# An Optimal Control Approach to Testing Theories of Human Information Processing Constraints

by

XIULI CHEN

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Computer Science
University of Birmingham
31 October 2014

# Acknowledgements

Undertaking this Ph.D has been a truly life-changing experience for me and it would not have been possible to write this thesis without the help and support of the kind people around me.

First and foremost, I would like to express my very special appreciation and thanks to my supervisor Professor Andrew Howes who has been a tremendous mentor for me. I have been extremely lucky to have a supervisor who cared so much about my work, who showed so much passion about the research and life, who responded to my questions and queries so promptly, and importantly who also sometimes be my friend. I could not have imagined having a better supervisor and mentor for my Ph.D study.

I was very much honoured to be involved in some projects with my supervisor and his colleagues. I must express my gratitude to my collaborators. Chapter 4 in this thesis is based on a draft 'A Model of Visual Search as Optimal Control Given Uncertainty in Peripheral Vision' submitted to Vision Research. Chapter 5 is based on a draft 'The Emergence of Interactive Behaviour: A Model of Rational Menu Search' submitted to CHI 2015. I thank Christopher W. Myers, Richard L. Lewis, Joseph W. Houpt, Gilles Bailly, Duncan P. Brumby and Antti Oulasvirta for their kindness and intelligence.

I am very grateful to the remaining members of my thesis group committee, Prof. Jeremy Wyatt and Prof. Russell Beale. Their academic support and input and personal cheering are greatly appreciated. I also would like to thank Benjamin R. Cowan, Chris Bowers and Bob Hendley for their friendly welcome when I first came to HCI and for being nice colleagues in past three years.

Also I would like to thank my friends. Thank Bin Wang for being there for me for so many years. Thank Juliusz Kopczewski for his weirdness. Thank Wen-Chi Yang for sharing his life philosophy. Thank Feng-Zhen Tang for her such sweet personality. Thank Li-Yan Song for sharing her enthusiasm for life and travel. Thank Jiang-Shan Yu for sharing his brave and confidence.

Special thanks go to my family, my patents, my grandma, my sisters and my brother.

Lastly, there were some big/small things that somehow leaded me to these wonderful four years. Thank my father for being willing to pay the expensive tuition fee for my Msc study in University of Sheffield. Thank my friend Xin Sun for giving me the website 'www.jobs.ac.uk'. Thank my exboyfriend Bin Wang for encouraging me applying this Ph.D position. Thank Prof. Alan Zinober from University of Sheffield for his extra long reference letter for my application of this Ph.D position.

OOO

# Abstract

This thesis is concerned with explaining human control and decision making behaviours in an integrated framework. The framework provides a means of explaining human behaviour in terms of adaptation to the constraints imposed by both the task environment and the information processing mechanisms of the mind. Some previous approaches tended to have been polarised between those that have focused on rational analyses of the task environment, on the one hand, and those that have focused on the mechanisms that give rise to cognition on the other hand. The former usually is based on the assumption that rational human beings adapt to the external environment by achieving 'goals' defined only by the task environment and with minimal consideration of the mechanisms of the human mind; while the latter focuses on information processing mechanisms that are hypothesised to generate behaviour, e.g., heuristics, or rules. In contrast, in the approach explored in this thesis, mechanism and rationality are tightly integrated. One interpretation of the framework decomposes the modelling problem into two parts: the state estimation problem and the optimal control problem. This thesis investigates a *state estimation and optimal control* approach, or *optimal control* approach for brevity, in which human behavioural strategies and heuristics, rather than being programmed into the model, emerge as a consequence of rational adaptation given a theory of the information processing constraints.

The thesis takes problems, each at different level, from visual perception (Chapter 4: the Distractor Ratio task) through immediate behaviour (Chapter 5: a Menu Search task), to probabilistic inference (Chapter 6: a information gathering and decision making task) and uses the optimal control approach to obtain the optimal policy; the behaviour of which is then compared to human behaviour.

The results show that (1) constraints imposed by human peripheral vision give rise to colour and letter search phenomena; (2) constraints imposed by peripheral vision, the statistics of the environment, and knowledge give rise to menu search phenomena; (3) constraints imposed by the information validities ranking discovery and the information cost give rise to certain decision

making phenomena. In each case, previous models, requiring the modeller to provide control rules or heuristics, are replaced with an optimal control model in which behaviour emerges from constraints.

Furthermore, the results show that the optimal control approach meets three challenges: (1) the challenge of finding emergent control strategies, rather than programming them in the model (2) the challenge of breadth of application and, (3) the challenge of testing different information processing theories by deriving optimal strategies.

Subsequently, the discussion considers the broader role of the optimal control approach in cognitive science. In particular, it focuses on the issue of whether or not people are optimal and the different roles of optimal state estimation and optimal control. It also, considers the possibility of integrated models of visual search and decision making.

# Contents

# List of Figures

# Chapter 1

# Introduction

One of the successes of cognitive science is the generation of a number of different approaches to explaining the mind. These include both *top-down* and *bottom-up* approaches (Chater & Oaksford, 1999; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; McClelland et al., 2010; Lewis, Howes, & Singh, 2014). *Top-down* approaches, sometimes referred to as *rational analysis*, *function-first* or, *purposive explanations*, are based on the assumption that rational human beings adapt to the task environment with minimal consideration of the mechanisms of the human mind. They focus on the problem that human beings are facing, e.g., the goal, the nature of the task environment. For example, according to Newell's knowledge-level analysis (A. Newell, 1982) the system structure is treated as a 'black-box'; its actions could be predicted by the rational solution to the task environment that the system is interacting with. Marr's computational level analysis (Marr, 1982) focuses on the nature of the problems that the visual system faces. The visual behaviour is then compared with an optimal solution to these problems. Other examples include Anderson's rational analysis (Anderson, 1990) and optimal foraging theory (Stephens & Krebs, 1986). Rational analysis offers the potential for deep 'why' explanations of human behaviour (Lewis et al., 2014). In fact, there is broad consensus that human cognition is adaptive (e.g., Brunswik, 1956; Marr, 1982; Stephens & Krebs, 1986; Anderson, 1990; Weber & Johnson, 2009; Griffiths, Chater, Norris, & Pouget, 2012). Based on this as-

sumption, the predictions are then considered more or less successful to the extent that human behaviour approximates the optimal solution for the task environment.

In contrast to *top-down* approaches, *bottom-up* approaches, sometimes referred to as *mechanism-first* or *mechanistic explanations*, focus on information processing mechanisms that are hypothesised to generate behaviour, e.g., heuristics, or rules. They carefully specify architectures, or algorithms for cognition with relatively little concern for why they work as they do. Various frameworks have been proposed to describe the role of cognitive mechanism. These include Newell's systems level analysis (A. Newell, 1982), Marr's algorithm and physical implementation level (Marr, 1982), Neural networks (McClelland, Rumelhart, & Group, 1986; Neal, 1995; MacKay, 1995). Investigations of the mechanism theory have typically involved proposals for information processing architectures that are hypothesised to simulate cognitive processes, for example, ACT-R (Anderson, Matessa, & Lebiere, 1997; Anderson et al., 2004), EPIC (Meyer & Kieras, 1997; Kieras, 1997) or connectionist architectures (McClelland, 1988). Mechanism focuses on 'how' explanation of human behaviour.

Unfortunately, research on *top-down* and *bottom-up* approaches tends to have been pursued separately. In fact, there is extensive and polarised debate about whether a *top-down*, *function-first* or *bottom-up*, *mechanism-first*, is a better approach for cognitive modelling. A broad perspective of this debate can be found in Griffiths et al. (2010); McClelland et al. (2010). They provided an overview of these two approaches respectively and their conceptual foundations, and give a flavour of the range of phenomena their modelling efforts have addressed respectively. Despite early promise (Anderson, 1990; Marr, 1982) there has arguably been little influence of one on the other (Bowers & Davis, 2012). On one hand, there is a danger of rational approaches being isolated from the consideration of the mechanism that is used to solve the problem. A closer investigation of mechanism offers deeper understandings of the information processing capacity. On the other hand, there is a danger of mechanism-first approaches resulting in ad-hoc mechanisms that are shaped as much by the modelers intuitions as by empirical evidence.

In contrast to the rational-first and the mechanism-first approaches, some more recent contributions to cognitive science have focused on explaining human decision making and control in a framework that combines the strength of both rational analysis and mechanistic approach to understand the mind. In these contributions, mechanism and rationality are tightly integrated using a framework first introduced by Baron and Kleinman (1969a), but subsequently neglected until Lewis et al. (2014); Howes, Lewis, and Vera (2009). This framework is referred to as 'Computational Rationality' by Lewis et al. (2014). This term is used in this thesis. The idea of this framework is that human behaviours are generated by cognitive mechanisms that are rationally adapted to the structure of both the mind (mechanisms) and the environment.

This thesis investigates one interpretation of the Computational Rationality framework, which decomposes the modelling problem into two parts: *the state estimation problem* and *the control problem*. The state estimation problem concerns how to integrate information into a task-relevant representation. The solution to the problem requires the formulation of a theory of information processing mechanisms (bottom up approach). For example, in a visual search task, the state estimation is constrained by the theory of visual system (e.g., acuity degradation away from fovea). Solving the optimal control problem determines what to do next given the state estimate. It also determines the rational strategy given the information processing limits imposed on the state estimate. The actions include overt task response (e.g., responds with 'Present' when a designated target is found), and intermediate information gathering actions (e.g., eye movements during a visual search task). The solution to the optimal control problem is determined by maximising a utility function and it is also, therefore, constrained by *top-down* considerations. Thereby provides a rigorous means of exploring computational rationality.

The *state estimation and optimal control* approach, or *optimal control* approach for brevity, presents a number of challenges. Three challenges have shaped much of my thesis work. These are described below.

The first challenge is to show how theories of information processing mechanisms

can give rise to control strategies without being programmed with those strategies. For the *optimal control* approach, the behavioural predictions are emergent from the rational analysis given the theoretical assumption of information processing mechanism. It aims to go beyond rational approaches that focus entirely on utility and environment, and that fail to address control. For instance, there are some approaches that focus on how to represent the environment but neglect how to use this representation to guide the selection of action, e.g., Bayesian observer/analysis of the environment, which offers an option to represent the environment and thus estimate the state of the world, but the action selection is then performed heuristically, e.g., by 'maximum a posteriori' (MAP), by maximising information gain or uncertainty reduction (Najemnik & Geisler, 2005, 2008; Myers, Lewis, & Howes, 2013).

The second challenge is to show that this approach can be applied across multiple levels of behaviours, e.g., perceptual, motor control learning, immediate behaviour routine tasks. The thesis takes problems, each at different level from visual perception (Chapter 4: the Distractor Ratio task) through immediate behaviour (Chapter 5: a Menu Search task), to probabilistic inference (Chapter 6: a information gathering and decision making task) and uses the *optimal control* approach to obtain the optimal policy; the behaviour of which is then compared to human behaviour.

The third challenge is to show how theories of information processing mechanisms can be tested by deriving optimal strategies. One advantage of the *optimal control* approach is that it assists addressing what is sometimes known as the strategy fitting problem. Testing a theory of human information processing and distinguishing it from alternative theories is challenging when it is difficult to determine behavioural correlates of the proposed mechanisms. In general, one reason for this difficulty is that neural information processing mechanisms admit a large range of strategies and it is these strategies that are more directly responsible for behaviour in interaction with the task environment. It is often the case that very different underlying neural mechanisms can support very similar spaces of possible strategies. It follows that it is not sufficient to merely choose a

mechanism and choose a strategy in order to explain behaviour. Rather, it is necessary to determine which strategies are efficient given which mechanism. By design, this exactly what the *optimal control* approach does and I am therefore optimistic about the potential of the approach as a means of distinguishing between alternative theories.

In what follows, Chapter 2: Background introduces the Computational Rationality framework, and some existing related work under this framework. Chapter 3, subsequently, introduces the *optimal control* approach, followed by three main chapters. Each chapter illustrates an application of the approach to a specific task.

In Chapter 4, the *optimal control* approach is used to test constraints imposed by human peripheral vision in a visual search task. Specifically, the model is used to explain phenomena associated with the distractor ratio paradigm in visual search (Bacon & Egeth, 1997; Shen, Reingold, & Pomplun, 2000). The model explains saccadic selectivity as a strategic response to uncertainty in the periphery. The emergent policy maximises reward by switching between selectivity favouring saccades to same-colour distractors and selectively favouring saccades to same-shape distractors. In other words, the model is used to test/identify the constraints of the visual system that are necessary for the human behaviour effects to arise.

In Chapter 5, the *optimal control* approach is used to test the hypothesis that in menu search tasks users rationally adapt to a combination of three sources of constraint (1) the ecological structure of interaction, (2) cognitive and perceptual limits, and (3) the goal to maximise the trade-off between speed and accuracy. An optimal control model of menu search was built and was tested with two studies of its predictions. The first study involved applying the model to a real world distribution of menu items and in the second study the models' predictions were compared to human data from a previously reported experiment (Bailly, Oulasvirta, Brumby, & Howes, 2014). The model was tested against existing empirical findings concerning the effect of menu organisation, menu length, and whether or not the target is a known word. The predictions of the model were largely supported by the experimental evidence.

In Chapter 6, the *optimal control* approach is used to test constraints imposed by the information usefulness discovery and the cost of information in a probabilistic inference paradigm. In the task, human information gathering and decision making behaviour, which was previously explained using a list of heuristics assumed people might adopt (Gigerenzer & Goldstein, 1996), was explained as the strategies that were flexibly developed in response to the distribution of task demands. The predictions of the model were largely supported by the experimental evidence. In addition, this approach offered novel predictions about the diversity and flexibility of decision-making strategies.

