of the experiment, accuracy on the task varied as a function of participants' RT. That is, the more time participants spent on the task, the greater the probability of them giving a correct response. However, beyond a certain amount of time, more time spent assessing the task did not increase accuracy. This is illustrated by a sigmoidal function of accuracy against RT, see Figure 5.2 below.

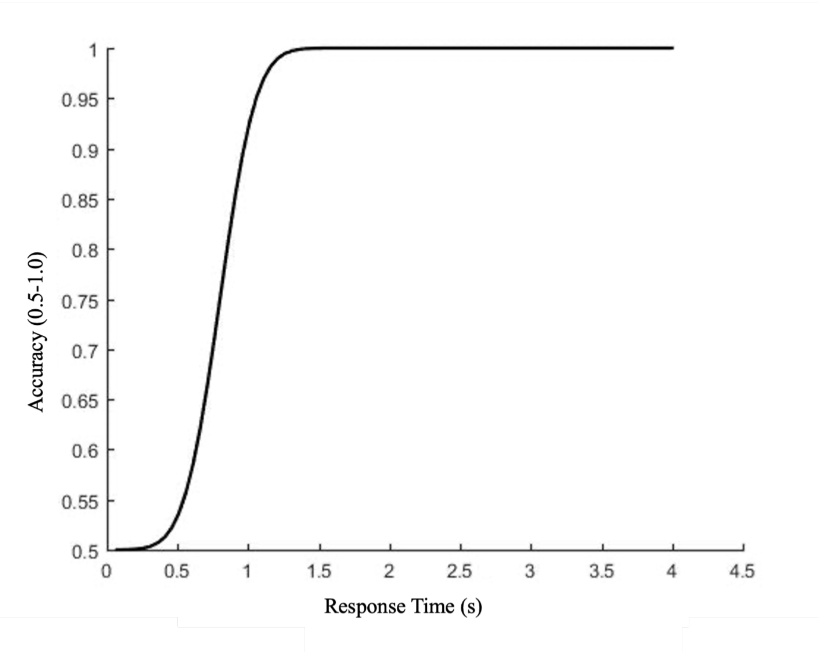

Figure 5.2. The relationship between RT and accuracy.

During the decision-making stage participants had a fixed time window in which to complete as many individual trials of a mental arithmetic task as they saw fit. Consequently, quicker response times yielded more trials during this stage. During each trial of this stage of the experiment, participants were shown two positive integers. Both integers were chosen so as to sum up to an integer within the range of 90-110, excluding 100. On each trial participants were tasked with determining as rapidly as possible, whether the two integers summed up to less or more than 100. Furthermore, a neutral reward structure was used for a correct response on a trial (reward = +1 point) and an incorrect response on a trial (penalty = -1 point). Overall rewards were then multiplied by the participants' overall performance during the second stage of the experiment. Jarvstad et al. (2012) then determined how well participants performed on the task for a given RT: efficiency function. Refer to Figure 5.3 below. The efficiency function for each participant was derived by taking into account the participant's time-accuracy function (Figure 5.2) and the reward structure for each trial.

Jarvstad et al. (2012) further examined whether participants' choices related to their efficiency functions (Figure 5.3). As efficiency functions are a method of determining which response time for a given participant maximises expected returns for the experiment, they can be used to identify optimal decision strategies. In other words, the researchers attempted to assess if participants were able to optimise their speed-accuracy trade-offs when different components of the trade-off were emphasised. Results showed that the majority of participants had used optimal or near optimal decision strategies. That is, participants were largely able to identify average RTs during the decision-making stage which maximised their expected returns per trial and therefore overall returns. This can equally be interpreted in terms of participants applying the MVT. Participants gave responses at specific times on each trial when the average rate of reward for that trial at a given response time equalled the rate of reward for the entire task. What these results reveal is that tasks involving a speed-accuracy trade-off have instances where performance can coincide with more optimal trade-offs. From an EAM perspective, the optimisation of this trade-off assumed to control task performance should also coincide with increased performance.

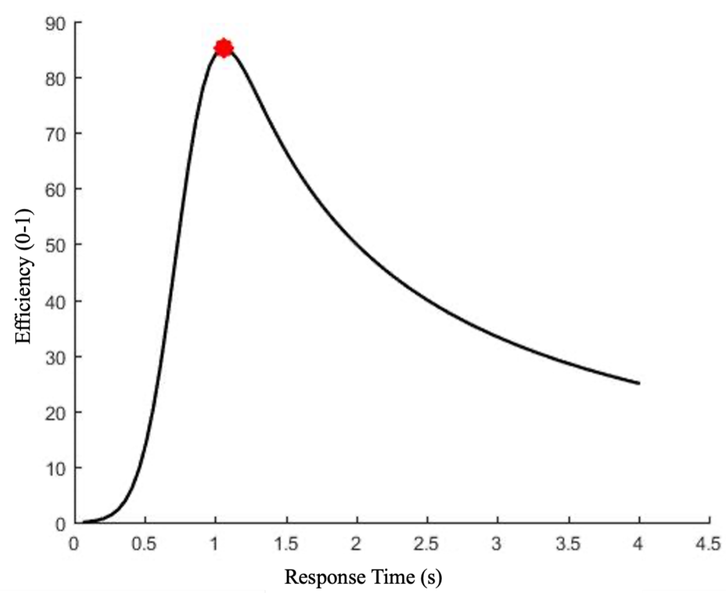

Figure 5.3. The above graph shows how the overall task performance for a given participant varies with response time. The red star indicates the response time associated with optimising overall performance on average.

In the context of foraging behaviour, a comprehensive EAM perspective provides a detailed account of performance drivers while providing tractable parameters. The perspective assumes that performance on foraging tasks is driven explicitly by changes in the speed-accuracy trade-off. Performance is not driven by changes in search strategy. Furthermore, increased performance should coincide with more optimal speed-accuracy trade-offs. These assumptions are testable EAM predictions and the main objective of this chapter. In order to assess this, I presented participants with a multi-patch foraging task. Participants were instructed to perform the task in two parts: during the first part of the experiment participants had to focus on accuracy and in the second part participants had to focus on speed. The aim was to produce a paradigm that allowed for a clear distinction in behaviour between the two parts of the speed-accuracy trade-off. The changes in participants' behaviour could then be studied in relation to the various EAM features. That is, if the EAM response threshold parameter drives the speed-accuracy trade-off in this task, changes in this parameter should be represented in accuracy and RT performance as participants focus on different aspects of the trade-off. Furthermore, if the task is based on an intra-task design with response feedback, regardless of the trade-off, there ought to be no dominate search strategy (Boot et al., 2006). In which case, changes in the underlying speed-accuracy trade-off should be more clearly reflected in performance. To clarify, the hypothesis is that performance on the foraging task will be driven by explicit changes in the speed-accuracy trade-off and not by search strategies.

5.1.1 Method

Participants

In total, 6 participants were recruited to take part in the experiment. The participants were recruited from the online participation platform Prolific and were paid approximately £15 each for their participation. A total of 6 participants were recruit, as preliminary analyses revealed that the experimental paradigm had fundamentally failed to functioned as intended and therefore did not merit further data collection. This is further discussed later in the chapter.

Design and procedure

The experiment was based on a within participants design with three factors: the number of foraging patches in a trial (V1), location of the target acorn (V2) and the speed-accuracy focus of the block. V1 had three levels: single patch, three patch and five patch trials. V2 had two levels: absent or present in single patch trials and one of three or five locations in multi-patch trials. V3 had two levels: at the start of the block participants were asked to either focus on responding as quickly or as accurately as possible. All factors were measured for their effect on search strategy, the amount of time participants spent searching the different patches and accuracy.

All trials commence with each foraging location identified by a 200px x 200px image of a tree. To begin foraging, participants must move their cursor over the image of the tree. This will then reveal a 4x6 grid of 25px x 25px multi-coloured leaves. If participants move the cursor beyond the boarder of the initial tree image, the grid of leaves image will be replaced by the tree image. On single patch trials participants are presented with a single image of a tree, refer to Figure 5.4 A and B. Participants must then identify whether or not there is a 25px x 25px image of an acorn amongst the leaves, refer to Figure 5.4 C. Participants indicated their response by either clicking the "Present" or "Absent" button onscreen. For three and five patch trials, foraging patches are presented in an evenly spaced semi-circular configuration, refer to Figure 5.4 D and E. For three and five patch trials, an acorn is always present in one of the foraging patches. For these trials participants provide a response by clicking the button onscreen containing the same label as the foraging patch they believe to contain the acorn.

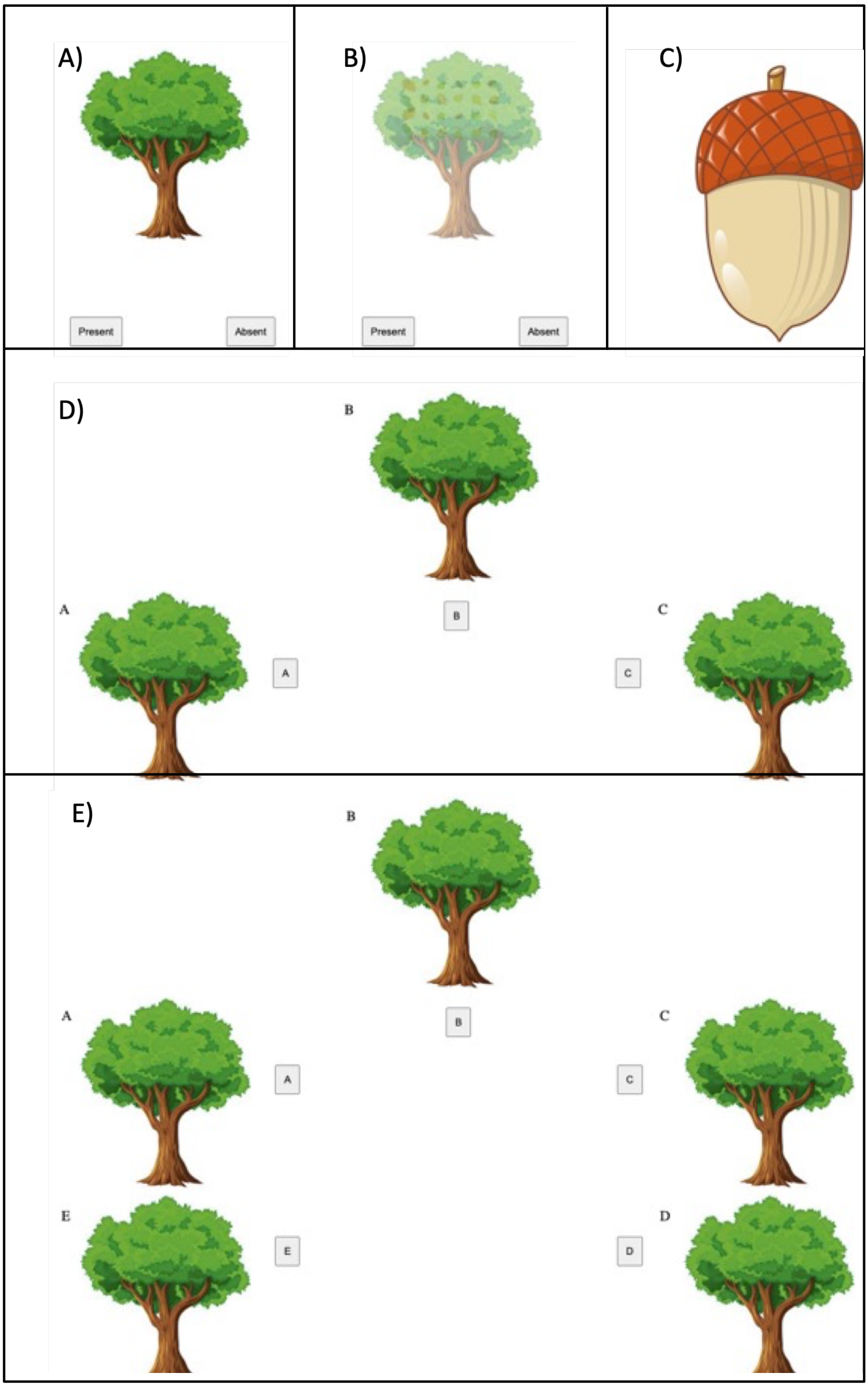

Figure 5.4. A) single patch trial. B) single patch trial revealing the hidden grid of leaves image with a low opacity level. C) enlarged image of the target acorn. D) three patch trial. E) five patch trial.

The experiment was presented in a fixed sequence of three consecutive blocks. The first block included the calibration stage of the experiment. During this block participants were shown an indefinite number of trials containing a single foraging patch and given 4 seconds to provide a response before the trial ended. The initial calibration block trial had both the image of the tree and the underlying grid of leaves image displayed with maximum opacity: opacity was initialised at 1 and had a range of 1-0 (1 = completely opaque, 0 = fully transparent). If participants provided an incorrect response, the opacity of the underlying grid of leaves image and the acorn was increased by 0.1. If participants provided a correct response, the opacity of the underlying grid of leaves image and the acorn was decreased by 0.1. If participants took more than 4 seconds to respond in a trial, an incorrect response was recorded and opacity increased by 0.1. If the current opacity level reached blow 0.5 and participants provided three separate responses on consecutive trials identified as incorrect-correct-incorrect, the calibration block ended and the opacity value of the underlying grid of leaves and acorn image for the remainder of the experiment was set to the value of the last calibration trial.

In the subsequent first experimental block, 15 single patch trials were shown with an acorn present and 15 trials with an acorn absent. Each foraging location on three patch trials was shown 10 times with the acorn present. Each foraging location on five patch trials was shown 6 times with the acorn present. This yielded 30 single patch trials, 30 three patch trials and 30 five patch trials, with a total of 90 trials for the first experimental block. This block, along with the second experimental block had no response time constraints. Single, three and five patch trials were randomly presented. During this block of trials participants were instructed to respond as accurately as possible. The second experimental block was identical to the first experimental block. During this block of trials participants were instructed to respond as quickly as possible. Participants were also shown their RT after each trial. Therefore, the experiment had 180 trials in total.

5.1.2 Results

Experimental results

The results show that for all participants no singular search strategy dominated. A strategy is defined as the order in which a search is performed. In relation to the present task, this corresponds to the order in which the different patches were searched: e.g. A then B and then C, and A again. It is important for the reader to note that on three and five patch trials there were an infinite number of ways for participants to search the different patches in respect to order and duration. Specifically, the number of times the most used search strategy was used as a proportion of the total number of times all the different search strategies were used across each block and trial type was never greater than 50% for each Participant. Please refer to Table 5.1 below. This indicates that the most used search strategy was used relatively infrequently. As such, these results appear to confirm that participants' foraging behaviour was not explicitly driven by a specific search strategy.

Table 5.1. The number of times the most used search strategy was used by all participants. Each row represents a separate participant.

Block 1 – 3 Patch	Block 1 – 5 Patch	Block 2 – 3 Patch	Block 2 – 5 Patch
10%	10%	23%	20%
13%	13%	16%	10%
27%	20%	24%	13%
23%	17%	14%	13%
17%	10%	23%	14%
30%	17%	30%	20%

Table 5.2 below shows summary statistics for STs for all participants across the different trial conditions.

Table 5.2. Average STs for all three trial types across the two blocks for each participant.

Participant	Block 1 - 1 Patch	Block 1 - 3 Patch	Block 1 - 5 Patch	Block 2 - 1 Patch	Block 2 - 3 Patch	Block 2 - 5 Patch
P1	2.15s	8.02s	10.46s	1.59s	4.41s	6.70s
P2	3.19s	7.22s	7.91s	2.89s	8.38s	8.17s
P3	2.93s	6.97s	7.76s	1.97s	8.04s	7.78s
P4	2.41s	5.71s	10.39s	2.27s	7.94s	7.11s
P5	2.87s	10.48s	8.38s	2.04s	5.17s	8.91s
P6	2.62s	8.38s	9.11s	2.51s	9.65s	10.6s

Table 5.3 shows summary statistics for accuracy rates for all participants across the different trial conditions.

Table 5.3. Accuracy rates for all three trial types across the two blocks for each participant.

Participant	Block 1 - 1 Patch	Block 1 - 3 Patch	Block 1 - 5 Patch	Block 2 - 1 Patch	Block 2 - 3 Patch	Block 2 - 5 Patch
P1	70%	77%	77%	80%	77%	60%
P2	70%	87%	53%	87%	77%	70%
P3	80%	97%	80%	90%	90%	53%
P4	73%	60%	37%	63%	63%	37%
P5	100%	77%	80%	100%	97%	77%
P6	93%	97%	97%	93%	100%	100%

Next, we wanted to assess whether instructing participants to emphasise different parts of the speed-accuracy trade-off in the different blocks produced observable differences in STs and accuracy. As the data was not normally distributed we performed a 2(block) x 3(trial types) Friedman Test to determine if there was a significant difference in STs across trial types between blocks. The results showed that there was a non-significant effect of trial type and block on overall STs. We additionally performed a 2(block) x 3(trial types) Friedman Test to assess if there was a significant difference in overall accuracy across trial types between blocks. The results showed that there was a non-significant effect of trial type and block on accuracy. Methodologically, results suggest that the experimental paradigm failed to produce the desired behavioural results. Specifically, there was no evidence confirming the expectation that participants would emphasise accuracy over speed in block 1 and speed over accuracy in block 2. Furthermore, the inferential analyses show no significant difference in behavioural responses between blocks.

I believe that the lack of a significant difference in STs between three and five patch trials may be due to a lack of clarity in the initial instructions for the experiment. As discussed later in the chapter, it became clear that participants were not following the instructions and just followed idiosyncratic aspects of a speed-accuracy trade-off. That is, participants were instructed to emphasize speed in the first block and accuracy in the second block. Instead, the results indicate that participants were using a uniform and non-specific strategy throughout the experiment. As participants were not following the task instructions, individual participants may have fixed their speed-accuracy trade-off across trial types and blocks (i.e., followed an idiosyncratic strategy), resulting in similar performance across trials and blocks.

At this stage it became clear that the averaged dataset was not producing the desired behavioural responses. Therefore, I wanted to assess whether the behavioural responses for each participant in each block reflected a difference in the underlying information processing architecture, that possibly resulted in similar behavioural responses in both blocks. It may be the case that although the observed behavioural responses were not significantly different across the blocks, they may be significantly different in terms of the underlying parameters used in the information processing of the stimuli in each block. In order to assess this, an individual level modelling approach was used given the small size of the dataset.

Modelling results

The present experimental results indicated that the current experimental paradigm did not merit the collection of additional data. This resulted in a relatively small dataset that provided insufficient data to produce satisfactory model fits to the data for standard EAM fitting procedures. Additionally, a closed form expression for the likelihoods could not be derived. Therefore, standard model fitting procedures were not applicable to our dataset. The subsequently defined model was therefore fitted to the data by estimating parameter values for the model that would produce datasets that most closely resembled the real data.

An EAM was used to access the underlying information processing architecture of the present task. This was because EAM parameters easily capture the speed-accuracy trade-off of tasks and require relatively less computational power to do so. The precise model used was the Random Walk Drift Diffusion Model (RDDM). The drift diffusion model belongs to the same class of models as the previously discussed LBA model in the previous chapter. The primary difference between the two models is that the trajectory for the evidence accumulator is not linear, as it is in the LBA model. Instead, the evidence accumulator alternates its trajectory between two separately defined response thresholds in a random manner, until the accumulator reaches one of the two response thresholds to trigger a response. As the response thresholds are separately defined in a RDDM, they better facilitate computational modelling where the goal is to separately manipulate two response thresholds in a model. This is contrary to the LBA, where the two response thresholds are defined as being at an equal distance from each other (Brown & Heathcote, 2008).

In this experiment, it is assumed that the different response thresholds in the RDDM represent the different response thresholds for the acorn is present and the acorn is absent responses. It is assumed that the response threshold for the acorn being absent in any given patch is higher than the response threshold for the acorn being present in any given patch. The assumption here is that uncertainty is greater for a decision regarding whether an acorn is absent within a particular patch, than deciding if an acorn is present in a particular patch. Consequently, this uncertainty increases the amount of accumulated evidence required to trigger the response that the acorn is absent. Additionally, it is assumed that the response thresholds for the acorn being absent or present in block 2 of the experiment are both lower than in block 1 of the experiment. The assumption here is that the emphasise placed on the speed component of the speed-accuracy trade-off in block 2, causes participants to forego the need to accumulate higher levels of evidence in support of a particular response in order to provide a quicker response. This therefore lowers the response thresholds in block 2 of the experiment relative to block 1. In effect, on three and five patch trials it is assumed that participants initially search each patch once, with the duration of the search and accuracy defined by the model parameters. The first patch searched on a trial is assumed to be random. If the target is believed to be present in a patch, the trial ends, and accuracy and RT is determined. If the target is believed to be absent in a patch, the participant moves onto another yet unsearched patch. If all patches on a trial are searched and believed to not contain the target, another randomly chosen patch is searched until the target is believed to be present in a patch. Refer to Figure 5.5 below for an illustration of the RDDM logic for each trial type.

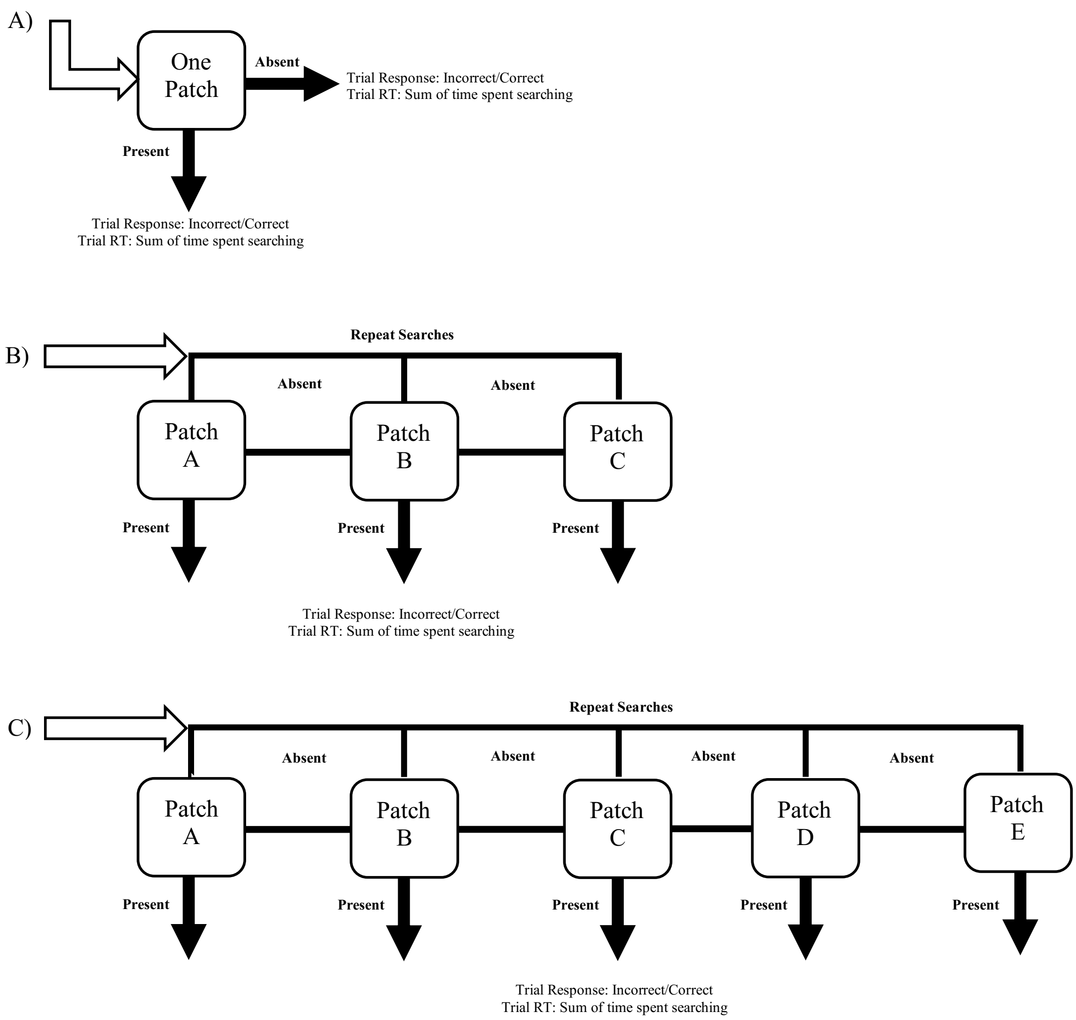

Figure 5.5. Illustrates the logic for the RDDM algorithms used to determine trial responses and RTs for all the trial types. A) Displays the logic scheme for one patch trials. B) Displays the logic scheme for three patch trials. C) Displays the logic scheme for five patch trials.

The simulation model algorithm was created with 8 parameters: a response threshold for when the acorn is absent and when the acorn is present response across both blocks; a separate drift rate for when the acorn is present and when the acorn is absent; a sampling noise parameter; a non-decision time parameter.