Finally in the discussion, Chapter 7, I review the extent to which the models in Chapter 4, 5 and 6 meet the three challenges set out earlier in the introduction: (1) the challenge of finding emergent control strategies, rather than programming them in the model, (2) the challenge of breadth of application and, (3) the challenge of testing different information processing theories by deriving optimal strategies.

# Chapter 2

# Background

The *state estimation and optimal control* approach, or *optimal control* approach for brevity, proposed here is motivated by the desire to explain human behaviour as a rational adaptation to constraints imposed by psychological mechanisms. For this reason the first part of this Background chapter will review work on the *computational rationality* framework, which has been proposed as an approach that combines the strengths of both rational and mechanism approaches to explaining behaviour (Lewis et al., 2014).

In addition, there are a number of precedents to the thesis work each of which exhibits some elements of the *optimal control* approach. These include analyses of (1) visual/audio stimulus detection (Tanner Jr & Swets, 1954), (2) eye movement prediction in visual search tasks (Najemnik & Geisler, 2005), (3) manual operator system (e.g., pilot performance) (Baron & Kleinman, 1969a), (4) a memory encoding task (Gray, Sims, Fu, & Schoelles, 2006), (5) Psychological Refractory Period task (Howes et al., 2009). The remainder of the Background chapter systematically reviews each of these precedents in turn.

## 2.1 Computational Rationality

One way to summarise the computational rationality framework is with Figure 2.1. In the figure there are four components required to explain behaviour: Ecology, Utility, Mech-

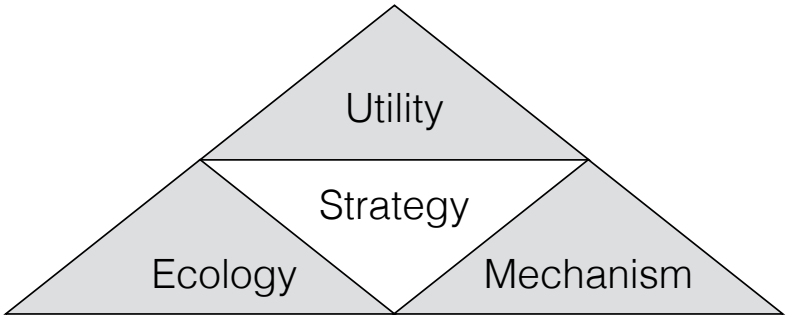



Figure 2.1: An illustration of the Computational Rationality framework (S. Payne & Howes, 2013). There are four components that are critical to understanding human control and decision making behaviours. Utility concerns what a person wants to do; Ecology concerns the constraints imposed by a person's experience of the task environment, including the immediate local task environment and environment experienced through a lifetime. Mechanism concerns the information processing system implemented in the human brain that determines what a person can do.

anism and Strategy. A space of strategies is determined by the three other components. Optimisation of the utility function leads to the selection of a strategy. The successful application of this framework requires a theory of each of the components. According to S. Payne and Howes (2013), the computational rationality framework is particularly inspired by Optimal Foraging Theory (Stephens & Krebs, 1986), and Cognitive Game Theory (Erev & Roth, 1998; Roth & Erev, 1995; Erev & Gopher, 1999). In the following sections I review the definition of each of these components.

### Utility

The study of utility is the study of quantitative representations of people's preferences and choices. For example, Expected Utility Theory (Von Neumann & Morgenstern, 1953) states that the decision maker chooses between risky or uncertain prospects by comparing their expected utility values. Expected Utility theory has had profound influence in many aspects of human behavioural science. Over the years, decision analysts have developed useful tools for assessing and applying knowledge about the utility structure of complex problems. Psychologists have approached explaining human decision making from the perspective of decision analysis. One of the examples is the cost-benefit approach (Beach

& Mitchell, 1978; Christensen-Szalanski, 1978; J. W. Payne, Bettman, & Johnson, 1988; V. L. Smith & Walker, 1993). According to this approach, individuals trade a strategy's costs against its benefits in making their decisions. People anticipate the 'benefits and costs of the different strategies that are available and choose the strategy that is best for the problem' (J. W. Payne, Bettman, & Johnson, 1993).

The Utility component in the framework provides a measure of the goodness of the strategies depending on the goal. The optimal strategy is obtained by finding a strategy that maximises expected utility. In practice, the utility function may play a number of theoretical roles. For example, in a laboratory experiment, it may correspond to the task goal (the objective utility). The assumption that could then be made is that the subjective utility exploited is consistent with the objective utility. Or in some cases it may require a theory of 'internal subjective utility', which might deviate from the objective utility function asserted by, e.g., experimental instructions. For example, it may specify desired speed-accuracy trade-offs, or temporal discount of the reward. Arguably this definition is related to notion of intrinsic motivation and reward (Singh, Lewis, Barto, & Sorg, 2010).

### Mechanism

The study of psychological mechanisms, whether memory, attention, motion perception, or movement control is of course the core of what academic psychology is about. Computational approaches to models of psychological mechanisms have given rise to a number of architectures. These include production system architectures ( using if-then rules and procedures that operates on them to explain thinking (e.g., Rosenbloom, Laird, & Newell, 1993; Kieras, 1997; Anderson et al., 1997)) and connectionist type architectures ( using artificial neural networks (e.g., McClelland et al., 1986) ) amongst others. Each architecture is defined as a set of mechanisms that process information and generates behaviours. Predictions are derived by programming the architecture with task knowledge and executing the program to generate simulated behaviour.

The Mechanism component in the framework refers to the architecture that maps from

perceptual inputs to interactive actions (e.g., eye movements, button presses) (Lewis et al., 2014). This includes mechanisms of cognitive and perceptual/motor systems, for example, human visual noise, motoric control noise, or memory limitations. With these considerations, the framework can be used to explore the range of possible strategies allowed by the architecture.

### Ecology

Ecology component in the framework refers to the ecological task environment. It is the definition of the environment to which the cognitive mechanism adapts to. The key point here is that the agent's strategy is adaptive to the task environment as *it has experienced*. In other words, the behaviours are partly determined by the statistics of the ecological environment.

For example, Anderson and Schooler (1991); Schooler and Anderson (1997) demonstrated that human memory behaves optimally. Specifically, human memory's 'forgetting' behaviours are adaptive to the ecological environment. They showed that the probability of needing the information in the ecological environment decays as a power-law function of time, which is consistent with human memory's forgetting behaviour (Anderson & Schooler, 1991; Schooler & Anderson, 1997). Therefore, the ecological environment is one of the elements to understand human behaviours.

Another example is related to perceptions of randomness (Hahn & Warren, 2009). For example, in a fair coin flipping task, people strongly believe the sequence HTHHTT (where H for heads and T for tails) is more likely to happen than the sequence HHHHHH. This kind of perception has long been regarded as bias. However, Hahn and Warren (2009) showed that the apparent bias is actually adaptive to the facts that (1) an individual experience of the coin flipping is only one trajectory through the whole probability tree, and (2) people have limited working memory or limited life time experience (i.e., this one trajectory of the probability tree is truncated). They showed that for a truncated trajectory of the whole probability tree, HHHHHH is actually less likely to happen than HTHHTT.

Therefore, to summarise, the Ecology considered by the optimal control approach is the ecological environment, as *it has been experienced*.

### Strategy

Before introducing the strategy space, three closely related terms used through this thesis are clarified. The first term is *heuristic*. In psychology, an *heuristic* refers to a set of simple, efficient rules used to solve the problem. An heuristic is a hypothesis about an available cognitive strategy. The second term is *policy*. In Reinforcement Learning, a *policy* is a complete specification of what to do in every possible state of a state space. The third term is *strategy*. Strategy is a relatively more general term, which refers to a prescription of what to do given a certain goal. In this thesis, *policy* and *strategy* are sometimes used interchangeably.

The three components above determine a space of strategies, which is the fourth component in Figure 2.1, and with a principle of rationality they then jointly determine the *optimal strategy* (also called an optimal policy). In other words, the space of strategies from which an individual will select is delineated by other three components. In absence of any one component, the strategy space would be inconsistent with human beings' and thus harder to explain their behaviour (S. Payne & Howes, 2013).

In the framework the strategies are an emergent consequence of rational adaptation to the other constraints in Figure 2.1. A predicted strategy can be derived through utility maximisation. In general, the optimal strategy could be derived by many formal approaches that require utility and reward functions, e.g., optimal control theory , dynamic programming; reinforcement learning. The goal of the thesis is to propose and test a general purpose means of finding optimal strategies for theories of the cognitive mechanisms.

In the next section various related approaches are reviewed.

## 2.2 Related work

The optimal control problem given constraints has been vigorously investigated in some fields (e.g., neurophysiology, computer science, engineering, ecology), however its full potential in developing models of human cognition has been realised on and off (e.g., Tanner Jr & Swets, 1954; Baron & Kleinman, 1969a; Baron, Kleinman, & Levison, 1970; Kleinman, Baron, & Levison, 1970; D. McRuer, 1980; Howes et al., 2009; Lewis et al., 2014). In this section below, it illustrates some exemplars, each of which exhibits some elements of the *optimal control* approach. Specifically, I list some preceding work of the *optimal control* approach with their strengths and weaknesses.

### 2.2.1 Signal Detection Theory

*Signal Detection Theory* (SDT) (Tanner Jr & Swets, 1954) is a early piece of work that shows that human behaviour is jointly determined by the cognitive architecture (e.g., a theory of visual or auditory perceptual system) and the task goal (the benefits of successful detections against the costs of false alarms).

The classic SDT task involves discrimination of a stimulus in which the presence of an auditory or visual stimulus is required to be detected against a background of noise. The 'Mechanism' assumption is that the external stimulus is transmitted by the sensory system and results in a measure of neural activity (state estimation). This transmission process is shaped by a theory of the mechanism of the sensory system. A judgement on the existence of a signal is then based on the measure of neural activity (a control problem given state estimation). Specifically, the model responds with 'present' if the measure of neural activity exceeds a threshold, otherwise 'absent'. The optimal decision strategy (linking) is the optimal threshold (strategy). The ecological constraint is the local task environment.

The utility function is assumed to trade the benefits of successful detections against the costs of false alarms. The optimal threshold is sensitive to the utility function. The

optimal detection threshold is derived mathematically by a statistical inference technique introduced by Peterson, Birdsall, and Fox (1954). The detection threshold is determined by both a parameter that represents the individual's sensory system and a parameter that represents a process of psychological control during the task.

SDT shows that behaviour can be explained as an optimal adaptation to organism and environment constraints, particularly noise. Behaviour is not simply a consequence of noise, it is a consequence of adaptation to noise. SDT is a early precedent that shows both external environmental noise and organism-internal noise could shape human behaviour. In this respect SDT is a means of explaining signal detection behaviour as a computationally rational adaptation (See Figure 2.1).

However, (1) SDT does not offer a general approach to explaining human behaviour as rational adaptation to more complex information processing mechanisms. For example, it is not possible to use for modelling a sequential control. In particular, the strategies do not inform the intermediate information gathering actions, e.g., gaze distributions. Also, (2) in SDT the perceptual control is abstracted to a single statistical parameter, and then embedded in a statistical inference structure. It thereby only makes limited assumptions about the detailed architecture of the perceptual system.

### 2.2.2 Ideal Observer Analysis

Ideal Observer Analysis (IOA) (Geisler, 2011) is a further development of Signal Detection Theory (SDT) (Tanner Jr & Swets, 1954), which has been applied to a range of visual search tasks. IOA was not only applied to signal detections but also to sequential control tasks, such as the gaze distribution during a visual search task. For example, in Najemnik and Geisler (2008), the task involved detecting the presence or absence of a Gabor patch against a noisy background. They found that the number, and spatial distribution of saccades during the task could be predicted by a IOA model. The model was sensitive to some known human visual constraints, e.g., the decreasing acuity with increasing eccentricity (Mechanism constraint). For the model, each saccade (action) was directed to the

location that was most likely to find the target, i.e., utility maximisation is regarded as information gain maximisation.

Most of the ideal observers are stated within the framework of Bayesian statistical decision theory. They assume that the visual search behaviour is informed by a Bayesian estimate of the current state of the world. For example, in Najemnik and Geisler (2005) the state estimate was calculated by optimally integrating information from the fovea, from the periphery, and from previous fixations according to its reliability. The reliability of information was determined by physiological and neural information processing constraints. The optimal state estimate was then used to guide action selection by, e.g., maximising the information gain of the target position.

The performance of an ideal observer is constrained by various sources of variability. These include, for example, (1) variability in the stimuli (e.g., photon noise, variability in scene illumination, variability due to the projection from a 3D environment to the 2D retinal images); (2) variability in the sensory neural representation (e.g., sensory neural noise), and (3) variability in the decoding circuits (e.g., decision and motor neural noise) (Geisler, 2011). They can be the basis of a deep investigation into the mechanisms of the human visual information processing system.

For the Ideal Observer Analysis, the state estimate is calculated by optimally integrating information from the fovea, from the periphery, and from previous fixations according to its reliability. The reliability of information is determined by physiological and neural information processing constraints. However, the optimal state estimate must be complemented with some search rules and/or stop rules. Specifically, Najemnik and Geisler (2005, 2008) calculated the optimal next fixation based on the optimal state estimate. To compute the optimal next fixation point, the ideal searcher considers each possible next fixation and picks the location that will maximise the probability of correctly identifying the location of the target after the next fixation. They calculated the one-step further posterior for each of next possible locations, and chose the one that maximised the information gain from the current posterior and next posterior. This approach is not optimal control

for two reasons. First, the search stopped where the maximum posterior probability exceeded a criterion, which is picked by matching the error rate of the human observers. This means that the stop rules are not learnt by optimal control. Secondly, as pointed out by Najemnik and Geisler (2005, 2008) themselves, the ideal searcher is not optimal because it only considers one fixation into the future.