As previously mentioned, due to the relatively small dataset collected for this experiment, the standard model fitting procedure could not be applied in this instance. Additionally, a closed form expression for the likelihood functions could not be derived. As such, an estimation procedure was used to fit our RDDM to each participant individually. Specifically, this procedure estimates the best fitting parameters for the RDDM model for each participant. The estimation procedure used was the Approximate Bayesian Computation (ABC) method.

ABC is fundamentally based on a simple rejection algorithm, comprising a series of steps (Csilléry, Blum, Gaggiotti, & François, 2010; Csilléry, François, & Blum, 2012; Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010):

1) summary statistics are produced for the collected dataset,
2) a simulation model is created and multiple iterations are ran with starting parameters initially picked from a uniform distribution,
3) the simulation results for each iteration are turned into summary statistics equal to those produced for the collected dataset,
4) if the difference between the summary statistics of the collected dataset and simulated dataset for a particular iteration is within a rejection threshold, the simulation model parameters that produced the simulation summary statistics for that specific iteration are kept. The other simulation model parameters are rejected,
5) the simulation parameters kept for each model parameter are then transformed into new parameter distributions,
6) the process is repeated from step 2 with parameters now picked from the new parameter distributions and with a lower rejection threshold.

The primary function of ABC is to update the prior information available on the parameter distributions of the simulation model, by refining it through the rejection algorithm. The aim is to obtain a more narrowly defined posterior distribution range for the model parameters that better estimate the true parameter values (Csilléry et al., 2010).

The summary statistics used for step 1 were five RT quantiles (10%, 30%, 50%, 70% and 90%) for both correct and incorrect responses across all conditions. For the present experiment step 2 was ran with 500 iterations and a set of uniform parameter distributions were defined based on trial-and-error. The difference between the real data and the simulated data summary statistics in step 4 was calculated as the square root of the squared difference between the collected and simulated summary statistics. The rejection threshold used in step 4, defined as ε, ranged from 7.5-0.5 and decreased in intervals of 0.5. This gave a total of 15 resampling steps for the ABC algorithm used. In step 5 the new parameter distributions were created through a Gaussian kernel transformation. The specific ABC estimation procedure used in this modelling exercise was a subset of the rejection algorithm: Partial Rejection Control (PRC). The ABC-PRC algorithm is similar to the ABC procedure outlined above, expect that step 5 involves resampling and weighting the parameter distributions (Sisson, Fan, & Tanaka, 2007, 2009).

The parameter values selected from the ABC posterior distributions for the model parameters were based on the mean value of these distributions. Refer to Figure 5.6 below for a sample illustration of the posterior distributions produced for the various model parameters by the ABC fitting procedure.

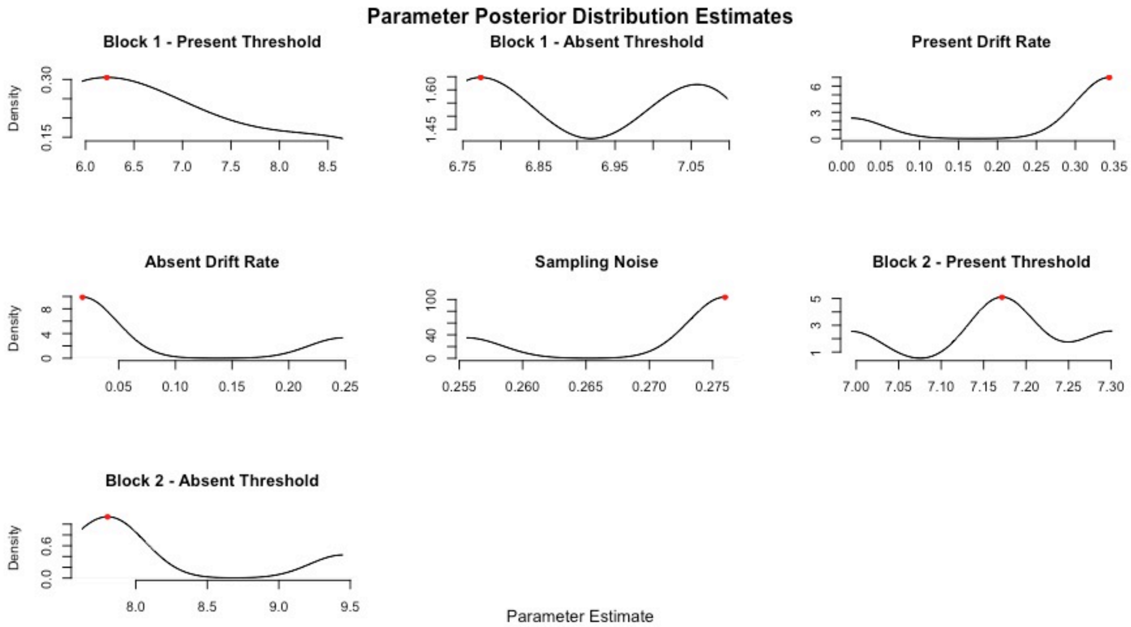

Figure 5.6. An illustration of the posterior distribution parameter estimates produced for the 7 model parameters, for one participant. The red circle on each distribution corresponds to its point of highest density.

Table 5.4 provides summary statistics for the estimated response thresholds.

Table 5.4. The mean estimated values for the parameters that varied between target present and absent search patches: target present, target absent response thresholds, target present drift rate and target absent drift rate, alongside HDIs from the model fitting procedure.

Parameters	Mean Estimates	HDIs
Block 1 - Present:	6.05	2.17 - 8.55
Block 1 - Absent:	7.45	2.45 - 9.06
Block 2 - Present:	6.16	2.47 - 8.64
Block 2 - Absent:	7.80	2.15 - 8.79
Drift - Present:	0.11	0.09 - 0.80
Drift - Absent:	0.04	0.04 - 0.81

In order to assess whether changes in the speed-accuracy trade-off between blocks were determined by changes in participants' response thresholds across the blocks, we performed a 2(blocks: 1 & 2) x 2(response threshold: present & absent) Friedman Test on the data as it was not normally distributed. This was done to assess whether there was a significant difference between present and absent response thresholds across blocks. As the present assumption is that the speed-accuracy trade-off is determined by changes in response thresholds and not changes in the rate of information processing, previous findings that participants' behaviour did not significantly differ across blocks should be reflected in response thresholds. That is, if changes in response thresholds determine the speed-accuracy trade-off and inferential results show no indication of a speed-accuracy trade-off being implemented, there should therefore not be a significant difference between response thresholds across blocks. Results showed that there was a non-significant main effect of block and response threshold. These results are therefore consistent with the findings from the experimental results that a speed-accuracy trade-off was not successfully implemented methodologically.

5.1.3 Discussion

The experimental results show that a speed-accuracy trade-off paradigm was not successfully implemented in this experiment. Conclusions on whether changes in the speed-accuracy trade-off drive foraging behaviour could not be made. However, descriptive results do support the assumption that search strategy alone does not drive foraging behaviour. Results showed that all participants utilised numerous search strategies throughout the experiment. For example, for any given block and trial type, the most used search strategy never represented more than

30% of all search strategies used by a participant. Although no significant behavioural differences were observed between blocks, it remained to be seen if there occurred significant differences between participants' response conservativeness between the experimental blocks. Participants' response conservativeness (speed-accuracy trade-off) was modelled as a participant's response threshold using a series of RDDMs and fitted to the experimental data using ABC. The results showed that there was no significant difference in participants' response thresholds between blocks.

The non-significant experimental and modelling results found are as expected. As if the experimental paradigm failed to implement a speed-accuracy trade-off, non-significant results are to be expected. However, I do not believe this extends to the descriptive results on search strategy. In conjunction with the results found by Boot et al. (2009), the intra-task nature of the present experiment did not produce a convergence of search strategies. These results are to be expected if foraging behaviour is assumed to be driven by the manipulation of the speed-accuracy trade-off (modelled as response conservativeness). Even though the present experiment showed no significant difference in response thresholds between blocks or trial types, a convergence of search strategies did not occur to attempt to drive behaviour in this task. As such, the similarity in behaviour between both blocks of the experiment could have been driven by a similarity in participants' response conservativeness between both blocks. Therefore, it is assumed that a significant difference in participants' response thresholds between blocks would consequently represent a significant difference in behaviour between blocks.

Methodologically, the implementation of a speed-accuracy trade-off was based on instructions. The experimental paradigm depended on participants strictly following the presented instructions to empathise a different aspect of the speed-accuracy trade-off in different blocks. Although these instructions were intended to be carefully followed by participants, there was no obligation or incentive for participants to do so. As such, a second experimental paradigm is proposed, whereby the implementation of the speed-accuracy trade-off is not dependant on the properties of the task, but on the inherent properties of the participants.

The literature has robustly shown that the speed-accuracy trade-off is emphasised in different ways by different age groups (Berchicci, Lucci, Pesce, Spinelli, & Di Russo, 2012; Tiedemann, Sherrington, & Lord, 2007; Zaal & Thelen, 2005). For example, young participants have been found to emphasise the speed component and more senior participants found to emphasise the accuracy component of the speed-accuracy trade-off (Berchicci et al., 2012). Ratcliff, Thapar, and McKoon (2004) applied the drift diffusion EAM to the data generated by younger college age participants and older (60-75) participants on a recognition memory task. The results found that older participants were characterised by higher response conservativeness (trading accuracy for speed) compared to the younger participants. Furthermore, these inherent characteristics of participants of different age groups have been found across a variety of different RT tasks (Zaal & Thelen, 2005). Another example of this effect is found in the Trial Making Tests. These tests are a neuropsychological instrument used as a method for identifying neurological and cognitive decline. They involve individuals connecting a series of approximately 25 enclosed numbers or letters in numerical or alphabetical order (Wagner, Helmreich, Dahmen, Lieb, & Tadi, 2011). Clinical results have found that older samples have increased competition times relative to older participants, even in the absence of motor, sensory or non-age related cognitive deficits (Bowie & Harvey, 2006; Wagner et al., 2011). This again is indicative of how different age groups intrinsically emphasise different aspects of the speed-accuracy trade-off on tasks.

The proposed second experimental paradigm will be identical to the first, except that the experiment will be ran using a mixed method and there will be no explicit speed-accuracy trade-off emphasis. Instead, three separate groups of participants will be tested, categorised by age group: 18-30, 35-45 and 50-65. The three groups will represent young, middle aged and senior participant groups. The assumption is that the younger age group (18-30) will inherently emphasise the speed component of the speed-accuracy trade-off and therefore have lower response conservativeness. The senior group (50-65) will inherently emphasise the accuracy component of the speed-accuracy trade-off and will therefore have higher response conservativeness. The middle-aged group will represent a balance between the two groups. As such the proposed associated differences in response conservativeness associated with the different components of the speed-accuracy trade-off will be most salient between the young and senior group.

The aim of the second experiment remains identical to the first, except that the speed-accuracy trade-off will now be implemented through the inherent characteristics of the different participant groups.

Section 5.2 Manipulating the Speed-Accuracy Trade-off Through Different Age Groups

Introduction

The purpose of the second experiment was to access whether the inherently different speed-accuracy trade-offs (response conservativeness) in the different age cohorts, is responsible for determining behaviour in foraging tasks. As such, the present experiment removed the speed-accuracy instructions between blocks and recruited participants from three distinct age groups. To clarify, the hypothesis is that performance on the foraging task will be driven by explicit changes in the speed-accuracy trade-off and not by search strategies.

5.2.1 Method

Participants

In total, 120 participants were recruited to take part in the experiment. 40 participants were recruited from three separate age groups: 18-30 (young), 35-45 (middle aged) and 55-65 (senior). All participants were recruited from the online participation platform Prolific. Each participant was paid approximately £7 for their participation.

Design and procedure

The experimental design was identical to the first experiment. However, the present experiment did not instruct participants to focus on a different component of the speed-accuracy trade-off on the different blocks of trials. Therefore, as there were three age groups performing the same task in the experiment, a mixed design was used.

5.2.2 Results

Experimental results

The descriptive results show that accuracy rates were largely similar for the young and senior age groups, with the senior age group performing slightly better across the different trial types. The middle age group had the highest accuracy rates of all the groups, across all three trial types. Additionally, accuracy rates were similar for three and five patch trials and only appeared to differ compared to one patch trials which had higher overall accuracy, refer to Figure 5.7 below.