This point will be a major focus of the study reported in Chapter 4.

### 2.2.3 Ideal Performer Model

The Ideal Performer model is related to Ideal Observer Analysis (Geisler, 2011) but focuses on the control problem, e.g., a machine learning method (reinforcement learning) is used to derive the optimal strategy of the behaviours. Gray et al. (2006) built the Ideal Performer model of the Blocks World task first introduced by Ballard, Hayhoe, and Pelz (1995); Ballard, Hayhoe, Pook, and Rao (1997).

In the Blocks World task reported in Gray et al. (2006), there are three windows in the interface: (1) a Target Window that contained a pattern of coloured items, i.e., 8 different coloured items randomly distributed across a $4 \times 4$ gridded window; (2) a Resource Window that linearly displayed the 8 coloured items in the Target window. (3) a empty $4 \times 4$ gridded window called Workspace Window. The participants are asked to replicated the coloured pattern in the Target window to the Workspace window by dragging items from the Resource window.

The different strategies to perform this task are related to how many blocks are able to be encoded in the memory after each visit of the Target Window. The more blocks being encoded correctly each time (i.e., a more memory-intensive strategy), the fewer transitions, both visual and motor transitions, across the windows would be required (i.e., less perceptual/motor intensive strategy). This is referred as the low-level cognitive interactive strategy (Strategy).

In the Ideal Performer Model (Gray et al., 2006), the mechanism assumptions concern (1) motor constraints, e.g., time cost for moving the blocks from the Resource window

to the Workspace window (based on Fitts' Law, Fitts, 1954); (2) the memory constraints, e.g., the time spent encoding an item; the time spent retrieving an item from memory; and the probability the retrieval will be successful (based on ACT-R, Anderson & Schooler, 1991). It then uses a reinforcement learning algorithm, Q-learning, to derive low-level cognitive interactive strategies that minimise the time cost (Utility). Ecology is the local task environment.

In the Ideal Performer Model (Gray et al., 2006), the optimal strategy for the action selection, i.e., how many blocks to be encoded at once, is learned with a reinforcement learning algorithm, Q-learning. In addition, the action selection is constrained by theoretical assumptions about human cognition and motor system. In this respect, it shares some similarities with the study present here. However, in the IPM the equations/parameters that described the constraints for encoding time, retrieval latency, and probability of recall were estimated by fitting the human data before the application of the optimal control on the task level. A short discussion about this parameter fitting is given below. In Gray et al. (2006)'s approach, the parameters of the architectural constraints are found by fitting the human data *before* that learning is used to find the optimal behaviour given the constraints. Another method of finding the parameter settings is by fitting the parameters in one condition, and then to see whether the model predicts the results from other conditions with the same parameter set. Both of these two methods drive the optimal control behaviours (strategies) for the architectures shaped by the fitted parameters. In contrast, in the studies presented in this thesis, I attempted to explore how the model performs across the parameter space. Specifically, optimal control learning is used to derive each optimal strategy for each parameter combination. Subsequently, the behaviour, and therefore the parameter set that best corresponds with human data, can be identified. One advantage of conducting the parameter fitting with the human data *after* the optimal control is that you get a better idea of whether the model predicts the qualitative phenomena regardless of the fit.

### 2.2.4 Cognitively Bounded Rational Analysis

In contrast to Ideal Performer Model (2.2.3), Howes et al. (2009) demonstrated that it could be problematic that simply choosing a single plausible strategy space, programming the architecture with this strategy space, and fitting quantitative free parameters to the data. In Howes et al. (2009), two different cognitive architectures, which differ in whether the cognitive processor had the capacity to select only one response at one time (ACT-R) or multiple responses (EPIC), were tested (Mechanism).

The model (Howes et al., 2009), was built on a Psychological Refractory Period (PRP) task. In the task, a verbal response was given to the pitch of a tone, whether the tone was high or low pitch, and a manual response was given to a visual pattern, whether the pattern contained a particular feature. The two responses must be ordered as demanded (e.g., a verbal response first). The participants were asked to do the task as quickly as possible with the correct order (Utility). The Ecology is the local dual-task environment. For each tested architecture, optimal programs were those that select scheduling signals and a wait process that maximises the utility. The optimal program selection is accomplished by using a tool called CORE (Constraint-based Optimal Reasoning Engine)(Vera, Howes, McCurdy, & Lewis, 2004; Howes, Vera, Lewis, & McCurdy, 2004a). In other words, CORE is used to generate predictions of the optimal behaviour, given theoretical assumptions. The theoretical assumptions of the constraints are specified to CORE in terms of relationships between events, e.g., the start times, the duration of processes, in the environment, tasks, and psychological processes (Howes et al., 2004a). Given these constraints on behaviour, CORE is then used to generate a prediction. First, for example for the dual task, the values of the start, durations, and other parameters, such as cost, are constrained by the posted equations, but their values are not yet uniquely determined. Then an optimal schedule is configured with an optimal scheduling algorithm. For example, optimal scheduling is used to find a schedule that maximise the utility and minimise the cost.

Howes et al. (2009) offered an example of testing different theoretical assumptions

about the cognitive architecture using the computational rationality framework, and contrast this to fitting a fixed cognitive architecture. However, the strategies (optimal programmes) were found by doing an exhaustive exploration of all possible strategies (using brute force). Hence, it is difficult to generalise to more complex tasks.

### 2.2.5 Optimal Human Operator

The term *Optimal human operator* or *Optimal controller* often appeared in control theory, man-machine system and automation literatures (Baron & Kleinman, 1969a; Baron et al., 1970; Kleinman et al., 1970; D. McRuer, 1980). A related literature is a perspective on Artificial Intelligence called 'bounded optimality' (S. J. Russell & Subramanian, 1995). The work has covered a wide range of topics, e.g., pilot performance (Baron et al., 1970), human spatial orientation space (Borah, Young, & Curry, 1988), human postural balance control (Kuo, 1995).

An optimal controller is a hypothetical device that performs a given task optimally given the available information and any specified constraints. In Kleinman et al. (1970)'s words, '...the basic assumption underlying is that the well-motivated, well-trained human operator behaves in a near optimal manner subject to his inherent limitations and constraints, and his control task'.

In Kleinman et al. (1970), the mechanism assumptions concerned the human visual and motor control systems, which were derived from empirical or theoretical accounts of human cognition. The mathematical representations of these mechanism assumptions were embedded into the optimal controller for the task. The tasks range from a simple single-axis control task (Kleinman et al., 1970) to complex manual control of the longitudinal position of a hovering XV-5A, a VTOL aircraft (Baron et al., 1970). The ecological constraints was that, e.g., they concern the flight dynamics of the controlled vehicles. The optimal operator's strategy was selected so as to minimise a utility function, which consisted of weighted sum of mean-squared tracking errors and/or control effort (e.g., time cost) (Baron et al., 1970).

This line of work is built on classical or modern control theory, but by including the consideration of the assumptions of human visual system and/or motor control system. There are two basic approaches for the control problem. One is based on classical multi-loop control theory (e.g., D. T. McRuer & Jex, 1967). The other one is rooted in modern control and optimisation theory (e.g., Kleinman et al., 1970). The former relies heavily on judgements concerning the closed-loop system structure (D. T. McRuer & Jex, 1967). For the latter, although it could be applied to a wide range of manual control situations, the dynamical systems being controlled are usually assumed to be linear (e.g., Kleinman et al., 1970). Due to this strong assumption about the dynamics of the system, the application of this model is limited. In contrast, for example, reinforcement learning assumes little about the dynamics of the system. Instead, it develops a good control function through on-line, trial and error learning.

## 2.3 Summary

In this chapter I have reviewed a number of approaches, in which the behavioural predictions are jointly determined by the cognitive architecture and the task environment. These approaches differ in two aspects as follows.

(1) How are the mechanism assumptions represented in the optimality problem specification? For example, in the signal detection theory, the external visual stimulus is transmitted as a neural measure given the assumptions of the perceptual system. The neural measure of real external stimulus is regarded as the *state estimation* of the real world situation. The *state estimation* is constrained by the theoretical mechanism assumptions. The state estimation then serves as the basis of the action selection. Also, in the ideal observer analysis (e.g., Najemnik & Geisler, 2005), the state estimation (through Bayesian integration) is obtained by integrating information from the fovea, from the periphery, and from previous fixations according its reliability. The reliability of information is determined by physiological and neural information processing constraints. The optimal state estimation

then again serves as the basis of the action selection.

(2) How are the optimal strategies derived? For each approach mentioned above, some form of optimality analysis was used to derive the predictions. These include (1) statistical inference techniques (e.g., Signal Detection Theory in section 2.2.1), Ideal Observer Analysis in section 2.2.2), (2) Control Theory and Estimation (e.g., Optimal human operator in section 2.2.5), (3) Machine Learning (e.g., reinforcement learning in section 2.2.3), a special designed tool, CORE (Cognitively Bounded Rational Analysis in section 2.2.4).

In what follows, I will explain how these precedents vary in their generality and, particularly, how they vary in their ability to meet the three challenges set out in Chapter 1.

The first challenge is to pursue an approach in which human behavioural strategies, rather than being programmed in the model, emerge as a consequence of utility maximisation. For the ideal observer analysis (section 2.2.2), after an optimal estimate of the world states, a heuristic rule was required to guide the action section, e.g., 'maximum a posteriori' or the information gain maximisation. For Ideal Performer Model (section 2.2.3), the strategy space was pre-delineated by a set of production rules, i.e., the memory strategy was based on ACT-R architecture in which the cognitive strategies were governed by a set of predefined rules.

The second challenge is to pursue an approach that can be applied across multiple types of human behaviours. Signal Detection Theory (section 2.2.1) could not be used to explain more complex human behaviours. For example, it was not possible to use for modelling sequential control, e.g., the saccadic sequence. Optimal Control Operator 2.2.5 failed this challenge too as it had strong assumptions about the dynamics of the system. For example, the system being controlled was usually assumed to be linear (e.g., Kleinman et al., 1970). For Cognitively Bounded Rational Analysis (section 2.2.4), the strategies (optimal programmes) were found by exhaustively exploring all possible strategies. Hence, it was difficult to generalise to more complex tasks.

The third challenge is to pursue an approach in which theories of information pro-

cessing mechanisms can be tested by deriving optimal strategies. For the Ideal Performer Model (section 2.2.3), the strategy space was pre-delineated by a set of production rules (ACT-R). In addition, The ACT-R parameters were estimated by fitting the outcome data, which made it a less convincing test of the theory.

In the next chapter I propose an approach that addresses these identified problems.

# Chapter 3

# Framework: state estimation and optimal control

This chapter is to provide a high-level overview of the *state estimation and optimal control* approach and its background. Greater details of the application of the approach to a specific task or a type of task are given in the three following chapters, Chapter 4, 5, 6.

## 3.1 Overview

An overview of the *state estimation and optimal control* approach, or *optimal control* approach for brevity, is presented in Figure 3.1. In the figure, the information from the external task environment (on the bottom left of Figure 3.1) is encoded by noisy perceptual processes to generate a task relevant perception. For example, if the task is to find a word starting with the letter 'X' from a list of words that are alphabetically ordered, then one possible task relevant perception would determine the first letter of the word fixated and how far this letter is from 'X'. Sometimes, this encoding process involves some knowledge possessed by the model besides the external perceptual information. For example, in the simple task mentioned above, the search behaviour would be greatly affected by model's knowledge of the alphabet table. This encoded task relevant perception is then integrated with the previous state into a new state representation (on the bottom

right of Figure 3.1). It thereby integrates information across the time steps to form a state representation.

Subsequently, the optimal controller (on the top right of Figure 3.1) chooses an action on the basis of the new state and the current *policy* (which determines a *state-action value function*). For example, it would either choose to gather more information or make a response decision. The *initial policy* has no control information and results in random action selection. State-action values are updated incrementally (learned) as reward and cost feedback is received from the interaction. Hence, through reward guided control optimisation, the *policy* converges on the *optimal control policy* that maximises the reward. This interaction is thereby defined as a *reinforcement learning* problem.

The function of the optimal controller (top right in Figure 3.1) is to choose which action to take next. It does so by looking up the value of each action available in the current state and picking the one with the highest value[1]. These values are called *Q-values* or *state-action values* and they are stored in a *Q-table*. The most important question to answer is then how these *state-action values* are learnt and why they therefore implement the optimal control policy. To answer this question, it requires us to introduce the concept of a Markov Decision Process (MDP), and introduce one of the reinforcement learning solutions, *Q-learning*.

It is worth noting that the Q-learning is not a theoretical commitment. Its purpose is merely to find the optimal control policy. It is not used to model the process of learning and is used instead to achieve methodological optimality (Oaksford & Chater, 1994). The use of reinforcement learning here is consistent with Gray et al. (2006) and Chater (2009)'s proposal for the role of reinforcement learning in rational analysis. In particular, in the *optimal control* approach, the task problems are defined as reinforcement learning problems, so any reinforcement learning solver/algorithm that is guaranteed to converge on the optimal policy is sufficient to meet the needs of the theory. Q-learning was chosen

[1]The problem is defined as a reinforcement learning problem. A reinforcement learning algorithm, Q-learning, among many others is chosen for the optimal controller. Therefore, the description of the optimal controller is based on the Q-learning.