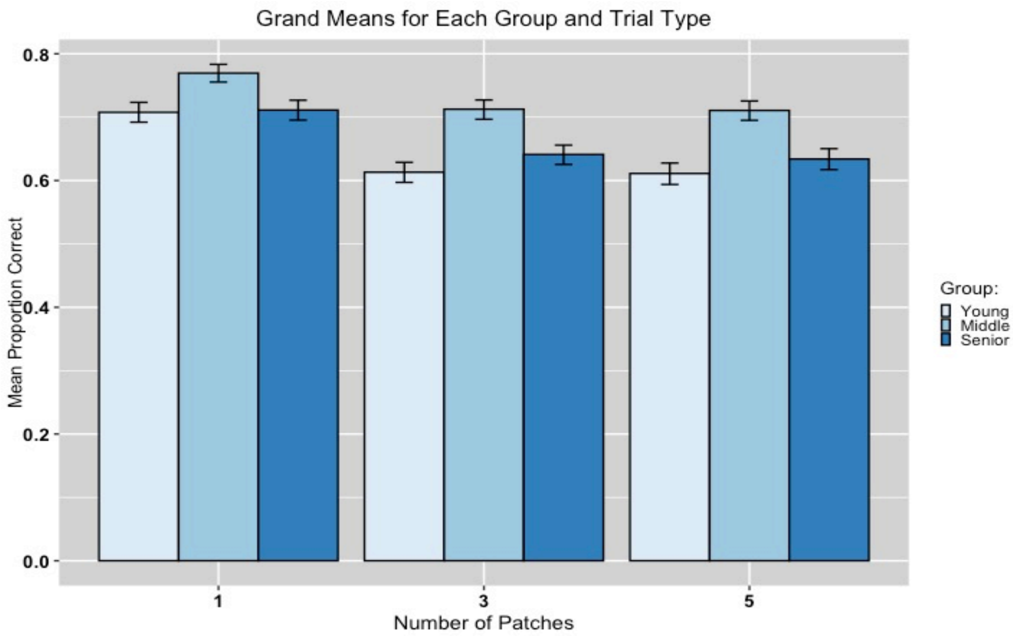

Figure 5.7. Accuracy rates for the three age groups across the three different trial types, with bootstrapped 95% confidence intervals.

Trial STs were defined as the total amount of time spent searching patches only. Trial ST distributions also showed that median STs increased incrementally with the number of patches in a trial. The young age group displayed the lowest median STs across the different trial types, with the senior age group displaying the highest median STs, refer to Figure 5.8 below.

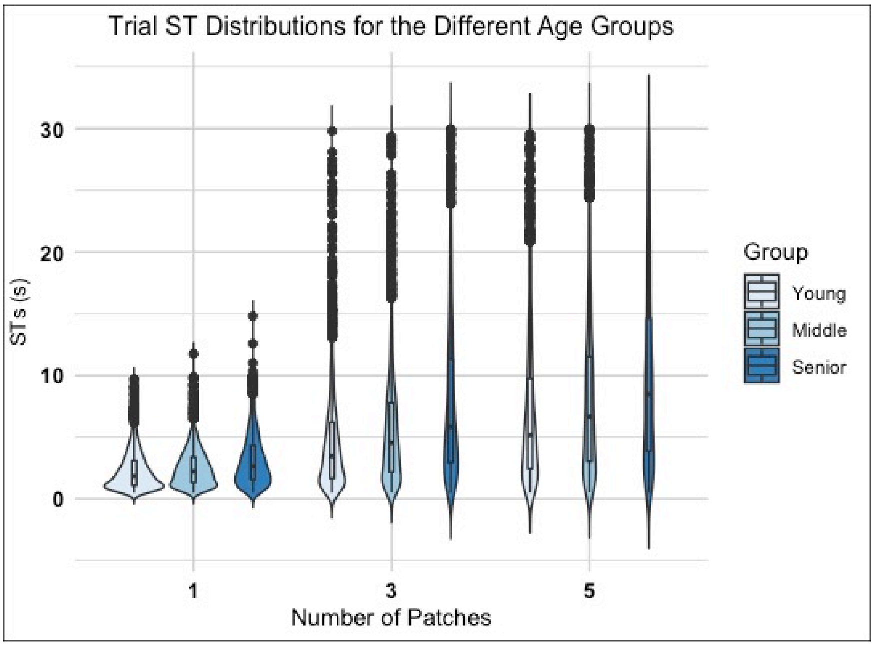

Figure 5.8. Violin plots of the ST distributions for the three age groups across the different trial types.

In this experiment we hypothesised that different age groups will inherently emphasise different aspects of the speed-accuracy trade-off. These preliminary descriptive results appear to support this. For example, the young age group consistently displayed the lowest median ST and corresponding accuracy rates across all three trial types. This corresponds with the young age group emphasising the speed component of the speed-accuracy trade-off. In other words, they were trading speed for accuracy in the task. At the opposite end, the senior age group had the highest median ST across all three trial types. However, although accuracy rates are above the young age group, they are below the middle age group. These results imply that relative to the young age group, the senior age group were emphasising the accuracy component of the speed-accuracy trade-off. In other words, they were trading accuracy for speed in the task. However, when comparing the senior and the young age groups, it is clear that both groups sub-optimised their speed accuracy trade-off relative to the middle-aged group. The middle-aged group consistently had higher accuracy rates compared to the other groups, while having median STs in-between the range of the young and senior age groups. These results suggest that rather than emphasising a particular component of the speed-accuracy trade-off, the middle age group were attempting to optimise this trade-off to yield the best performance.

To further explore the difference in performance between the middle and other age groups, we analysed the differences in the types of correct and incorrect responses. In the present experiment, there were only two ways of correctly identifying the target acorn. For example, either participants correctly identified the target acorn the first time it was revealed to them (termed first-go accuracies) or participants correctly identified the target acorn after the target acorn was revealed to them more than once (termed repeat accuracies). Refer to Figure 5.9 below.

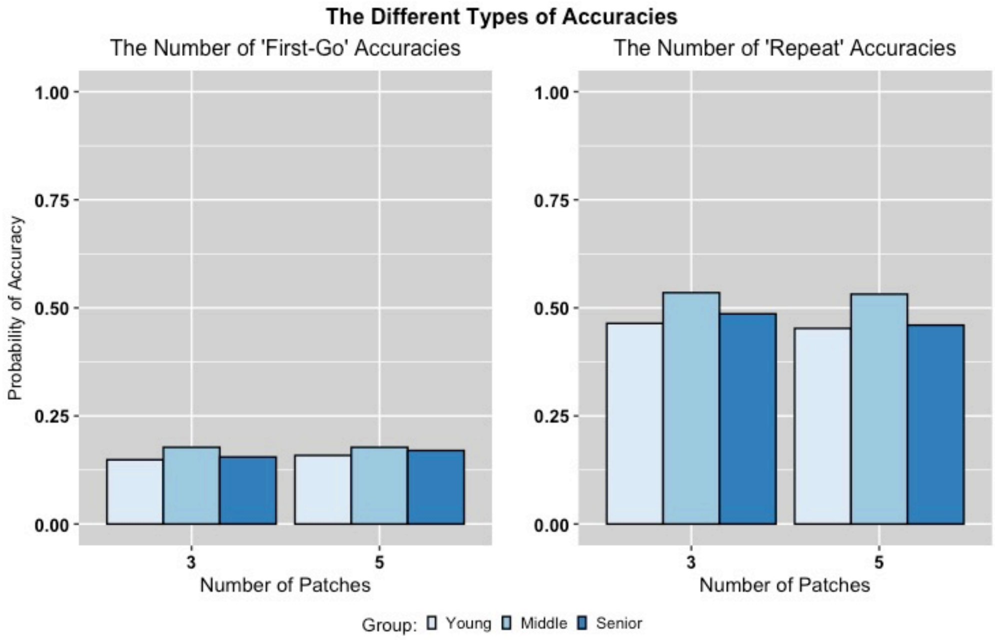

Figure 5.9. The left figure shows the proportion of correct responses that resulted from participants correctly identifying the acorn on its first appearance. The right figure shows the proportion of correct responses that resulted from participants correctly identifying the acorn after more than one appearance.

Additionally, in the present experiment there were only two types of target identification errors and although both ultimately resulted in incorrectly identifying the target acorn in the correct patch, this misidentification resulted from two types of search errors. Either participants did not search the patch containing the acorn (termed skipped errors) or participants identified the acorn as being in the incorrect patch (termed missed errors). Please refer to Figure 5.10 below.

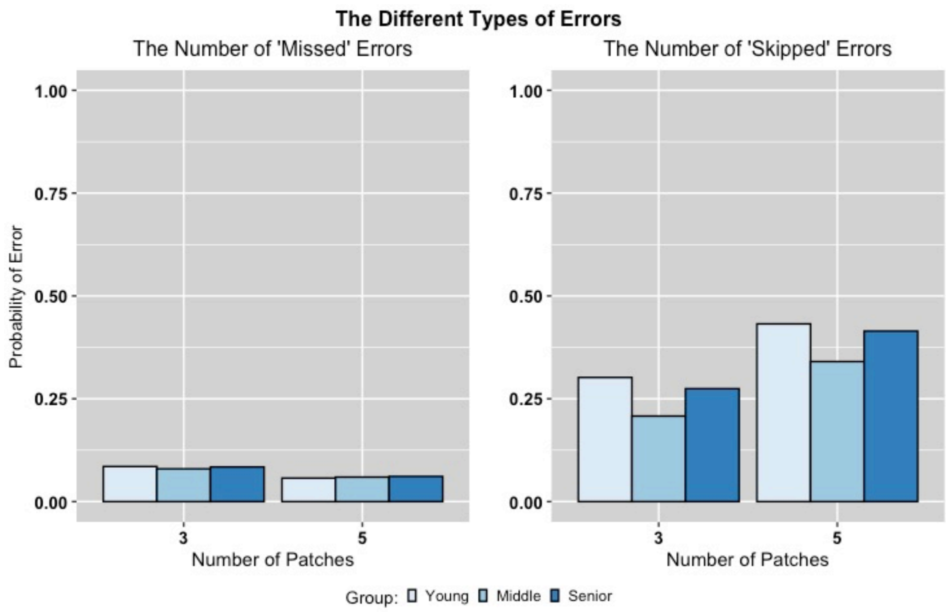

Figure 5.10. The left figure shows the proportion of errors that resulted from participants identifying the acorn in the incorrect patch. The right figure shows the proportion of errors that resulted from participants not searching the patch containing the acorn.

These results show that the middle age group's performance in the task was categorised by relatively lower errors resulting from not searching the patch containing the acorn and relatively higher repeat searches. The performance of the young and senior age groups on the other hand was categorised by similar levels of correct and error response patterns. In conjunction with the previous descriptive results, changes in the speed-accuracy trade-off appear to be driving different behaviours in the task that are not related to search strategy, but still determine performance in the task. To assess the significance of these behavioural changes, I subsequently conducted a series of inferential analyses.

To assess whether there was a significant difference in performance between the three age groups, I conducted a series of inferential analyses. Firstly, to assess if there was a significant difference between the three age groups in terms of STs, I conducted a 3(trial type) x 3(age group) mixed ANOVA on STs. There was a significant main effect of trial type, $F(1.566, 183.278) = 315.643$, $p = <.001$, $\eta_p^2 = .73$. There was also a significant interaction between trial type and age group, at $F(3.133, 183.278) = 12.613$, $p = <.001$, $\eta_p^2 = .18$. To further assess the significant difference in STs between trial types, a post-hoc Bonferroni–Holm *t*-tests was conducted, refer to Table 5.5 below.

Table 5.5. Post-hoc comparisons of the within-groups factor of trial type.

		Mean Difference	SE	*t*	Cohen's d	p
Patch 1 ST	Patch 3 ST	-4.10	0.30	-13.711	-1.25	< .001
	Patch 5 ST	-7.51	0.30	-25.089	-2.29	< .001
Patch 3 ST	Patch 5 ST	-3.41	0.30	-11.378	-1.04	< .001

There was also a significant main effect of age group, $F(2, 117) = 23.533$, $p = <.001$, $\eta_p^2 = .29$. To further assess the significant difference in STs between age groups, post-hoc Bonferroni–Holm *t*-tests was conducted, refer to Table 5.6 below.

Table 5.6. Post-hoc comparisons of the between-groups factor of age group.

		Mean Difference	SE	*t*	Cohen's d	p
Middle	Senior	-3.07	0.60	-5.126	-0.47	< .001
	Young	0.83	0.60	1.390	0.13	0.502
Senior	Young	3.90	0.60	6.516	0.60	< .001

Inferential results on STs for the different trial types are as expected: the more patches there were to search, the more time participants spent searching and consequently the longer it took to provide a trial response. Results on ST differences between age groups showed that STs were indeed different between the age groups, supporting the claim that the different age groups would empathise different aspects of the speed-accuracy trade-off and thereby produce distinct STs. The greatest mean difference was between the senior and young age groups, who are expected to display opposing behaviours.