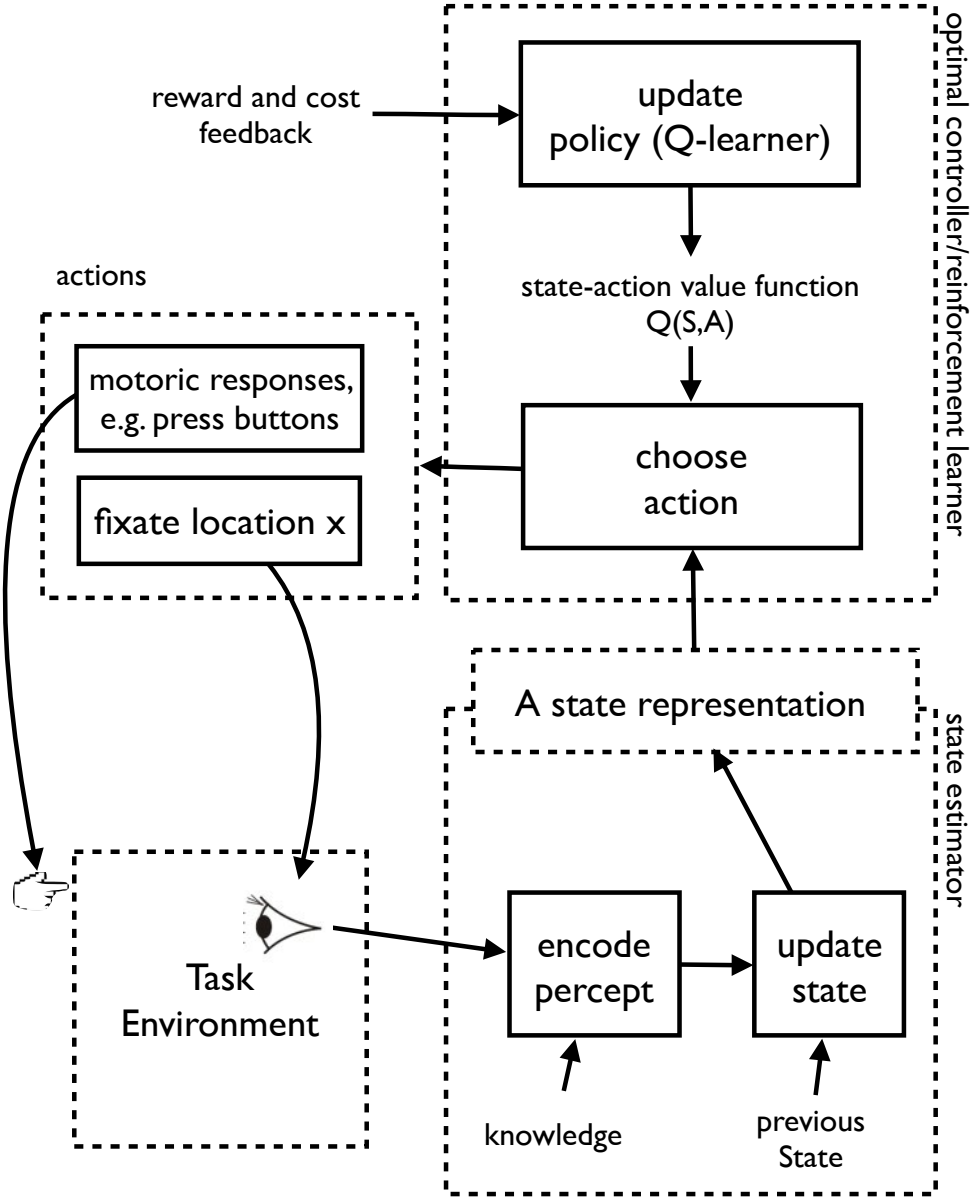



Figure 3.1: An overview of the state estimate and optimal control approach. The state estimator (bottom right) encodes information through perceptual mechanisms. The optimal controller (top right) chooses actions on the basis of a state-action value function. It learns this function with Q-learning given reward and cost feedback.

due to its simplicity and because it has a formally proved optimality guarantee. However, it is known that Q-learning can easily lose viability for large problems, so a more efficient learning mechanism will be pursued in the future (this will be discussed in the Discussion chapter). There are numerous algorithms based on Q-learning, which are more powerful (although also more complicated). The key idea motivating these algorithms is to represent the value functions and policies approximately (for a review see Busoniu, Babuska, De Schutter, & Ernst, 2010). For example, one variation involves value function approximation, and one of the most popular and straightforward ways to integrate approximation in Q-learning is using gradient descent. However, these more powerful algorithms have no advantage for the scientific analyses conducted in this thesis over Q-learning where the primary requirement is a guarantee of optimality.

In the remainder of this chapter, in Section 3.2, an overview of the state estimation is given. The basic concepts of reinforcement learning are reviewed in Section 3.3, including the concept of a Markov Decision Process (MDP). In Section 3.4, a reinforcement learning algorithm, Q-learning, is reviewed.

## 3.2 State Estimation

The state estimation component of the model represented in Figure 3.1 (bottom right) implements the psychological constraints imposed by the human visual and cognitive mechanisms. The output of the state estimation is a task relevant state representation, which serves as the basis for the action selections in the control problem. Two questions that naturally follow are (1) what is the information content of the state, (2) how is the state represented. The answers to these questions are critical for understanding the derived optimal policy, and for understanding the tested theories of visual and cognitive mechanisms. It is known that the problem of how to design the state is often challenging in the application of the reinforcement learning (Sutton & Barto, 1998).

The state design is usually task-oriented and theory-oriented. For example, one of the

guidelines for the state design is that the state includes whatever information is *available* to the agent (Sutton & Barto, 1998). It is assumed that the information is given by the information processing systems that is part of the environment that the agent adapts to. Therefore, the availability of the information to the agent is constrained by both the tested theories of human visual and cognitive mechanisms and the task. One example is that the visual information gained at one fixation is constrained by the theory of human visual system. Another example concerns which level of information needs to be considered in the state. For example, you want to find your key on a messy desk with dozens of objects spreading over the desk. If the hypothesised theory is that people are able to remember which objects they have seen, then it follows that the state representation needs to be rich enough to distinguish different objects, e.g., phone, pen, mug. Also, it may be necessary to represent relationships, for example, knowledge that the key is more likely to be close to the phone. The model based on this theory would then find optimal control behaviours for these state representation. An alternative theory is that people are not able to remember which objects they visited, but only remember whether the target object has been seen so far. For example, if the goal is to find the 'key' then it only need encode that 'pen' and 'mug' are not the target. In this case, the state representation only needs 3-levels for each location (e.g., NaN for don't know, 1 for target object and 0 for non-target object).

Furthermore, in extreme cases the state might only include immediate sensations, such as visual perceptions. At another extreme the state might be a highly processed version of original sensations. For example, a Bayesian updated posterior of a set of hypothesis given all the observations. A full understanding of the theoretical implications behind different state representations is critical.

In this thesis, I explored these aspects of the state estimation. For example, in Chapter 4, the state does not encode the colour information, red or green, or shape information, cross or circle, for the objects visited. Instead, a relevance score to the target based on the colour and shape information is used. In addition, the information, i.e., the relevance estimate for each location, across the time steps is integrated according to the error variances

(the measurement noise). In Chapter 5, again only the relevance estimate information of the objects visited to the target is kept in the state representation. The state is a vector, each element of which represents the information (the relevance estimate) obtained for each location of the display. Along with the fixations, the information obtained for each location is added to the state vector. Greater details will be given in the three main chapters following.

## 3.3 Reinforcement Learning and Markov Decision Processes

In this section the basic concepts of reinforcement learning are reviewed (a more complete review can be found in e.g., Sutton and Barto (1998)).

### 3.3.1 Agent-environment interaction

The agent-environment interaction process as shown in Figure 3.2 captures the basic elements of a reinforcement learning problem. In the figure, an agent interacts with the environment through three signals, state transitions, rewards, and actions. At each time step $t$ the agent finds itself in a state, $s_t$. The agent then chooses an action $a_t$ at time step $t$. As a (partial) consequence of the action $a_t$ chosen, the environment gives feedback to the agent with a reward/penalty $r_{t+1}$ and transitions to a new state $s_{t+1}$.

By interacting with the environment, the agent aims to maximise the total amount of reward it receives over the long run. In order to do that, the agent needs to find the *optimal policy*; a *policy* is defined as the mappings from the states to the probabilities of choosing the actions available.

During this process, the states are served as the basis of the action selection. The state is defined as a signal that summaries the information that is available to the agent. The least compact way to represent the state is a list of everything happened so far, i.e., a history of states and actions up until time step $t$. The most compact state representation is

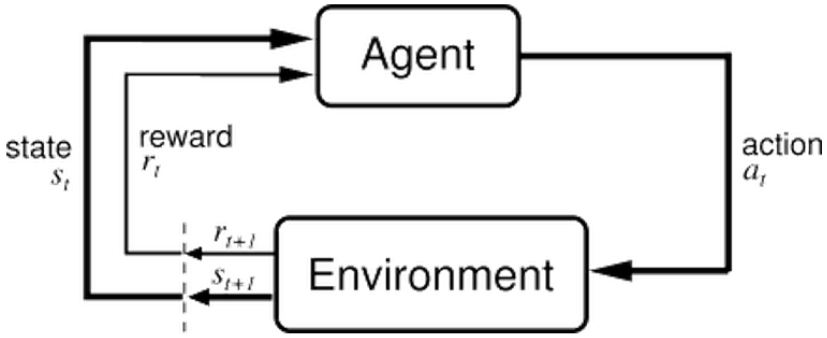



Figure 3.2: The agent-environment interaction in Reinforcement Learning (Figure from Sutton & Barto, 1998)

called a Markov state. A Markov state is a state that summarises all relevant information in the past. All that matters for the future is the current state. A common example used is 'the current position and velocity of a cannonball'. 'The current position and velocity of a cannonball' is regarded as a Markov state, because it summarises everything that is required for future flight. It does not matter how that position and velocity was obtained (Sutton & Barto, 1998). Formally, if the environment has the *Markov property*, i.e., the states are Markov states, the environment dynamics can be described by

$$Pr = \{s_{t+1} = s', r_{t+1} = r \mid s_t, a_t\} \tag{3.1}$$

instead of

$$Pr = \{s_{t+1} = s', r_{t+1} = r \mid s_t, a_t, r_{t-1}, s_{t-1}, a_{t-1}, ...r_1, s_0, a_0\} \tag{3.2}$$

As shown in Equation 3.1, the response of the environment at time $t + 1$ depends only on the previous state $s_t$, and the action chosen from that state $a_t$. In contrast, Equation 3.2 is for a non-Markov environment, where the response of the environment depends on everything that happened so far. The interaction process shown in Figure 3.2 can be modelled as a Markov Decision Process (MDP) if the states have the Markov property. A formal introduction to MDPs is given below.

### 3.3.2 Markov Decision Processes

The basic elements for a Markov Decision Process (MDP) are $(\mathcal{S}, \mathcal{T}, \mathcal{A}, \mathcal{R})$, where $\mathcal{S}$ represents the state space, $\mathcal{T}$ is a function that governs the state transitions (more details later), $\mathcal{A}$ represents the action space, and $\mathcal{R}$ is the reward function. If state space $\mathcal{S}$ and action space $\mathcal{A}$ are finite, then it is called finite a Markov Decision Process. Finite MDPs are particularly popular in reinforcement learning (Sutton & Barto, 1998).

Due to the Markov property, a MDP could be described as a one-step dynamical system as follows.

$$\mathcal{P}^{a}_{s,s'} = Pr\{s_{t+1} = s' \mid s_t = s, a_t = a\} \tag{3.3}$$

The state transition function, $\mathcal{T}$ is a function that maps the state-action pairs to the new states. The state transition function is usually represented as *transition probabilities*, $P_{s,s'}(a)$, where $s \in \mathcal{S}$, $s' \in \mathcal{S}$, and $a \in \mathcal{A}$. $P_{s,s'}(a)$ is the probability of entering state $s'$ after taking action $a$ at state $s$ at previous time step. The reward function $\mathcal{R}$ is represented as follows.

$$\mathcal{R}^{a}_{s,s'} = E\{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\} \tag{3.4}$$

Equation 3.3 and Equation 3.4 specify the dynamics of a MDP. The agent's goal is to maximise the reward over the long run given the Markov Decision Process (MDP). One way to specify an agent's behaviour is in terms of a *control policy*, which specifies what to do for each state in the state space. Formally, a *policy* $\pi$ is a function from the state space $\mathcal{S}$ to the action space $\mathcal{A}$. In reinforcement learning, the agent's objective is to learn a control policy that maximises some measure of total reward accumulated overtime. The most prevalent measure is $R_t$ (Sutton & Barto, 1998), as shown in Equation 3.5.

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ... = \sum_{n=0}^{\infty} \gamma^n r_{t+n+1} \tag{3.5}$$

$R_t$ is a discounted sum of reward received over time, where $\gamma$ is called the *discount rate*,

$0 \leq \gamma \leq 1$. The discount rate concerns how future rewards are considered for the value of current state. The closer $\gamma$ approaches to 1, the stronger future rewards are considered. $R_t$ is also called the *return*.

Now we are in a place to talk about the value of a state. The value of a state depends on how the agent behaves after that state, so the value of a state is associated with a control policy $\pi$. The value of a state $s$ under a policy $\pi$, denoted as $V^\pi(s)$, is the expected return when starting in state $s$ and following the policy thereafter as in Equation 3.6

$$V^\pi(s) = E_\pi\{R_t \mid s_t = s\} = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s\} \tag{3.6}$$

where $E_\pi$ denotes the expected value given that the agent follows policy $\pi$. We call the function $V^\pi$ the state-action value function of policy $\pi$.

A policy $\pi$ is defined to be better than or equal to a policy $\pi^{'}$ if and only if $V^\pi(s) \geq V^{\pi^{'}}(s)$ for all $s \in \mathcal{S}$. Therefore, there is always at least one policy that is better than or equal to all other policies. This is an optimal policy, $\pi^*$ as defined in Equation 3.7

$$\pi^* = \max_\pi V^\pi(s) \tag{3.7}$$

for all $s \in \mathcal{S}$. Therefore, the goal of the agent is to find this optimal policy. This is the job of a reinforcement learning algorithm.

An important property of Markov Decision Processes (MDPs) is that the optimal policy is well defined and guaranteed to exist. In particular, the Optimality Theorem from dynamic programming guarantees that for a discrete state Markov decision process there always exists a deterministic policy that is optimal. Furthermore, a policy is optimal if and only if it satisfies the following relationship:

$$Q_\pi(x, \pi(x)) = \max_{a \in A}(Q_\pi(x, a)) \qquad \forall x \in S \tag{3.8}$$

where $Q_\pi(x, a)$ is an *action-value function*, which is defined as expected return given that

the agent starts at state $x$, applies action $a$, and follows policy $\pi$ thereafter. Equation 3.8 states a policy is optimal if and only if in each state the policy specifies an action that maximises the local 'action-value'.

If an MDP is completely known (including transition probabilities and reward distributions), then the optimal policy can be computed directly using techniques from dynamic programming. However in many cases the dynamics of the environment are unknown. Under this circumstance, the optimal policy cannot be directly computed, but can be learnt by exploring the environment by trial-and-error.

In the next section, Q-learning, which can be used to find an optimal action-selection policy for (finite) MDPs, is introduced.

## 3.4 Q-learning

Pseudo code for a simple, one-step Q-learning algorithm is shown in Box 3.4, where $s \in \mathcal{S}$; $a \in \mathcal{A}$, $\alpha$ is called the *learning rate*, $\gamma$ is called the *discount factor*, which determines the importance of future rewards. Q-learning estimates the optimal state-action value function directly. A *Q-table* is used to store the state-action value function. An optimal control policy is derived from the Q-table using the greedy strategy.

At the beginning there is no control information in the $Q - table$, unless some prior knowledge about the task is available. The control information, i.e., the state-action values, in the Q-table is then updated by interacting with the environment. To do this the model is usually trained with the simulated experience of the environment, for as many trials as it requires.

After initialisation of the Q-table, e.g., uniformly zero, the model enters the learning loop. On each trial, the model starts with an initial state, $s_0$. It then selects an action $a$ based on a policy. For example, an $\epsilon$-greedy policy means that the model chooses a greedy action according the existing state-action values with probability of $1 - \epsilon$ and chooses a random action otherwise.

**Box 3.4: Pseudo code for one-step Q-learning algorithm**

- Initialize $Q(s, a)$ arbitrarily, e.g., set zero for each $(s, a)$ pair.
- Repeat (for each trial):
  - Initialise $s_0$, or randomly choose one of the states
  - Repeat (for each step of the trial):
  - Choose $a$ from $s$ using policy derived from Q-table (e.g., $\epsilon$-greedy)
  - Take action $a$, observe $r$, $s'$
  - $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \times max_{a'} Q(s', a') - Q(s, a)]$
  - $s \leftarrow s'$
- until $s$ is a terminal state

Once executed the action $a$ from a state $s$ and resulted in a transition to another state $s'$ and an immediate reward $r(s, a)$, the estimate for state-action value $(s, a)$ is then updated as follows. $U(s') = max_{a \in A}[Q(s', a)]$ is regarded as the value estimate of the next state $s'$. $r$ is the immediate reward. The estimate for the state-action value is then obtained by combining these two factors, as shown in the equation below.

$$r + \gamma \times U(s') \tag{3.9}$$

This is an unbiased estimator for $Q^*(s, a)$ when $Q = Q^*$, since that $Q = Q^*$ is defined as follows.

$$Q^*(x, a) = E[r(s, a) + \gamma V^*(T(s, a))] \tag{3.10}$$

where $V^*(s) = max_{a \in A}[Q^*(s, a)]$. The one-step estimate is combined with the old esti-

mate for $Q(s, a)$ using a weighted sum:

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha[r + \gamma U(s')] \tag{3.11}$$

where $\alpha$ is the learning rate.

For any finite Markov Decision Process, Q-learning's optimality is guaranteed on one condition that in the limit every state-action pair is tried infinitely often and if the learning rate decreases according to a proper schedule.

# Chapter 4

# An Optimal Control Model of the Distractor Ratio Paradigm

In this chapter the *state estimate and optimal control* approach, or *optimal control* approach for brevity, is applied to a visual search task. Unlike many approaches to modelling visual search, the optimal control approach requires no heuristic decision assumptions, rather behaviour is an emergent consequence of adaptation to the visual system and reward. I demonstrate that an optimal control model that is bounded by hypothesised limitations of the human visual system and that maximises reward can explain empirical evidence concerning saccadic selectivity in visual search. The model generates an estimate of the stimulus based on noisy percepts and makes optimal control decisions about where to saccade and when to respond with a target present or target absent action. The model explains saccadic selectivity as a strategic response to uncertainty in the periphery. A comparison that contrasts two models of uncertainty in peripheral vision shows that while a *spatial smearing* model predicts the magnitude of the saccadic selectivity, a *spatial swap* model does not. In the discussion, I contrast the optimal control approach with approaches that require heuristic decision rules.

## 4.1 Background

Visual search is a psychological process that involves attending to parts of a visual scene in service of a task. A typical visual search task might, for example, involve a search for the location of a named item on a kitchen table, perhaps in service of a manual task (Hayhoe & Ballard, 2005, 2014). Alternatively, it might involve detecting the presence or absence of a Gabor patch against a noisy background (Najemnik & Geisler, 2008), or it might involve a conjunction search task, such as detecting the presence or absence of a red letter X amongst a display of red and green Xs and Os (Shen et al., 2000; Wolfe, 2014).

Eye movements during visual search are required because vision is bounded by constraints on visual acuity, where acuity decreases rapidly with eccentricity from the fovea. Consequently, eye movements are essential if the high resolution fovea is to be used to gather reliable information (Geisler, 2011). Irrespective of the particular task, the key visual search questions are often (1) *why* do people choose to look where they do and (2) *what* determines when they act. The current chapter answers these questions by deriving the optimal control policy given a theory of uncertainty in peripheral vision and a theory of reward.

There are at least three different approaches to modelling eye movements. In the first are what Kowler (2011) describes as *map-based approaches*, such as salience maps (Itti, 2007; Itti & Koch, 2000) and activation (or priority) maps (Pomplun, Reingold, & Shen, 2003; Wolfe, 2007), where information is accumulated and processed to produce a map. Peaks in the map represent areas/items that differ from their surroundings, that contain attributes of the target, or both. Map peaks are used to guide search through the display using some peak selection rules, such as a greedy heuristic (Pomplun et al., 2003) or a winner-take-all heuristic (Itti & Koch, 2000; Wolfe, 2014). In general, the map-based approach assumes that saccades are programmed to move the fovea to areas of a stimulus that stands out from its surroundings.

In the second approach, *optimal state estimation* (Najemnik & Geisler, 2008; Myers,

Gray, & Sims, 2011), it is assumed that saccade programming is informed by a Bayesian estimate of the current state of the world. The state estimate is calculated by optimally integrating information from the fovea, from the periphery, and from previous fixations according to its reliability. The reliability of information is determined by physiological and neural information processing constraints. However, the optimal state estimate must be complemented with search rules and/or stop rules. The search rules for the MAP (Maximum A Posterior) result in the searcher always fixating the location with maximum posterior probability. This is more or less like a 'winner-take-all', or a 'greedy' heuristic. One example of how MAP fails to explain human data is the 'centre of gravity effect'. For example, at one update, the posterior gives a set of probabilities of target locations with two close peaks; one is slightly higher than the other. If using MAP (the greedy heuristic), the model would go to one of the slightly higher peak. An optimal control approach, in contrast, would learn to go to a central point between these two peaks, as it in a way reduces the uncertainty for both peaks. This point was also made by Najemnik and Geisler (2005, 2008). A reasonable search rules might be to direct attention to areas of uncertainty so as to achieve a reduction in uncertainty that would not be possible from the current fixation position. Najemnik and Geisler (2005) found that the number, and spatial distribution, of saccades to find a target could be described by a model in which each saccade was directed to the ideal location (i.e., a location that maximises information gain for the target's location). Their model was sensitive to known human constraints on vision (e.g., decreasing contrast sensitivity with increasing eccentricity from the fovea). However, this approach fails to meet the criteria of full optimal control for two reasons. First, the search stopped where the maximum posterior probability exceeded a criterion, which is picked by matching the error rate of the human observers. This means that the stop rules are not learnt by optimal control. Secondly, as pointed by Najemnik and Geisler (2005, 2008) themselves, the ideal searcher is optimal because it only considers one fixation into the future.

In the third approach, *optimal control* (Trommershäuser, Glimcher, & Gegenfurtner,

2009; Sprague, Ballard, & Robinson, 2007; Hayhoe & Ballard, 2014; Nunez-Varela & Wyatt, 2013), it is assumed that saccades, and other actions, are directed so as to maximise reward. In this approach the assumption is that the purpose of vision is not to form the best possible estimate of the world but rather to determine the best choice of action. The optimal control approach is distinguished from map-based approaches and from optimal state estimate approaches because control/guidance optimises a reward signal rather than optimises some property of the state. The maximum reward attainable by an individual will of course be bounded by the level of noise during the encoding of visual information, just as it is in the optimal state estimation approach. However, unlike optimal state estimation, the optimal control approach does not require heuristic control; rather, the control policy is derived given assumptions about the visual system. Policies can be derived through the use of a reinforcement learning algorithm (Sprague et al., 2007; Hayhoe & Ballard, 2014), though it is possible that the policy may be acquired by other learning mechanisms, for example, by cultural transmission, through instructions, or by evolution. Reinforcement learning algorithms have been proposed both as means of explaining human learning processes (Dayan & Daw, 2008) and as means of deriving rational analyses of what a person should do in particular task environments (Chater, 2009). In the latter case the derivation of an optimal control policy given a theory of the perceptual constraints might provide an explanation for why visual search behaviours, such as saccadic selectivity (Shen et al., 2000), are computationally rational (Lewis et al., 2014; Howes, Lewis, & Singh, 2014; Howes et al., 2009).

In the current chapter I build on the optimal control approach by finding a policy that maximises reward given a bounded visuo-cognitive system[1]. The behaviour of the optimal policy demonstrates trial-by-trial adaptation to changes in the ratio of available environmental features. More specifically, the optimal control approach is used to explain phenomena associated with the distractor ratio paradigm (Bacon & Egeth, 1997; Shen et al., 2000; Zohary & Hochstein, 1989); phenomena that have previously been given

[1]We refer to 'optimal control' so as to contrast the approach to optimal state estimation. However, the solutions that we are able to find with function approximation are only asymptotically optimal.

interpretations in terms of map-based approaches.

The optimal control model has a simple structure that decomposes visual search into a *(non-optimal) state estimation* mechanism and *optimal control*. The state estimation integrates perceptual evidence into a task-relevant representation of the external stimulus. The optimal control analysis determines overt task responses and information gathering actions given the state in order to maximise the utility (Stengel, 1994; Baron & Kleinman, 1969a). The structure of the optimal control approach affords the formulation and exploration of a number of interesting theoretical questions concerning visual search phenomena. In this chapter, I use modelling to identify the constraints of the visual system that are sufficient for saccadic selectivity effects to arise.

The model minimises assumptions about the visuo-cognitive system and shows how behaviours can be derived from adaptation to these assumptions. The small number of assumptions are intended to abstractly characterise the key properties of a foveated visual system in which uncertainty increases in the periphery (Swets, Tanner Jr, & Birdsall, 1961). The model assumes two sources of uncertainty. The first is *feature noise* and the second is *spatial noise*. Feature noise is a source of uncertainty that leads to the reduction in acuity of items in the periphery according to a hyperbolic function (Strasburger, Rentschler, & Jüttner, 2011). Feature noise negatively impacts visual acuity of a letter irrespective of the spatial surrounds of the letter. *Spatial noise* leads to locational uncertainty through the combination of incompatible information from adjacent items in the stimulus (Yu, Dayan, & Cohen, 2009). A comparison that contrasts two models of *spatial noise* in peripheral vision is done in this chapter. To foreshadow the results, the model demonstrates that saccadic selectivity in the distractor ratio paradigm is a signature of optimal control in the face of *spatial smearing* in the periphery. A comparison to a model that is subject to *spatial swapping* uncertainty shows that while uncertainty due to *spatial smearing* predicts the same level of saccadic selectivity observed in humans, the *spatial swapping* model does not.

In the following sections I first introduce the distractor ratio paradigm and phenomena

and then introduce the theory and optimal control model. I next describe the results and discuss their implications.

## 4.2 Distractor Ratio Paradigm

The distractor ratio paradigm is a task in which participants are asked to search for a target amongst a number of distractors that vary across two or more feature dimensions (e.g., colour and orientation). A typical stimulus for the task is shown in the left panel of Figure 4.1, the target is the red letter O (red-O). The distractors either share the same colour with the target or share the same shape with the target. The result is that people find the target more quickly and with fewer eye movements when there is an extreme ratio of same-colour to same-shape distractors. For example, in Figure 4.1, it can be seen that the target, red-O, can be found quickly in Figure 4.1A (left panel) and Figure 4.1C (right panel) but is relatively difficult to find it in Figure 4.1B (middle panel). The term *minority set* is used to refer to either shape or colour depending on its numerosity.