To assess if there was a significant difference between the three groups in terms of accuracy, we also conducted a 3(trial type) x 3(age group) mixed ANOVA on accuracy. There was a significant main effect of trial type, $F(1.562, 182.738) = 156.445$, $p = <.001$, $\eta_p^2 = .57$. A non-significant interaction between trial type and age group was found. To further assess the significant difference in accuracies between trial type, post-hoc Bonferroni–Holm *t*-tests were conducted, refer to Tab 5.7 below. There was a non-significant main effect of age group.

Table 5.7. Post-hoc comparisons of the within-groups factor of trial type.

		Mean Difference	SE	*t*	Cohen's d	p
Patch 1	Patch 3	0.07	0.01	7.060	0.65	< .001
	Patch 5	0.19	0.01	17.576	1.60	< .001
Patch 3	Patch 5	0.11	0.01	10.515	0.96	< .001

These results show that participants' accuracy rates varied significantly when performing single, three or five patch trials. However, although there is a visible illustrative difference between the age groups in terms of overall accuracy, the strength of this difference is not statistically significant. In combination with the previous results, these findings also suggest that the emphasis on different components of the speed-accuracy trade-off by the different age groups, can drive difference patterns of performance.

Modelling results

The ABC modelling procedure was performed on all the data for each of the three age groups separately and followed the same procedure as in Experiment 1 of this chapter. The ABC modelling fits yielded good results for overall accuracy and mean RT across the three groups.

Here RTs explicitly refer to the time from the start of a trial to when a response was provided. The young age group yielded good fits, refer to Figure 5.11 below.

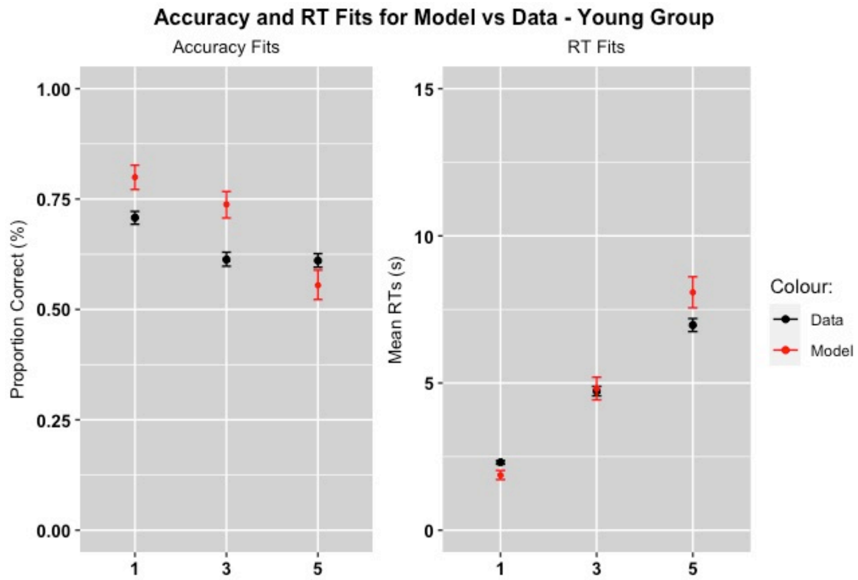

Figure 5.11. Model vs data fits for the young age group for accuracy and mean RT across the one, three and five patch trial types. Model and data points are shown with 95% confidence intervals.

The middle age group yielded satisfactory fits, refer to Figure 5.12 below.

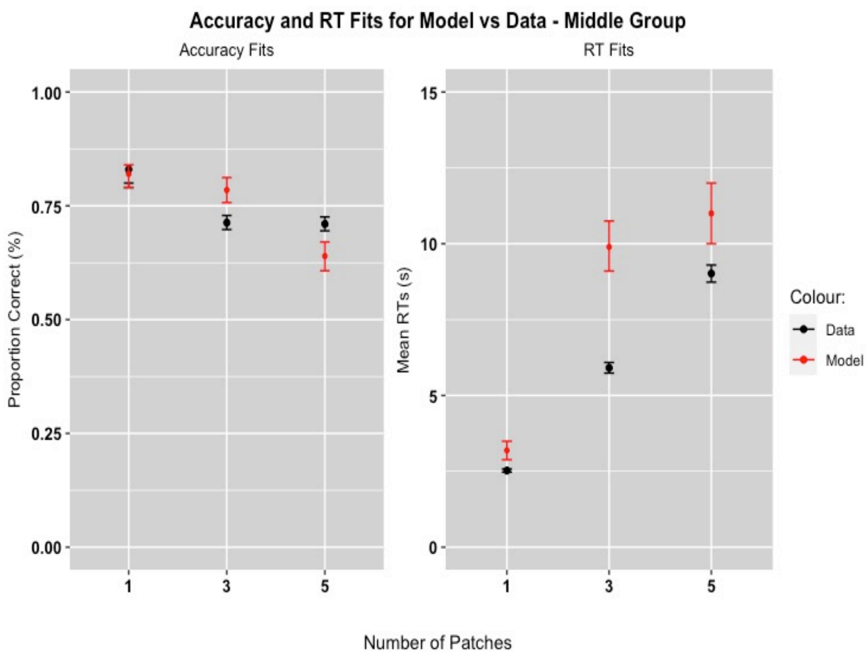

Figure 5.12. Model vs data fits for the middle age group for accuracy and mean RT across the one, three and five patch trial types. Model and data points are shown with 95% confidence intervals.

The senior age group yielded good fits, refer to Figure 5.13 below.

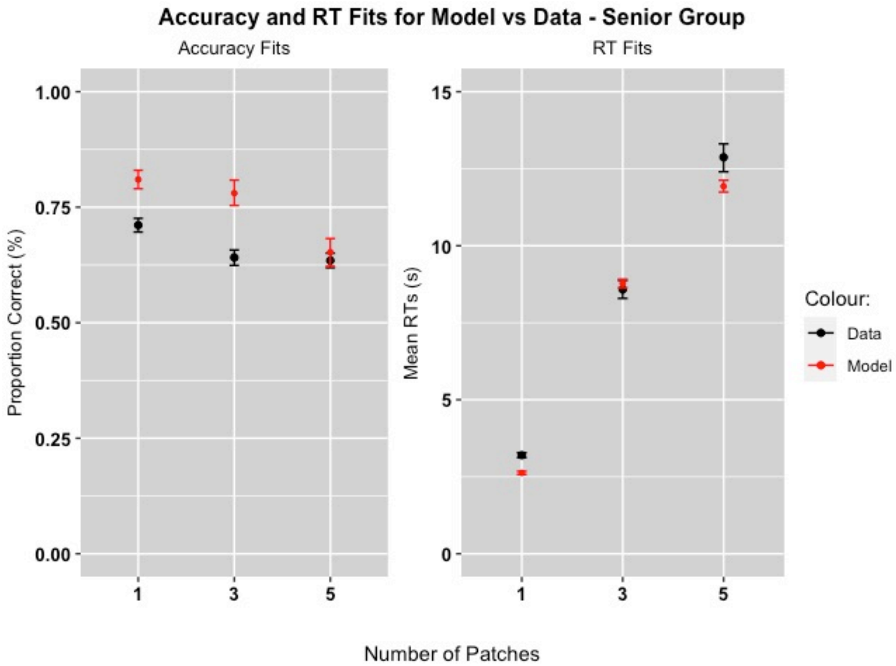

Figure 5.13 Model vs data fits for the senior age group, for accuracy and mean RT across the one, three and five patch trial types. Model and data points are shown with 95% confidence intervals.

If behaviour in this task is driven by changes in parameters controlling the speed-accuracy trade-off (response conservativeness) and not changes in search strategy, then parameter values should not change between trial types. As more patches to search in a trial only adds to a chain combination of single patch search strategies, not to the development of a new search strategy. As such, in addition to finding different response conservativeness parameter values in the three age groups that represent and drive the different behaviours of these groups, these values should not change as the number of patches to search in a trial increases. In other words, a participant's level of response conservativeness when searching a trial with a single patch, determines the method of searching that patch. As the number of patches to search in a trial increases, the same level of response conservativeness is applied to the additional patches. This results in all patches being searched independently as they would be on single patch trials. This results in a chain combination of single patch search strategies on three and five patch trials. This is distinctly different from unique foraging search strategies across multiple patches that involve patch searches that depend on searches in previous patches.

Refer to Table 5.8 below for ABC estimates of the main parameters concerned. Estimated mean parameter values gradually increase for the present and absent thresholds from young to senior age group. This is consistent with the view that younger participants trade speed for accuracy in the speed-accuracy trade-off and therefore have lower response conservativeness, whereas the senior age group behave in the opposite manner. The absent threshold is also higher than the present threshold across all groups. This indicates that response conservativeness is high when deciding that the target acorn is not present. Drift rates on the other hand appear to be similar across the young and middle age groups, but markedly lower for the senior age group.

Table 5.8. Estimated mean parameter values (M) and associated 95% high density intervals (HDIs) for the three age groups.

	Threshold - Present	Threshold - Absent
Young	M = 4.81 HDI = 4.61-5	M = 8.08 HDI = 7.97-8.20
Middle	M = 5.99 HDI = 3.35-8.62	M = 8.33 HDI = 7.92-8.75
Senior	M = 6.33 HDI = 6.09-6.57	M = 9.1 HDI = 8.92-9.28

An analysis of main effects appears to support these findings. A 2(present and absent threshold) x 3(age groups) mixed ANOVA was conducted to assess the differences in estimated threshold values across the groups. There was a significant main effect of threshold, $F(1, 3598) = 137.618$, $p < .001$, $\eta_p^2 = .04$, with the absent threshold being significantly higher ($M = 6.94$, $SD = 2.07$) than the present threshold ($M = 6.62$, $SD = 1.79$) Additionally, a significant interaction was found between threshold and age group, $F(2, 3598) = 2446.050$, $p < .001$, $\eta_p^2 = .58$.

There was also a significant main effect of age group, $F(2, 3598) = 3104.754$, $p < .001$, $\eta_p^2 = .63$. To further assess the significant difference in age groups, post-hoc Bonferroni–Holm *t*-tests were conducted, refer to Tab 5.9 below.

Table 5.9. Post-hoc comparisons of the between-groups factor of age group.

		Mean Difference	SE	t	Cohen's d	p
Middle	Senior	-2.509	0.036	-69.451	-1.157	< .001
	Young	2.419	0.036	66.957	1.116	< .001
Senior	Young	0.090	0.036	2.493	0.042	0.038

These results confirm the differences that exist in the ABC estimated parameter threshold values. Response conservativeness appears to be set differently for deciding whether the target acorn is present or absent, with response conservativeness being set higher for deciding if the target is absent within a given trial patch. Results also show that response thresholds were highest for the senior age group and lowest for the young age group. For the middle age group, this may represent a more balanced and optimal setting of the speed-accuracy trade-off that resulted in better overall performance in the task.

5.2.3 Discussion

Our descriptive results show that different age groups were categorised by different speed-accuracy trade-offs. The young age group traded speed for accuracy, in direct contrast to the senior age group who did the opposite. However, the middle age group showed a more balanced trade-off that resulted in higher accuracy rates than the other age groups, but with median STs between the other groups. Although this behaviour by the middle age group is not a result of participants behaving strictly optimally, it may represent an attempt at being more optimal than the other groups in terms of their overall performance (Ratcliff et al., 2004; Starns & Ratcliff, 2010). However, this is representative of a more general point: changes in the setting of the speed-accuracy trade-off drive overall performance in foraging tasks. This point assumes that irrespective of the optimal or sub-optimal nature of the trade-off set by participants, any change in the trade-off will be reflected in changes in overall performance.

These results further detail the asymmetrical relationship that can exist between pairs of response thresholds as modelled in EAMs. For example, in the present experiment the response threshold for the target is absent response was notably higher than the target is present response. Starns & Ratcliff (2010**)** found that across a series of two-alternative forced choice memory recognition tasks, the older participant groups were primarily categorised by wide response boundaries. This was representative of a higher response threshold for both binary responses in the tasks. However, our present results show that although this generally holds true for inter-group differences, such as pairs of response thresholds for the senior vs young age group, there is a clear asymmetrical relationship between threshold pairs. It may be the case that the degree of this asymmetrical relationship between threshold pairs may be a more indicative measure of overall task performance, than the overall difference between response thresholds.

It is important to note here that the type of errors and accuracies committed are assumed to be determined explicitly by the degree of response conservativeness for the target is present and absent response thresholds. For example, increasing the target is present threshold, and thereby increasing response conservativeness, leads to more repeated searches. This is because participants require more information or "evidence" in favour of that response, which leads to longer STs and more "repeated" searches. However, lowering this threshold leads to less information or "evidence" being required in order to trigger the associated response. As such, STs are shorter and less patches are searched, resulting in more "skipped" errors.

The present experiment did not make explicit use of time-bond trials in an attempt to get participants to forcefully manipulate the speed-accuracy trade-off to determine their performance. Instead, the differences in trade-offs were assumed to be inherent to different age groups. This is a crucial point, as it also assumes that if performance on a foraging task was governed by changes in search strategies, there ought to be a negligible difference in response conservativeness between age groups. As Boot et al. (2009) found, in the presence of response feedback in inter-task visual searches, search strategies eventually converge on a default strategy. In the present experiment where the paradigm was based on intra-task visual searches with response feedback, response strategies were accepted to not converge (Boot et al., 2009). However, if a dominant search strategy is not present to drive behave and ultimately performance, what does? Our results show that three distinct speed-accuracy trade-offs are present between the three groups that correspond with three distinct performance patterns.

This ties back to GUTs, optimality and their relevance in driving foraging behaviour. The present experimental paradigm does not have time-controlled trials or trials that are affected by previous performance in any form. In effect, participants are not penalised or rewarded for spending more or less time on any one patch. As such, optimising mean STs in an attempt to maximise some reward, as suggested by GUTs, is not relevant to this specific task. Consequently, optimality in the present task satisfies no task specific objective and consequently gives no task specific advantage to participants. As although we may assume participants are attempting to respond as accurately as possible, it is also an assumption to believe that participants are trying to do so in as little time as possible. As the present task places no emphasis on time, I believe this assumption does not hold for the majority of participants. As such, the view that foraging behaviour in non-time bound trials at the very least, are determined largely by participants attempting to find an optimal foraging strategy does not appear plausible. This is explicitly different from the view expressed here that the middle age group represents a more optimal speed-accuracy trade-off setting relative to the other two groups. This is because the speed-accuracy trade-offs set in the present task are assumed to be inherent to the individual participant groups and not set by participants only in response to the task itself. In this sense, trade-offs are not actively set or learned by participants, as assumed in the literature.

Additionally, there are other issues with the assumption of optimality in the present task. McNamara & Houston (1985) argue that assumptions surrounding foraging behaviour usually assume that information on certain key environmental parameters are known to the forager. However, this is not always the case. Taking the MVT as an example, the researchers factored in a behavioural rule that both allowed the forager to learn the key parameters of its environment and optimally exploit what was learned. The researchers derived that if foragers learned as suggested by the MVT, it would take an infinite amount of time for their behaviour to converge on the optimum. This is also because learning the key parameters from an ever-changing environment to determine the optimal strategy may never be attained. As the environment and the associated parameters required are never stable (Pierce & Ollason, 1987). As such, understanding overall performance in the present task as being derived from an active attempt at trying to find the optimal foraging strategy does not seem plausible.

On the other hand, an EAM perspective provides the most plausible explanation for the results found. From an EAM perspective, the only properties of the task under the participant's control are how much evidence should be required before triggering one of the two binary responses. In other words, when the participant should decide that the acorn is present or absent in a patch. This is a participant's response conservativeness. For the present task, this would represent the same level of response conservativeness set throughout the task for each patch searched. That is, because each trial type contains one or more of the same patches, all patches are governed by the same response thresholds. As such, the argument is made that the process of foraging in a region with one patch is identical to foraging in a region with multiple patches, except that foraging behaviour is concatenated across the different patches in the region. These results explain the consistent relative performance between groups across the different trial types. As the degree of response conservativeness remained fixed throughout the experiment across the different age groups, a similar pattern of performance across the three age groups remained fixed and scaled accordingly.

As the first experiment showed, no single search strategy dominated across the three and five patch trials. In line with (Boot et al., 2009), search strategies remained diverse across the intra-task trials. A similar pattern of results can be expected for the present expanded second experiment that is largely identical to the first. Furthermore, if overall performance was determined by simple changes in search strategies, what we ought not to see are categorical differences in response thresholds across the three different age groups, that clearly correspond with differences in performance across the different age groups. However, the three age groups show clear use of a speed-accuracy trade-off. Young participant traded speed for accuracy, whereas the senior age group did the opposite. Furthermore, the middle age group showed a more balanced approach to this trade-off and therefore, as expected performed better overall. Additionally, these observed differences in performance attributable to differences in the speed-accuracy trade-off are backed up by different response thresholds derived from the three age groups. Simple strategy changes do not explain these differences, nor how they are related to performance. However, from an EAM perspective, observed differences in response conservativeness directly control speed-accuracy trade-offs observed in this experiment.

These findings are also in line with those reported by Hommel, Li and Li (2004). These researchers found that in single-feature and conjunction-feature search tasks the primary difference in participants as age increased from a sample of 6 to 89 years old, was performance impairment due to target absent trials. From an EAM perspective, this behaviour is captured and explained through an asymmetrical relationship between binary response threshold pairs that yield higher response conservativeness for target absent trials. Furthermore, as age increases general response conservativeness increases and this asymmetrical relationship between threshold pairs still holds. The result is that older participants take more time to action a response in general and also take even longer to action a target absent response. From a purely search strategy-based perspective this behaviour is not captured or explained.

In conjunction with the results found by Bogacz et al. (2006), if foraging behaviour, or indeed visual search behaviour in general, is dependent on variations in response conservativeness and not the rate of information process, this may have implications for clinical research. Distinct differences in the behaviour of more senior individuals during cognitive related tasks may not be strictly due to a decrease in the rate of information processing associated with cognitive decline in older age. Instead, distinct differences in performance in these tasks may be a result of an increased level of response conservativeness, possibly brought on by more life experiences where delayed responses have been the more prudent course of action. This may have wider implications for the diagnostic literature. It may be relevant to distinguish between the rate of information processing and response conservativeness associated with task performance when assessing older participants. As such, exploring response thresholds in older participants may help in differentiating between cognitive decline associated factors and standard response conservativeness associated factors, especially in visual searches.

The argument remains that participants may still be in the process of learning the “optimal” search strategy and any changes in EAM response thresholds are due to a training of search strategies. Therefore, from this perspective when the optimal search strategy is found, the respective response thresholds will settle to new values. I believe this to be an inadequate explanation. Firstly, it does not explain the categorical and monotonic differences in EAM response thresholds or the consistent asymmetrical relationship between response threshold pairs across age groups. As Pierce and Ollason (1987) argue, the assumption that participants learn and train optimal search strategies depends on the assumption that key parameters from the environment remain fixed. Even though the present experiment represents a standardised visual foraging task, the assumption that these parameters are stable in this task and therefore derivable, assumes that they are known and were factored into the design of the experiment. However, this is not the case. Another crucial factor is the ecological validity of learning optimal search strategies beyond experimental settings. The plausibility of participants being able to learn all key parameters within a real-world environment that is constantly changing in order to derive optimal search strategies, is very low.

In conclusion, an EAM account of visual foraging behaviour provides the most complete account of the age-related factors associated with such tasks. Additionally, an EAM account sheds further light on the dominance of speed-accuracy trade-offs in driving performance in visual search tasks, as opposed to visual search strategies. The present findings also have wider implications for the diagnostic literature in relation to better differentiating between genuine cognitive decline in older participants and distinct changes in response conservativeness.

Chapter Six: General Discussion

In this thesis I attempted to show the novel insights that can be found as a result of either redesigning experimental paradigms so that they better suit modelling procedures or by applying a specific class of RT models to data. An initial argument was made to show that while speed accuracy trade-offs are useful in determining how participants form response strategies and what may be driving performance, in certain cases the relationship between accuracy and response is not linear. In these cases, there must be a meaningful way of combining accuracy and response. It is in these cases that RT modelling can prove to be a significantly insightful analysis tool. More specifically, EAMs have been a class of RT models proven to be highly robust, accurate and insightful in these instances (Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998). Furthermore, applying a computational framework allows for more thorough in-depth analyses of the underlying cognitive information processing structures (Harding et al., 2016; Townsend, 1990). Additionally, as proposed in the opening chapter of this thesis, applying a computational framework causes high-level cognitive tasks to be reduced to more rudimentary forms. This minimises disruptive high-level cognitive effects, such as framing effects and interpretation biases (Guest & Martin, 2021; Van Rooij & Blokpoel, 2020).

In the first part of this thesis, I showed how EAMs can be used to provide useful insights into the underlying cognitive processes associated with a classic cognitive bias: the CF. Previous experimental paradigms focused on improving the widely used high-level descriptive scenario-based task to assess the occurrence of the CF. In order to apply an EAM to the task, the experimental paradigm was transposed to the psychophysical domain. This represents an immediate benefit of applying RT models to higher-level cognitive tasks. Specifically, transposing such tasks to the psychophysical domain reduces them to low-level task and minimises unwanted higher-level effects, such as interpretation issues and uncontrolled levels of subjectivity in the task. Another significant advantage associated with creating experimental paradigms suited to modelling, is the opportunity for larger datasets through more iterations of the task. As shown in chapter two, the traditional scenario-based task used to assess the CF could successfully be transposed to a psychophysical task that allowed for more variations in experimental conditions, without jeopardising the internal validity of the task itself. In effect, this allowed for large scale iterations and therefore larger datasets.

Preliminary EAM results revealed that committing the CF fallacy was associated with a distinct EAM parameter: the rate of information processing. Specifically, these results revealed that the underlying cognitive systems were processing stimuli at a slower rate during CF eliciting trials, compared to non-CF eliciting trials. This is significant because the stimuli that constituted the CF eliciting trials also constituted the non-CF eliciting trials. As such, the rate at which information from the stimuli was processed for the two trial types was expected to be similar. Subsequent attempts were then made to show how EAMs can be combined to produce more sophisticated and insightful RT models. These more elaborate modelling exercises revealed that changes in the rate of information processing associated with committing the CF were linked to a specific information processing order. Moreover, this extended EAM revealed new insights into the underlying cognitive architecture associated with the CF phenomenon, that could not be derived from traditional experimental paradigms and procedures. They permitted a thorough analysis of the underlying information processing structures underlying the CF and revealed new findings. Specifically, findings that link the CF not to an error in combining conjunct probabilities into conjunction probabilities in general, but to a specific information processing order that produces a processing bias that facilitates the CF.

Reducing the traditional high-level CF task to a psychophysical experiment did produce a more iterative version of the CF task, which minimised unwanted secondary effects. However, transposing the task to a completely different domain removes the ability to assess whether there exist other high-level situations where the CF is not present. In other words, exhaustively determining whether the CF is present in all high-level scenario-based situations is not possible in this new domain. A natural extension to this second chapter is expanding on the modelling results. Although the modelling procedure in chapter two revealed that an information processing bias is associated with committing the CF, the effects of the biasing parameter have not been thoroughly explored. For example, the findings suggest that a bias parameter, φ, controls CF rates. However, the exact nature of this relationship is currently undefined.

Modelling results for the first experiment in the second chapter were exploratory and were aimed to allow some insights into the underlying information processing architecture of the present CF task. The second experiment in this chapter allowed a more elaborate modelling procedure involving more complete EAMs. However, the selected model (S-BSP model), which had an architecture based on SFT, is not strictly speaking a serial processing model. That is, while the two LBA components which constitute the model are activated one after the other, the precise way in which these components interact, through a bias parameter, is not a feature of standard SFT. This naturally poses the question of whether the proposed S-BSP model has an architecture that aligns with standard SFT. Furthermore, the assumption of selective influence in SFT, where each individual feature of a stimulus is processed separately in a serial manner, appears to be violated by the S-BSP model. This is because the S-BSP model does not assume that each individual feature of the stimulus is processed individually, but rather that separate groupings of the stimuli are serially processed individually. I argue that the S-BSP model satisfies all of these assumptions, when viewed from the perspective of a coactive model. In SFT, coactive models form a broad class of models, which (arguably) combine serial and parallel SFT processing architectures in yet unexplored ways. I believe that the S-BSP model represents a class of coactive models that function serially, but have additional features and parameters (like the biasing parameter), which function as an additional communication channel between both processing channels. Additionally, the S-BSP model may also capture a more elaborate aspect of the selective influences assumption, whereby processing channels are not only responsible for processing individual features, but also individual feature groupings.