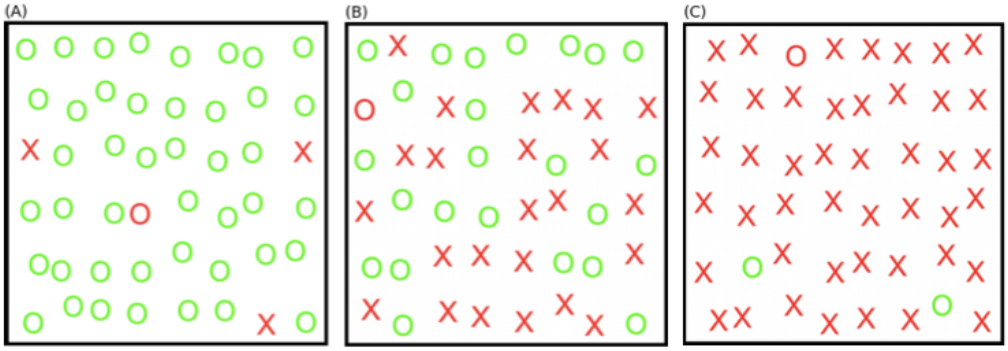



Figure 4.1: The distractor ratio paradigm. The goal is to determine whether a red-O is present or absent. The 'distractor ratio' is the number of distractors that are of the same colour relative to the number of distractors that are of the same shape. Figure (A) same-colour:same-shape=3:45; (B) same-colour:same-shape=24:24; (C) same-colour:same-shape=46:2

Several studies have demonstrated this effect (Bacon & Egeth, 1997; Zohary & Hochstein, 1989; Shen et al., 2000; Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & Van der Heijden, 1995; Poisson & Wilkinson, 1992). Specifically, the detection is relatively easy

for displays with extreme distractor ratios (i.e., either the same-colour or same-shape distractors are rare), but relatively difficult for displays in which the two types of distractors are equally represented (Bacon & Egeth, 1997). The distractor ratio is defined as the ratio of the number of different types of distractors. Figure 4.1A (left panel) has a ratio of 3:45; Figure 4.1B (middle panel) has a ratio of 24:24; Figure 4.1C (right panel) has a ratio of 46:2.

Results from Shen et al.(2000)'s study are shown in Figure 4.2. For both target present and target absent the search time and the number of fixations are lowest when the ratio is extreme (toward the ends of the x-axis) and they increase as the ratio approaches 1(the centre of the x-axis). This effect of distractor ratio is greater for target-absent than for target-present trials. Both effects are significant and there is an interaction.

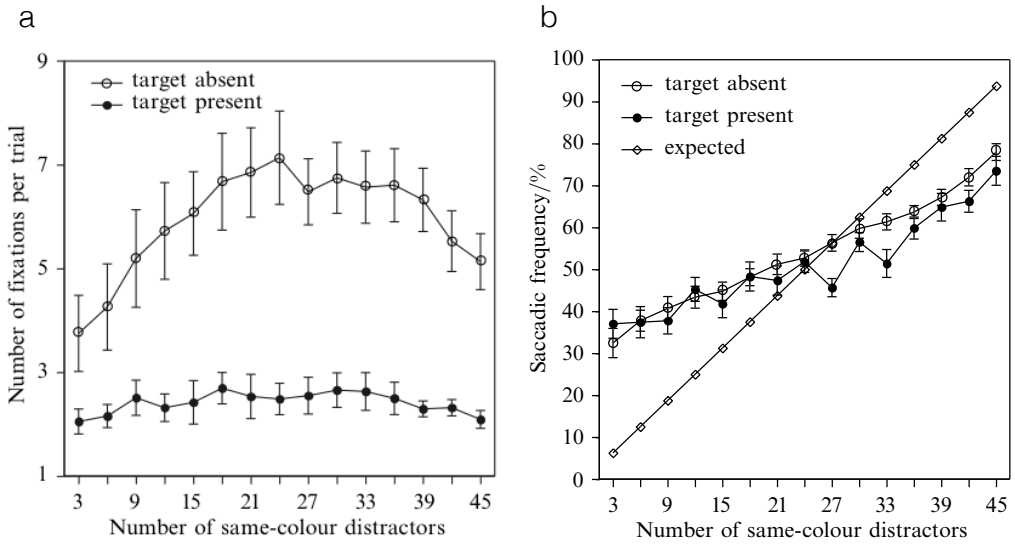



Figure 4.2: Empirical results from Shen et al. (2000) with 95% C.I.s. (a) The distractor ratio effect is evident in the mean number of fixations per trial plotted against the number of same-colour distractors. (b) The saccadic frequency (a measure of saccadic selectivity) is plotted against the number of same-colour distractors.

In addition to an effect on the overall number of saccades, studies have demonstrated an effect of *saccadic selectivity*, which is shown in Figure 4.2b. In Figure 4.2b, the frequency of saccades to the same-colour distractors is plotted against the number of same-colour distractors. It can be seen that saccades are to the minority set of features (whether

colour or shape) with a higher frequency than would be expected by chance (i.e., the line labelled as 'expected'). Specifically, when the number of same-colour distractors is in the minority (where x-axis$< 24$), the saccadic frequency to the same-colour distractor is higher than the expected. When the number of same-colour distractor is in the majority (where x-axis$> 24$), meaning that the same-shape distractors are the minority, the saccadic frequency to the same-colour distractor is low than the expected, which means that the saccadic frequency to minority distractor is higher than the expected. When the number of same-colour distractors is low then participants were more likely to saccade to same-colour distractors and when the number of same-colour distractors was high participants were more likely to saccade to same-shape distractors.

An important feature of the DR paradigm is that minority set of distractors have the same number of target feature matches as the majority set. A consequence of this is that there is no set bias: if the target is present it is in the majority and minority sets in all trials. The set itself is therefore not a cue to the location or presence of the target. Additionally, when a target is absent from a trial, all items have an equally good match to the target irrespective of set.

One of the reasons that the distractor ratio task is interesting is that there are a number of theoretical accounts. For example, saccadic selectivity in the distractor ratio paradigm has been hypothesised to result from a salience-map based approach (Theeuwes, 1991; Shen et al., 2000). The theory was that as a feature becomes the minority set it becomes increasingly physically conspicuous and draws attention and saccades to the locations. Another approach has been to describe saccadic selectivity as a consequence of optimal state estimation. The optimal state estimation approach computes the maximum *a posteriori* (MAP) likelihood that a target is in a particular location given a set of noisy perceptual samples (Myers et al., 2013; Najemnik & Geisler, 2008). Saccades to the MAP likelihood are then preferred. However, both of these approaches require heuristic assumptions about the control, assumptions that I show how to avoid with the optimal control approach.

## 4.3 Theory

The theory presented here is that the saccadic selectivity observed in the DR task is an optimal response to reward and *spatial noise* in peripheral vision. Two theories are defined, each of which is based on a model of the *spatial noise*. These two models are referred to as *spatial smearing* and *spatial swap*. With the exception of the smearing/swap assumption all other assumptions of the two theories are the same. The spatial smearing based theory is introduced first, followed by the spatial swap based theory.

### 4.3.1 Theory 1: spatial smearing based theory

Spatial smearing combines the shape information from each letter in the scene with the shape information from adjacent letters (Yu et al., 2009). Here I assume that it does the same to the colour of the letter, combining information suggesting a red colour with information suggesting a green colour, for example, where those items are adjacent. A consequence of spatial smearing is that, appearing in the periphery, a viewer of a red-X surrounded by green-Os will be uncertain as to whether the red-X really is an X or a green-O because information from the 8 surrounding Os will be mixed with information from the X. In contrast, when a letter is surrounded by the same letters in the same colour then there is less uncertainty about what the letter is. For these reasons, a viewer is less certain about the colour and shape of minority set items in peripheral vision than they are about the colour and shape of majority set items in peripheral vision.

Spatial smearing has been proposed as a potential mechanism behind crowding effects (Yu et al., 2009). The theory was motivated by the observation that the neurons that process letters have large receptive fields so that a stimulus at any one point will potentially lead to activation of receptive fields covering a number of adjacent points. Yu et al. (2009) therefore proposed that the probability of correctly encoding a stimulus was not only a function of the stimulus itself but also a function of its neighbouring stimuli. Specifically, they assumed that the response to a stimulus was a weighted sum of the response of the

populations of neurons responsible for the stimulus itself and its adjacent letters. Yu et al. (2009) demonstrated that this spatial uncertainty model could account for performance on a task in which a foveated letter, e.g., 'S', has either compatible 'S' letters (flankers) or incompatible 'H' letters to each side. In this flanker task participants have been shown to be less accurate at identifying the foveated stimulus when it is flanked by incompatible letters Eriksen (1995). This type of neural interference has also been used in accounts of the size and spacing requirements for letter identification in peripheral vision (Song, Levi, & Pelli, 2014). Spatial smearing is related to the observation of feature *mixing* reported by Golomb, L'Heureux, and Kanwisher (2014). They showed that when participants were presented with four colours after a saccade, one of which was in a precued target location, they would report a colour systematically shifted in colour space towards the colour of the retinotopic distractor. Subsequent experiments revealed similar effects after split visual attention.

The theory of spatial smearing fits well with feature integration theory (A. M. Treisman & Gelade, 1980) as well as with Rensink (2000)'s proposal of proto-objects. In feature integration theory illusory conjunctions are theorized to occur when items containing different features are adjacent. Similarly, preattentive items, or *proto-objects* have limited spatial coherence and may lead to incorrect spatial assignment for different features in adjacent objects due to the volatility of feature information, also potentially leading to illusory conjunctions. Spatial smearing provides one potential way of computationally producing illusory conjunctions or feature location uncertainty.

The theory then is that, because of spatial smearing of both letter shape and colour, a viewer will experience elevated uncertainty for any distractor ratio letter that is in the periphery and surrounded by letters with a different shape and/or colour. Spatial smearing will therefore mean that viewers will experience higher uncertainty for minority set items and may adaptively saccade to these items in order to resolve uncertainty. In addition, a viewer will also experience higher uncertainty overall with stimuli that have a distractor ratio closer to 1 than with an extreme distractor ratio. This is because with a distractor

ratio close to 1 many red-Xs and green-Os are mixed together in the stimulus. Informally, the theory therefore predicts that participants should take longer to determine that a target is absent when the distractor ratio is close to 1 because there is greater uncertainty in the scene.

The challenge for the rest of the chapter is to formalise the above intuition as a computational model in which the behavioural phenomena emerge given these very simple assumptions about reward and spatial smearing in peripheral vision. The hypothesis of the optimal control approach is not that participants saccade to uncertainty; it is that control optimisation to maximise reward rate leads participants to saccade to minority set items and to generate more saccades, overall, for distractor ratios close to 1. The saccades help maximise reward rate by resolving shape and colour uncertainty caused by peripheral smearing.

### 4.3.2 Theory 2: spatial swap based theory

An alternative model of the spatial noise in low level vision is described below. The model, called the *spatial swap* model, is motivated directly by the fact of illusory conjunctions (A. Treisman & Schmidt, 1982; Wolfe, 2014) and also by Golomb et al. (2014)'s study of feature binding errors after eye movements and shifts of attention in which they found evidence for both spatial swapping and feature mixing errors. The key idea in this model is that spatial uncertainty is a direct consequence of entirely changing a feature at one location into one in an adjacent location. These spatially defined feature swaps occur when the feature of one object is probabilistically sampled from the features from its proximal objects. This happens probabilistically in the model such that for an object closer to the fovea, the feature is more likely sampled from its true location and less likely to be sampled from a distant part of the stimulus. The further away from the fovea and the smaller the distance between adjacent items, the more likely it is that the feature of the object is sampled from neighbouring parts of the stimulus. Therefore, instead of a gradient mixture of feature information across adjacent locations, as is the case in the smearing

model, the spatial swap model results from changed features between adjacent items, resulting in complete illusory conjunctions. Another way to describe the spatial swap model is that all of the feature information at one location is transposed to an adjacent location.

A key property of the spatial swap model is that the sampling probabilities for the spatial noise follows a Gaussian fall-off, so over many samples, the information from each location should match the weighting given by the Gaussian filter in the spatial smearing model. In other words, over many trials both models generate the same distribution of spatial noise for a given location in a given stimulus. However, on any individual trial and at any individual location, where the spatial swap model tends to lead to a higher probability of an illusory conjunction, the spatial smearing model tends to lead to an increased probability of uncertainty as to whether either feature is present. While the spatial swap model tends to lead to either a randomly alternating percept of red or green, the smearing model tends to lead to a continuous percept that is somewhere between red and green.

Although the spatial swap model seems intuitively appealing – because it leads directly to illusory conjunctions – illusory conjunctions may in fact be a feature of memory, not of low level vision (Wolfe, Reinecke, & Brawn, 2006; Henderson & McClelland, 2012). A. Treisman and Schmidt (1982) observed Illusory conjunctions when people were asked to report what they *remembered* seeing in a display that had subsequently been removed. Henderson and McClelland (2012) concluded that without a dual task procedure, the number of illusory conjunctions was significantly fewer than would be expected by chance, suggesting that illusory conjunctions may be related to issues of memory load and response selection. It is therefore not at all obvious that illusory conjunctions are an important limitation on a visual search task but a contrast between spatial smearing and spatial swapping models may reveal more about its role in visual search.

## 4.4 The optimal control model

An overview of the optimal control model is presented in Figure 4.3. The model represents the key processes and representational structures for a 9-item distractor ratio task rather than the full 48-item task but the same structure is used for stimuli of any dimensions. The stimulus (on the bottom left of the figure) is encoded by noisy perceptual processes to generate colour and shape percepts (not represented) which are integrated into a relevance estimate. The new relevance estimate is then integrated with the previous state into a new state representation. It thereby integrates information across the time steps. Given the new state, the control policy (the upper right) guides action selection which either chooses to saccade, maintain fixation, or respond with 'present' or 'absent'. The initial policy has no control information and results in random saccades and decisions. However, through reward guided control optimisation, the policy converges on the optimal control policy.