EAMs form part of a class of models based on the laws of CPT and although these models can be used to find insights into human cognition, their underlying probability structure can be constrained through the QPT framework. That is, models based on QPT can both capture and explain a variety of cognitive effects captured by CPT and others not yet captured by CPT. The QPT account of interference effects between two questions or processes is an example of this. The QPT framework can account for and explain interference between two questions or processes through incompatibility. However, CPT and its present extensions do not capture such interference effects or theoretically permit them. This difference in predictions on interference effects functions as a constraint on all probability theories and resulting models that do not account for such effects.

The results in this thesis do not put QPT in direct opposition to CPT. Instead, they show that there exists a board hierarchy of probabilistic frameworks that can account for non-normative behaviours, such as interference effects. Therefore, models based on CPT or its extensions, like evidence accumulation RT models, do not represent a flawed interpretation of cognitive processes or non-normative behaviour, but instead represent one perspective for capturing these phenomena. Furthermore, just as Costello and Watts (2014) extended CPT in their modelling exercise, future attempts may be made to extend CPT or reconcile its differences with certain aspects of QPT to account for interference effects. Nonetheless, subsequent results in this thesis show that theoretical and model constraints brought on by interference effects do not consistently hold for EAMs.

One limitation of the series of experiments conducted in Chapter 3 is to do with the operationalisation of positive and negative interference. Although from a quantum perspective interference can be interpreted as being either positive or negative, from a cognitive perspective these characterizations are not as clear. Although the argument made in this chapter is that the direction of the inference is associated with how an association between scenarios is set up, not all of the present results support this. It may be the case that more explicit or direct associations are required to consistently elicit a positive or negative interference between the events in each scenario. A final notable challenge remains with understanding how the mind can behave in a quantum-like manner, given that all the evidence we are aware of point towards an understanding of brain neurophysiology as purely classical. I certainly do not suggest that the brain behaves in a quantum manner at the neuronal level, but quantum theory does appear to be useful in being able to capture various aspects of cognitive processing – how this comes about remains a challenge in such research.

The S-BSP model detailed in the second chapter of this thesis describes the CF phenomena as being a result of some interference in the serial evidence accumulation process. The next step to this modelling exercise was to assess what other non-normative phenomena this combined EAM can extend to. Given that the model gives a serial account of non-normative behaviour, trying to capture non-normative behaviour that occurs due to a serial processing account was the logical next step for this model. One such example is the suggested quantum interference effect observed by Kvam et al. (2015). While the quantum interference effect observed by Kvam et al. (2015) represents a phenomenon with a serial processing account that ought to be captured by the S-BSP model, it also functions as a constraint on EAMs. This is because EAMs are underlined by CPT and therefore, according to the quantum perspective, should not be able to capture the observed quantum effects.

The results from chapter four are not decisive. However, they do suggest that the quantum interference effect observed by Kvam et al. (2015), can still be observed when an experimental paradigm based on a standard EAM is used. Additionally, the argument is made that the observed effect of interference in RTs is hard to distinguish from an effect of response priming. Also, the findings in this chapter revealed that when the original experimental paradigm was altered to represent a more standard EAM processing account of the effect, the effect could still be observed. However, the present experiments had a largely similar response setup for participants compared to the original experiment, which I argue is highly conducive to response priming in the condition where an effect of interference is suspected to occur. It is argued that this effect of response priming may be responsible for causing a difference in RTs between conditions and not a quantum related effect. These results imply that the observed "quantum" interference effect is not unique to quantum experimental and modelling frameworks. Additionally, these results show that EAMs do not appear to be as constrained by interference effects as suggested by the quantum approach. Instead, they reveal how extensions of CPT can overcome these constraints and provide a simpler processing account of relatively complex phenomena. Nonetheless, there remains scope to further extend the S-BSP model to the results of Kvam et al. (2015). Developing an experimental paradigm that simultaneously allows for the quantum and S-BSP models to be fitted to the data, would allow for a more direct comparison of both approaches.

More generally, the findings from chapter four reveal the strength of simple processing accounts like EAMs. They also show how such models can be combined into more sophisticated processing models that can account for more complex phenomena otherwise unexplained by CPT. The first example of this is the S-BSP model account of the CF. The fourth chapter is another example of how a simpler processing model account can capture effects that are assumed to be incompatible with CPT, such as the interference effect. Although neither of the experiments in this chapter captured only the EAM features to assess if some effect of interference could still be observed, the consistent pattern of results even when the paradigm shifted from a largely EAM based framework to a quantum one, strongly suggests that the effect will be present in an entirely EAM based paradigm. This chapter does not attempt to put the CPT perspective against the QPT perspective. Rather, it shows how extending processing accounts of cognitive phenomena can bridge the gap between different perspectives to provide an alternative, and possibly simpler, explanation of cognitive effects like the CF and interference effects. Ultimately both perspectives provide viable methods for interpreting cognitive phenomena.

In Chapter 4, one of the main limitations is the lack of direct model comparisons. In this chapter, I argue and show that more standard EAMs are able to capture "quantum" effects without the need to resort to an explicit quantum EAM. I show that this can be achieved by making incremental changes to the experimental paradigm introduced by the original researchers (Kvam et al., 2015). Although I have presented behavioural results consistent with my argument, there remains the issue of whether a standard EAM based on classical probability theory can indeed computationally capture these effects. This is a valid point, but it remains beyond the scope and timeline of this PhD thesis. There is also the possibility that the changes introduced in the series of experiments in this chapter fundamentally altered the original experimental paradigm to such an extent, that it no longer adequately represents the original task. I do not believe that this is the case. Firstly, if this assumption were true, finding behavioural results largely identical to the original experiment would call into question the effect itself. That is, if the experimental paradigms I introduced in this chapter are inconsistent with an interference effect, then why is it that the behavioural results I observed are so similar to the original results? As I argue in the chapter, I believe this is because the experimental paradigm facilitates response priming, which produces the interference effect proposed by the original authors and that this is a feature of the original task, still present in the altered paradigms explored in this chapter. I believe this is the reason why similar results are observed both in my experiments and the original task.

EAMs can also be applied to assess intrinsic age-related factors associated with task performance. As shown in chapter five, EAMs best capture and represent the performance drivers in a basic foraging task. Additionally, the perspective helps differentiate between information related processing performance and response conservativeness related performance. In this fifth chapter, I show how unlike in chapter two where EAMs are used to explain performance beyond the scope of standard speed-accuracy trade-offs, EAMs can be used as a method of further analysing the trade-off itself. Specifically, EAMs reveal that during search tasks, the main driver of performance is how an individual's speed-accuracy trade-off is set, as determined by their level of response conservativeness.

Another crucial point that is echoed in chapter two, is how high-level cognitive processes can be reduced to low-level ones, while more clearly and accurately capturing performance drivers. In chapter five the main opposing argument is that foraging behaviour is determine by high-level search strategies of some sort. However, the findings in this chapter reveal that a simpler speed-accuracy trade-off account of performance best captures performance and the unique inter age group characteristics, when an EAM framework is applied. As previously mentioned, this shows how computational models can supplement simpler accounts of performance, to reveal an additional layer of performance drivers without introducing less tractable higher-level concepts.

Finally, in the fifth chapter I showed how EAMs can separate information related performance drivers and response conservativeness related drivers. Specifically, separating out these two characteristics of performance allows for a more accurate assessment of the influence of factors in and out of a participant's control. In the context of a clinical population, the findings from chapter five show that applying an EAM approach can assess whether information processing related factors are substantially affected, in line with cognitive decline or disease (Starns & Ratcliff, 2010). As EAMs can isolate the information processing feature of a task from the rest of an individual's performance, to better determine if it is functioning at an average or relatively standard level. I believe that this represents a highly impactful and exciting area of investigation, and may represent some of the first findings on the diagnostic potential of an EAM perspective. A natural next step for this project is to test a relevant clinical population, to determine if the EAM presented in this chapter can indeed differentiate between task performance indicative of clinical cognitive decline and elevated response conservativeness in older populations. Furthermore, there remains scope to improve the model fits in this fifth chapter, by possibly using more augmented ABC model fitting procedures.

The main argument of chapter five is that search strategies are controlled by the speed-accuracy trade-off. Furthermore, in a foraging task like the one in the present set of experiments, no particular search strategy is expected. Although the present set of experiments used an experimental paradigm with multiple foraging regions, there remains the possibility that the introduction of more foraging regions would elicit more coordinated foraging behaviour, i.e. behaviour which would reflect particular search strategies. This remains to be determined and represents a future avenue of research. In future experiments, time-bound trials could be introduced to assess optimality. Another point concerns optimality and time limitations. Although not introducing time-bound trials was an intended feature of the present experiments, it removed the possibility of evaluating the optimality of the observed behaviour using more standard procedures (i.e. whether participants could minimise the number of patches searched and the amount of time spent searching, while maximizing the likelihood of correctly identifying the target). This also represents a promising avenue for future research.

The principal theme throughout this thesis has been how EAMs can provide valuable and unique insights into the underlying cognitive processing systems behind various behaviours. That is, how a simple processing account can be extended through EAMs to provide viable explanations for a variety of complex cognitive effects, by reducing the mechanisms underlying them to simple accumulated processing features. Furthermore, combining simple EAM processing accounts into more sophisticated models can possibly help capture and explain cognitive phenomena beyond the current scope of its underlying CPT framework. Additionally, the thesis shows how such a simple processing account can even capture and represent general performance drivers.

References

Ahn, W. K., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology*, *31*(1), 82–123. https://doi.org/10.1006/cogp.1996.0013

Albar, F. M., & Jetter, A. J. (2009). Heuristics in decision making. *PICMET: Portland International Center for Management of Engineering and Technology, Proceedings*, 578–584. https://doi.org/10.1109/PICMET.2009.5262123

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112-149.

Anderson, J. R. (1991). The Adaptive Nature of Human Categorization. *Psychological Review*, *98*, 409-429.

Berchicci, M., Lucci, G., Pesce, C., Spinelli, D., & Di Russo, F. (2012). Prefrontal hyperactivity in older people during motor planning. *NeuroImage*, *62*(3), 1750–1760. https://doi.org/10.1016/j.neuroimage.2012.06.031

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy?: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision*, *9*(3), 1–16. https://doi.org/10.1167/9.3.7

Boot, W. R., Kramer, A. F., Becic, E., Wiegmann, D. A., & Kubose, T. (2006). Detecting transient changes in dynamic displays: The more you look, the less you see. *Human Factors*, *48*(4), 759–773. https://doi.org/10.1518/001872006779166424

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols*, *1*(5), 2277–2281. https://doi.org/10.1038/nprot.2006.390

Broadbent, D.E. (1958). Perception and communication. London: Pergamon

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

Busemeyer, J. R. & Bruza, P. (2011). Quantum models of cognition and decision making. Cambridge University Press: Cambridge, UK.

Busemeyer, J. R., Huang, J., Ebelt, Z. & Pothos, E. M. (in press). *The quantum sequential sampler. In Proceedings of the 2023 Conference of the Cognitive Science Society.*

Costello, F. J. (2009). How probability theory explains the conjunction fallacy. *Journal of Behavioral Decision Making*, *22*(3), 213–234. https://doi.org/10.1002/bdm.618

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*(3), 463–480. https://doi.org/10.1037/a0037010

Costello, F., & Watts, P. (2017). Explaining High Conjunction Fallacy Rates: The Probability Theory Plus Noise Account. *Journal of Behavioral Decision Making*, *30*(2), 304–321. https://doi.org/10.1002/bdm.1936

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking and Reasoning*, *14*(2), 182–199. https://doi.org/10.1080/13546780701643406

Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology and Evolution*, *25*(7), 410–418. https://doi.org/10.1016/j.tree.2010.04.001

Csilléry, K., François, O., & Blum, M. G. B. (2012). Abc: An R package for approximate

Bayesian computation (ABC). *Methods in Ecology and Evolution*, *3*(3), 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Dawson, N. V., & Arkes, H. R. (1987). Systematic errors in medical decision making: - Judgment limitations. *Journal of General Internal Medicine*, *2*(3), 183–187. https://doi.org/10.1007/BF02596149

Dakic, B., Paterek, T., & Brukner, C. (2014). Density cubes and higher-order interference theories. *New Journal of Physics, 16*, 023028.

Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, *41*(4), 1095–1110. https://doi.org/10.3758/BRM.41.4.1095

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*(5), 1–36. https://doi.org/10.1167/11.5.14

Epping, G., Fisher, E. L., Zeleznikov-Johnston, A. M., Pothos, E. M., & Tsuchiya, N. (2023). A quantum geometric framework for modeling color similarity judgments. *Cognitive Science*, e13231.

Evans, N. J. (2019). A method, framework, and tutorial for efficiently simulating models of decision-making. *Behavior research methods*, *51*, 2390-2404.

Evans, N. J., Dutilh, G., Wagenmakers, E. J., & van der Maas, H. L. J. (2020). Double responding: A new constraint for models of speeded decision making. *Cognitive Psychology*, *121*. https://doi.org/10.1016/j.cogpsych.2020.101292

Franco, R. (2009). The conjunction fallacy and interference effects. *Journal of Mathematical Psychology*, *53*(5), 415–422. https://doi.org/10.1016/j.jmp.2009.02.002

Feynman, R.P., Leighton, R., & Sands, M. (2005). The Feynman Lectures on Physics. The Definitive and Extended Edition. Addison Wesley, Boston.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364.

Gigerenzer, G., & Gaissmaier, W. (2015). Decision Making: Nonrational Theories. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Second Edi, Vol. 5). https://doi.org/10.1016/B978-0-08-097086-8.26017-0

Gould, S. J. (1988). The streak of streaks. The New York Review of Books.

Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585

Hardy, L. (2002). Why quantum theory?. In *Non-locality and Modality* (pp. 61-73). Springer, Dordrecht.

Harding, B., Goulet, M.-A., Jolin, S., Tremblay, C., Villeneuve, S.-P., & Durand, G. (2016). Systems Factorial Technology Explained to Humans. *The Quantitative Methods for Psychology*, *12*(1), 39–56. https://doi.org/10.20982/tqmp.12.1.p039

Haven, E. & Khrennikov, A. (2013). Quantum Social Science. Cambridge: Cambridge University Press.

Hertwig, R., & Gigerenzer, G. (1999). The ‘conjunction fallacy’ revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*(4), 275–305. https://doi.org/10.1002/(sici)1099-0771(199912)12:4<275::aid-bdm323>3.3.co;2-d

Hikmat Haji, S., & Mohsin Abdulazeez, A. (2021). Comparison Of Optimization Techniques Based On Gradient Descent Algorithm: A Review-Palarch’s. *Journal Of Archaeology Of Egypt/Egyptology*, *18*(4), 2715–2743.

Hommel, B., Li, K. Z. H., & Li, S. C. (2004). Visual search across the life span. *Developmental Psychology*, *40*(4), 545–558. https://doi.org/10.1037/0012-1649.40.4.545

Izzo, Z., Zou, J., & Ying, L. (2021). *How to Learn when Data Gradually Reacts to Your Model*.

Retrieved from http://arxiv.org/abs/2112.07042

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. Psychological Review, 397-402.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. Journal of Experimental Psychology: General, 126, 278-287.

Kahneman, D. (2001). Thinking fast and slow. Penguin: London, UK.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Kahneman, D., Tversky, A., & Tversky ', A. (1979). Prospect Theory: An Analysis of Decision under Risk ECONOMETRICA ICI VOLUME 47 MARCH, 1979 NUMBER 2 PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK. *Source: Econometrica*, *47*(2), 263–292.

Kolmogorov, AN (1933/1950). Foundations of the theory of probability. N.Y. Chelsea Publishing Co.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), 10645–10650. https://doi.org/10.1073/pnas.1500688112

Lee, C. M. & Selby, J. H. (2017). Higher-Order Interference in Extensions of Quantum Theory. Foundations of Physics, 47, 89-112.

Lieder, F. & Griffiths, T. L. (2019). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, 1-85.

Lopes, L. L. (1982). Doing the impossible: A note on induction and experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 626—636.

Maguire, P., Moser, P., Maguire, R., & Keane, M. T. (2018). Why the conjunction effect is rarely a fallacy: How learning influences uncertainty and the conjunction rule. *Frontiers in Psychology*, *9*(JUL). https://doi.org/10.3389/fpsyg.2018.01011

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. Serial dependence in position occurs at the time of perception. Psychon Bull Rev **25**, *2245–2253* (2018). https://doi.org/10.3758/s13423-018-1454-5

Mcnair, J. N. (1982). *Optimal Giving-Up Times and the Marginal Value Theorem Author (s): James N . McNair Source : The American Naturalist , Apr ., 1982 , Vol . 119 , No . 4 (Apr ., 1982), pp . 511-529 Published by : The University of Chicago Press for The American Society*. *119*(4), 511–529.

McNamara, J. M., & Houston, A. I. (1985). Optimal foraging and learning. *Journal of Theoretical Biology*, *117*(2), 231–249. https://doi.org/10.1016/S0022-5193(85)80219-8

Moro, R. (2007). On the nature of the conjunction fallacy. *Harbin Gongye Daxue Xuebao/Journal of Harbin Institute of Technology*, *39*(11), 1710–1714. https://doi.org/10.1007/s

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research*, *49*(10), 1286–1294. https://doi.org/10.1016/j.visres.2008.12.005

Nowakowska, A., Clarke, A. D. F., & Hunt, A. R. (2017). Human visual search behaviour is far from ideal. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1849), 1–6. https://doi.org/10.1098/rspb.2016.2767

Oaksford, M. & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann.

Pierce, A. G. J., & Ollason, J. G. (1987). Eight Reasons Why Optimal Foraging Theory Is a Complete Waste of Time Published by : Wiley on behalf of Nordic Society Oikos Stable URL : http://www.jstor.org/stable/3565560 REFERENCES Linked references are available on JSTOR for this. *Oikos*, *49*(1), 111–118.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of "rational" decision theory. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1665), 2171–2178. https://doi.org/10.1098/rspb.2009.0121

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*(3), 255–274. https://doi.org/10.1017/S0140525X12001525

Pothos, E. M., Lewandowsky, S., Basieva, I., Barque-Duran, A., Tapper, K., & Khrennikov, A. (2021). Information overload for (bounded) rational agents. *Proceedings of the Royal Society B*, *288*, 20202957.

Pothos, E. M., & Busemeyer, J. R. (2022). Quantum Cognition. *Annual Review of Psychology*, *73*, 749–778. https://doi.org/10.1146/annurev-psych-033020-123501

Pothos, E. M., Waddup, O. J., Kouassi, P., & Yearsley, J. M. (2021). What Is Rational and Irrational in Human Decision Making. *Quantum Reports*, *3*(1), 242–252. https://doi.org/10.3390/quantum3010014

Ramirez, J. C. & Marshall, J. A. R. (2017). Can natural selection encode Bayesian priors? *Journal of Theoretical Biology*, *426*, 57-66.

Ratcliff, R., & McKoon, G. (2008). Drift Diffusion Decision Model: Theory and data. *Neural Computation*, *20*(4), 873–922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424. https://doi.org/10.1016/j.jml.2003.11.002

Reyna, V. F., & Rivers, S. E. (2009). *Current Theories of Risk and Rational Decision Making Adolescent Risk Taking and the Science of Judgment and Decision Making*. *28*(1), 1–10.

Sagi, D., & Julesz, B. (1987). Short-range limitation on detection of feature dif- ferences. *Spatial Vision*, *2*(1), 39-49.

Scherer, L. D., Yates, J. F., Baker, S. G., & Valentine, K. D. (2017). The Influence of Effortful Thought and Cognitive Proficiencies on the Conjunction Fallacy: Implications for Dual-Process Theories of Reasoning and Judgment. *Personality and Social Psychology Bulletin*, *43*(6), 874–887. https://doi.org/10.1177/0146167217700607

Shiffrin, R. N. (2021). Is it reasonable to study decision-making quantitatively? *Topics in Cognitive Science*, 12541.

Simon, H. A. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69, 99-118.

Sinha, U., Couteau, C., Jennewein, T., Laflamme R., & Weihs, G. (2010). Ruling out multi-order interference in quantum mechanics. *Science*, *329*, 418.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1760–1765. https://doi.org/10.1073/pnas.0607208104

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2009). Sequential Monte Carlo without likelihoods (Proceedings of the National Academy of Sciences of the United States of America (2007) 104, 6, (1760-1765) DOI:10.1073/pnas.0607208104). *Proceedings of the National Academy of Sciences of the United States of America*, *106*(39), 16889. https://doi.org/10.1073/pnas.0908847106

Smith, E. E., Shafir, E., & Osherson, D. (1994). Similarity, plausibility, and judgments of probability. In P. N. Johnson-Laird & E. Shafir (Eds.) Reasoning and decision making, Oxford: Blackwell.

Sorkin, R. D. (1994). Quantum mechanics as quantum measure theory. *Modern Physics Letters A*, *9*(33), 3119-3127.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors*, *5*, 19-31.

Sperling, G. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica*, *27*, 285-292.

Stafford, T., Pirrone, A., Croucher, M., & Krystalli, A. (2020). Quantifying the benefits of using decision models with response time and accuracy data. *Behavior Research Methods*, *52*(5), 2142–2155. https://doi.org/10.3758/s13428-020-01372-w

Stanovich, K., & West, R. F. (1998). Individual differences in framing and conjunction effects. *International Journal of Phytoremediation*, *21*(1), 289–317. https://doi.org/10.1080/135467898394094

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*(2), 377–390. https://doi.org/10.1037/a0018022

Tenenbaum, J. B, Kemp, C., Griffiths, T. L., & Goodman, N. (2011). How to grow a mind: statistics, structure, and abstraction. Science, 331, 1279-1285.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science*, *28*(3), 467–477. https://doi.org/10.1016/j.cogsci.2004.01.001

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235-255.

Tiedemann, A. C., Sherrington, C., & Lord, S. R. (2007). Physical and psychological factors associated with stair negotiation performance in older people. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *62*(11), 1259–1265. https://doi.org/10.1093/gerona/62.11.1259

Timmis, M. A., Turner, K., & Van Paridon, K. N. (2014). Visual search strategies of soccer players executing a power vs. placement penalty kick. *PLoS ONE*, *9*(12), 1–16. https://doi.org/10.1371/journal.pone.0115179

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) be Distinguished. *Psychological Science*, *1*(1), 46–54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Treisman, A.M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282-299.

Trueblood, J. S., Yearsley, J. M., & Pothos, E. M. (2017). A quantum probability framework for human probabilistic inference. J*ournal of Experimental Psychology*: *General*, *146*, 1307-1341.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory : Cumulative Representation of Uncertainty Author (s): AMOS TVERSKY and DANIEL KAHNEMAN Published by : Springer Stable URL : http://www.jstor.org/stable/41755005 Accessed : 13-04-2016 07 : 57 UTC Your use of the JSTOR archive. *Springer*, *5*(4), 297–323.

Tversky, Amos, & Kahneman, D. (1983). Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment S. *Psychological Review*, *90*(4), 293–315. Retrieved from https://doi.org/10.1037/0033-295X.90.4.293

Van Rooij, I., & Blokpoel, M. (2020). Formalizing Verbal Theories: A Tutorial by Dialogue. *Social Psychology*, *51*(5), 285–298. https://doi.org/10.1027/1864-9335/a000428

Valone, T. J. (2006). Are animals capable of Bayesian updating? An empirical review. *Oikos,* *112*, 252-259.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. Proceedings of the National Academy of Sciences, 111, 9431–9436.

Wagenmakers, E. J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin and Review*, *15*(6), 1229–1235. https://doi.org/10.3758/PBR.15.6.1229

Wagenmakers, E. J., Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

Wagner, S., Helmreich, I., Dahmen, N., Lieb, K., & Tadi, A. (2011). Reliability of three alternate forms of the trail making tests A and B. *Archives of Clinical Neuropsychology*, *26*(4), 314–321. https://doi.org/10.1093/arclin/acr024

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*. https://doi.org/10.1186/1471-2105-11-116

Whitney, D. (2012). S3-1: The Serial Dependence of Visual Perception. *I-Perception*, *3*(9), 582–582. https://doi.org/10.1068/if582

Wolber, M., & Wascher, E. (2003). Visual search strategies are indexed by event-related lateralizations of the EEG. *Biological Psychology*, *63*(1), 79–100. https://doi.org/10.1016/S0301-0511(03)00028-0

Yearsley, J. M., & Trueblood, J. S. (2018). A quantum theory account of order effects and conjunction fallacies in political judgments. *Psychonomic Bulletin and Review*, *25*(4), 1517–1525. https://doi.org/10.3758/s13423-017-1371-z

Zaal, F. T. J. M., & Thelen, E. (2005). The developmental roots of the speed-accuracy trade-off. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1266–1273. https://doi.org/10.1037/0096-1523.31.6.1266

Zhu, J., Sanborn, A. N., & Chater, N. (2020). The Bayesian Sampler: generic Bayesian inference causes incoherence in human probability judgments. *Psychological Review*, *127*, 719-746.