Now examine the state in Figure 4.3 a little more closely. The state represents the outcome of two fixations. The first fixation, represented by a black plus sign, was in the bottom right of the stimulus and resulted in the state labelled 'previous state' (at the bottom left of the figure). As a consequence the previous state contained the three green-Os in the bottom right of the stimulus.

The control policy then selected to saccade to the centre of the stimulus where a red-X was perceived. The relevance estimates from both fixations encoded uncertainty about whether the stimulus contained red-Xs or green-Os in the top-left corner due to both feature noise and spatial smearing, which mixed features of the red-X and surrounding green-Os. The result of these two fixations is integrated into a single state representation where the central red-X and bottom-right green-Os are represented with a high degree of certainty.

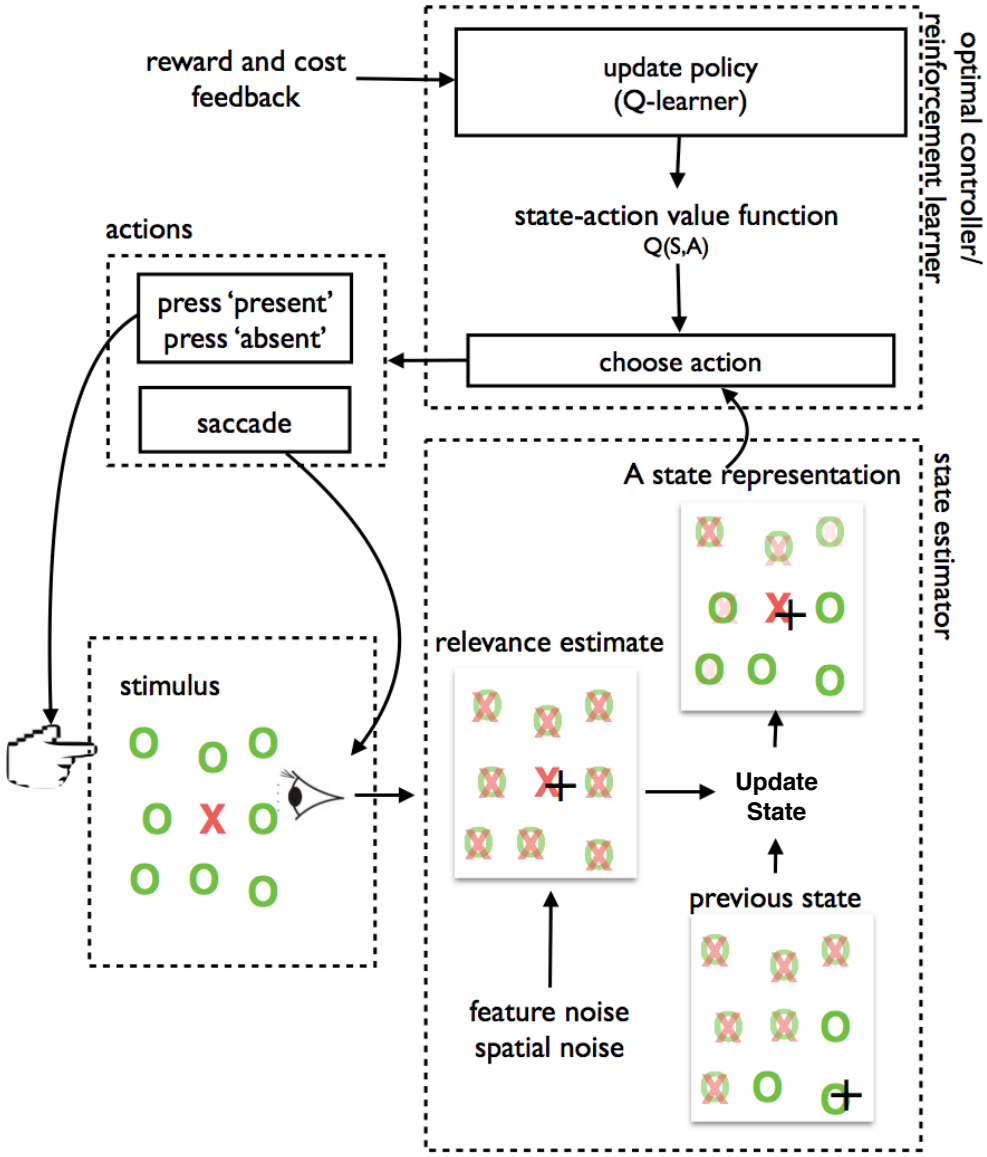



Figure 4.3: An overview of the optimal control model. Given the stimulus on the bottom left, perception encodes noisy colour and shape information. A relevance estimate is generated based on the noisy colour and shape percepts, which is then integrated with the previous relevance estimate (the bottom right). The symbol '+' represents the fixation location. The control policy (the upper right) guides action selection which either chooses a saccade (perhaps a null saccade which maintains the current fixation) or chooses a present/absent response (the upper left). Through reward/feedback guided control optimisation, the policy converges on the optimal control policy. All predictions are generated using the optimised control policy and greedy action selection.

### 4.4.1 State estimation

The stimulus (the real display) is represented as two binary matrices, $m_c$ and $m_s$, one for colour and the other for shape. In each of these matrices, $1$ represents the target feature; $0$ represents the non-target feature. A colour percept $p_c$ and a shape percept $p_s$ are then generated by corrupting the external stimuli, $m_c$ and $m_s$, according to the models of *spatial noise* and *feature noise*. The colour and shape percepts, $p_s$ and $p_c$, are then integrated to form a relevance estimate $R$. This relevance estimate is further integrated with the relevance estimate accrued from previous fixations to form the current *state*.

The state estimation process described above is used for both the *spacial smearing* model and the *spatial swap* model. A 5-step state estimation walkthrough is given below. With the exception of the smearing assumption all other assumptions of the two models are the same, these two types of the spatial noises are described in one walkthrough. The readers need to keep in mind that only one type of the spacial noise is chosen for one of these two models.

#### Step 1: True display

The true display is represented as two binary matrices, one representing colour $m_c$ and the other one representing shape $m_s$. In both $m_c$ and $m_s$, $1$ represents the target feature; $0$ represents the non-target feature. For example, if the target is a Red X, then a Green X at location $[i, j]$ is represented as $m_c[i, j] = 0$ and $m_s[i, j] = 1$.

#### Step 2 (spatial noise1): spatial smearing

The spatial smearing noise is added to these two true-display matrices, $m_c$ and $m_s$, independently. After adding the spacial noise we have two percepts $\mu_c$ and $\mu_s$. Specifically, for each location $[i, j]$ of the matrix $m_f(f \in c, s)$ we first calculate a $\mu_f[i, j](f \in c, s)$. $\mu_f[i, j]$ is a weighted sum of $m_f[i, j]$ and its 8 adjacent locations. The weighting function is a Gaussian kernel with a standard deviation that is a function of the visual angle between the fovea and $[i, j]$. This means that how the percept of the item at location $[i, j]$

is affected by the neighbouring items depends on how far the location $[i, j]$ is away from fovea. The standard deviation of the Gaussian kernel is $\sigma(\theta, \delta_s)$, which is represented in Equation 4.1 below with $\delta = \delta_s$. In the equation, $\theta$ is the visual angle between the fovea and $[i, j]$, and $\delta = \delta_s$ is used to scale the effect of the visual angle to the fovea.

$$\sigma(\theta, \delta) = 1 - acuity(\theta, \delta) \tag{4.1}$$

$$acuity(\theta, \delta) = \begin{cases} (\delta \times \theta)^{-0.29} - 0.32, & \text{if } \theta \neq 0. \\ 1, & \text{if } \theta = 0. \end{cases}$$

As $\theta$ increases the acuity decreases and the standard deviation $\sigma$ of the Gaussian kernel increases, which means that the percept of the item at $[i, j]$ suffers greater interference from surrounding items (Strasburger et al., 2011).

The parameters in Equation 4.1 are from Osterberg (1935), where they were estimated based on cone and rod density on the retina. According to this function the acuity is 1.0 in the fovea and falls to 0.5 at 5 degrees of eccentricity from the fovea. There is evidence suggesting that acuity could diminish with eccentricity differently for different features, e.g., colour, shape and size (Kieras & Hornof, 2014; Anstis, 1974; Johnson, 1986). While the model includes a parameter to weight the colour and shape percepts I did not explore it here. Here I assume that both colour and shape are subject to the same acuity function.

**Step 2 (spatial noise2): spatial swap**

For the spatial smearing model above, a weighted sum of the neighbouring items, $\mu_f[i, j]$, where $(f \in c, s)$, is calculated for each location $[i, j]$ based on the Gaussian kernel. This means that the percept feature for location $[i, j]$ is a gradient mixture of feature information across adjacent locations. In contrast, for the spatial swap model, the spatial noise is represented as stochastic changes of a feature value at one location into a value of an adjacent location. In other words, spatial swap noise was added to the models percept by allowing the feature value for each position to be sampled from nearby positions with

some probabilities. These probabilities are sampled from the same Gaussian kernel with the standard deviation shown in 4.1, which are dependent on the distance between the location $[i, j]$ and the fovea and the distance between the surround location to the location $[i, j]$. The closer the location $[i, j]$ to the fovea, the smaller the standard deviation of the gaussian kernel is and the more concentrated theses probabilities toward the location $[i, j]$, which means that the more likely that the feature of the location $[i, j]$ is sampled from the value of the location $[i, j]$. After the spatial swap, each location $[i, j]$ has a percept feature value 0 or 1 that is probabilistically sampled from the item itself or its surrounding items. To be consistent with the spatial searing model, we used the same label $\mu_f[i, j]$. Therefore we have $\mu_c$ and $\mu_s$ to represent two percepts obtained after the spacial swap.

**Step 3: Feature noise**

Remember that the spatial noise is added to the real display matrices $m_c$ and $m_s$ independently. Now we have two percept values $\mu_c[i, j]$ and $\mu_s[i, j]$ for each location $[i, j]$ after the spatial noise. In this step, we are adding the feature noise. This reflects the degradation of acuity with eccentricity. Two *percept* matrices corrupted by both spatial noise and feature noise are obtained after this step, $p_c$ for colour and $p_s$ for shape. Each cell of these two percepts, $p_c[i, j]$ and $p_s[i, j]$, is generated by adding the feature noise to $\mu_c[i, j]$ and $\mu_s[i, j]$ respectively, as seen in Equation 4.2 below.

$$p_{c,s}[i,j] = \begin{cases} min(max(0, \mu_{c,s}[i,j] - \mathcal{N}(0, \sigma(\theta, \delta_f))), 1), & \text{if } \mu_{c,s}[i,j] > 0.5 \\ min(max(0, \mu_{c,s}[i,j] + \mathcal{N}(0, \sigma(\theta, \delta_f))), 1), & \text{if } \mu_{c,s}[i,j] \leq 0.5 \end{cases} \tag{4.2}$$

where $\mathcal{N}(0, \sigma(\theta, \delta_f))$ is the acuity equation shown in Equation 4.1 with $\delta = \delta_f$; $\sigma$ is a function of the visual angle $\theta$ between the fovea and $[i, j]$, and $\delta_f$ is a weight that is used to scale the effect of the visual angle to the fovea.

Each percept of $p_c$ and $p_s$ is now represented as a matrix of values ranging from 0 to 1. A consequence of the noises is uncertainty in whether the content of the location is red

or green, or whether it is a letter O or a letter X. The extreme values, 0 and 1, represent strong evidence that the feature is either red or green or O or X, while 0.5 represents the absence of evidence in favour of either feature value.

**Step 4: Combining colour and shape percepts**

The noisy percepts for colour and shape, $p_c$ and $p_s$, are combined to generate a relevance estimate $R$. Each value of the matrix, $R[i,j]$, is defined as the *Euclidean distance* between points $(p_c[i,j], p_s[i,j])$ and $[1,1]$ (the target item). Here I assume that colour and shape are equally weighted in calculating the relevance score, i.e., $k_c = k_s = 1$. Specifically, colour and shape are combined using Equation 4.3 to give $R[i,j]$.

$$R[i,j] = \left| \sqrt{k_c \times (p_c[i,j] - 1)^2 + k_s \times (p_s[i,j] - 1)^2} \right| \tag{4.3}$$

**Step 5: Integration across fixations**

On each fixation $t$, the model obtains a relevance estimate, $R_t$, as described in the steps 1-4. These relevance estimates, $R_t(t = 1, 2, ...n)$, are then integrated across fixations to get the state $S_n$. Each element of the state $S_n[i,j]$ is an estimate of the letter-colour at location $[i,j]$ up until time step $n$, except the last element which represents the current fixation location.

The information integration process is described below. On one fixation $t$, we have the previous estimate $S_{t-1}[i,j]$ and its uncertainty $\delta_{t-1}[i,j]$. $S_{t-1}[i,j]$ is updated given a new observation $R_t[i,j]$ with measurement noise, $v_t[i,j]$. The measurement noise $v_t[i,j]$ is a compound product of the feature noise and spatial noise mentioned above. The noise is assumed to be white, Gaussian noise, i.e., $v_t[i,j] \sim \mathcal{N}(0, P)$. The standard deviation $P$

is related to the acuity function defined by Equation 4.1 with $\delta = \delta_f$.

$$\begin{cases} S_t[i,j] = S_{t-1}[i,j] + K_n(R_t[i,j] - S_{t-1}[i,j]) \\ \delta_t[i,j] = (I - K_n)\delta_{t-1}[i,j] \\ K_n = \dfrac{\delta_{t-1}[i,j]}{\delta_{t-1}[i,j] + P} \end{cases} \tag{4.4}$$

### 4.4.2 Reward

It is assumed that people adapt to a subjective internal reward function which involves maximising the balance between speed and accuracy. The reward function used in Q-learning was a $-1$ penalty for each step, a $+10$ reward for a correct response; and $-10$ penalty for an incorrect response. The optimal policy derived is the state-action mapping that maximises the reward. This means that an extra step will not be 'necessary' if it does not increase the accuracy by 10%. The model achieved 96% accuracy, compared with 98% of the participants. The choice of the reward function was guided by the fact that the number of fixation that people used was from 2 to 8.

### 4.4.3 Optimal control

Given the state provided by the estimation model above the *control policy* chooses the next action.[1] The space of possible actions consists of an action for each fixation location plus two response actions: 'present' and 'absent'.

The optimal control policy is learnt by one reinforcement learning algorithm, Q-learning, for more details refer to section 3.4. The optimal policy is determined by selecting the best action in each state based on the state-action value function.

[1] I use a function approximation approach by gridding the colour and shape values provided by the state estimation model.

### 4.4.4 Summary

In summary, the optimal control policy predicts a sequence of saccadic eye movements and a present/absent decision given a distractor ratio stimulus. The policy is the optimal policy for an MDP defined by states that are informed by a perceptual function that is either subject to spatial smearing or swap and a reward function that defines the trade-off between speed and accuracy. Two theories were defined. With the exception of the smearing/swap assumption all other assumptions of the two theories are the same. Therefore, any difference in the correspondence of the behaviour of the models to human performance will provide an indication as to the validity of this theoretical assumption. In what follows I will compare these two theories by deriving the optimal control policy for both and then measuring its behaviour.

## 4.5 Method

The visual angle between the centres of neighbouring items was set to 2 degrees (to capture the materials in Shen et al., (2000)). Spatial noise was the only free parameter explored. As mentioned above, the level of spatial smearing is defined by the rate at which the variance of a Gaussian kernel, applied at each location, increase with eccentricity, see parameter $\delta_s$ in Equation 4.1. 5 levels of $\delta_s$: $[0.01, 0.5, 1, 1.5, 2]$, was explored; the optimal policy was derived for each. The performance of the best fitting model is reported.

## 4.6 Results

Both the spatial smearing and the spatial swap models were tested on the full 48-item and a 9-item version of Shen et al. (2000) distractor ratio task. In the following, the results for the 48 letter task are presented in section 4.6.1. In particular, I first give the results for the spatial smearing model, showing that the results correspond well with human data. In section 4.6.2, I present results for the spacial swap model, which shows that the saccadic

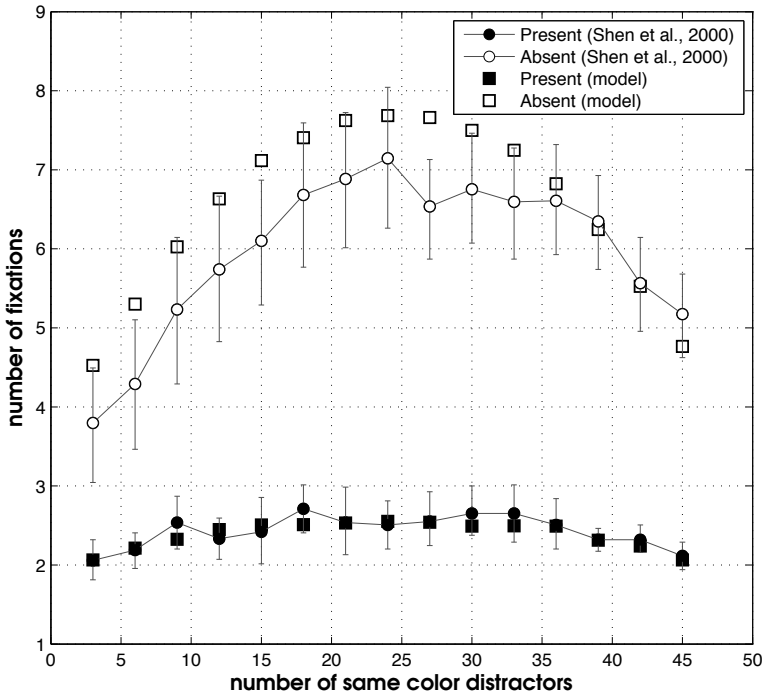



Figure 4.4: The number of fixations required by the optimal control policy for each level of distractor ratio.

selectivity predictions of the model are significantly less good than the spatial smearing model. In section 4.6.3, the model comparison is shown for a 9-letter task. The advantage of the 9-item variant of the task is that it allows a more comprehensive exploration of the parameter space. I firstly show that throughout its parameter range, the spatial swap model is unable to account for the magnitude of human saccadic selectivity. In contrast, for the spatial spearing model, it does predict a saccadic selectivity that is consistent with human performance.

### 4.6.1 Results for the 48 letter task

Q-learning converged on an approximately optimal policy within 10 million trials and it achieved an accuracy of .96 in its last 10000 trials. This is the proportion of trials on which it correctly determined whether the target was present or absent. In comparison people achieved an accuracy of .98 (Shen et al., 2000).

The number of fixations taken by the optimal policy is shown in Figure 4.4. This

shows that the model predicts the same distractor ratio effect as humans do for both the target present trials and for the target absent trials $R^2 = .93$. In both cases more fixations are required for ratios close to 1 (number of same colour distractors = 24). In addition, the model predicts the interaction between target present and target absent stimuli, with more fixations and a more accentuated distractor ratio effect for target absent stimuli.

Also, in Figure 4.4 it can be seen that the human, but not the model, the target absent DR effect is asymmetric. People require more fixations to determine that the target is absent when there are more same colour distractors. As mentioned, while the model does include a parameter to weight the colour and shape percepts I did not explore it here.

The model's prediction of the effect of distractor ratio on saccadic selectivity is shown in Figure 4.5, along with the human data. The left panel shows the human data and model predictions for target present saccadic selectivity and the right panel shows the human data and model predictions for target absent saccadic selectivity. The goodness of fit between the model's saccadic selectivity prediction and the human data was $R^2 = .72$. While the absolute prediction of the target present saccadic selectivity (left panel) is good for the majority of the distractor ratio range, the prediction is less good at extreme distractor ratios. The same is true for the target absent distractor ratios (right panel) but, in addition, here the model over predicts the distractor ratio bias throughout the high colour distractor ratio range. I suspect that this asymmetry is again due to the differences in acuity, for humans, between colour and shape, which were not modelled.

### 4.6.2 Model comparison

In this section I contrast the performance of the spatial smearing model with a model in which uncertainty in the periphery causes spatial swaps and therefore illusory conjunctions. While the effect of distractor ratio on the number of fixations was very similar to the spatial smearing model and correlated strongly with the human performance, the saccadic selectivity predictions were less good. The saccadic selectivity of the model is shown in Figure 4.6; in the figure it is clear that the predicted saccadic selectivity is much

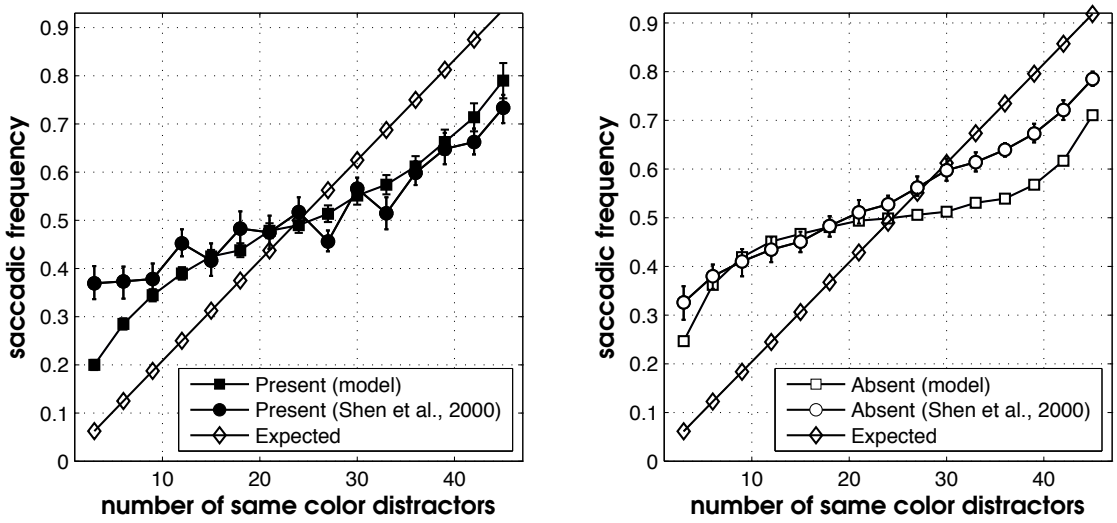



Figure 4.5: Saccadic selectivity against distractor ratio level for the spatial smearing model applied to the 48 letter task. The left panel shows the human data and model predictions for target present and the right panel is for target absent.

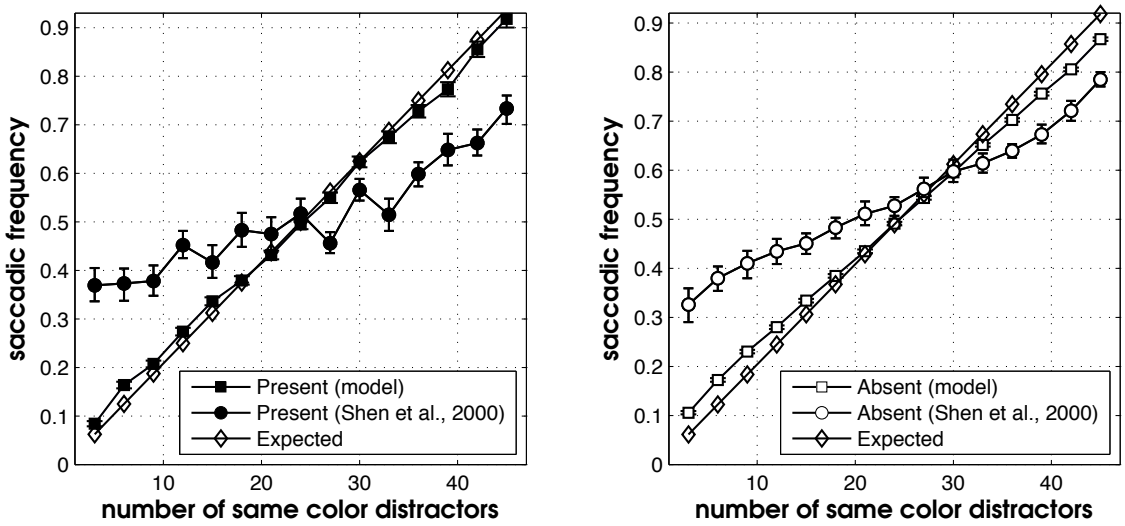



Figure 4.6: Saccadic selectivity against distractor ratio level for the spatial swap model applied to the 48 letter task. The left panel shows the human data and model predictions for target present and the right panel is for target absent.

lower than the observed on both the target present and target absent trials. The maximum saccadic frequency predicted in either condition was under 8% and the proportion of variance predicted was $R^2 = 0$.

### 4.6.3 Results for spatial swap model performance on a 9 letter task

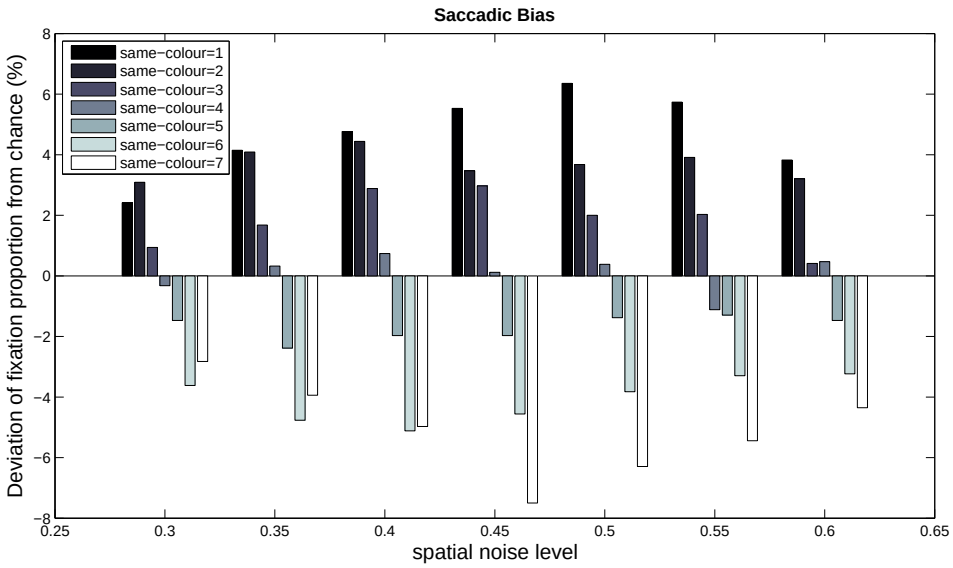



Figure 4.7: Saccadic selectivity against distractor ratio level and noise level for the spatial swap model variant applied to the 9 letter task.

In order to further test the robustness of the spatial swap model comparison, further tests using a 9 letter variant of the distractor ratio task was conducted. The purpose is to validate whether the spatial swap model would be unable to account for the magnitude of human saccadic selectivity throughout its parameter range. The benefit of using the 9 letter variant was that simulations could be run over a larger space of possible parameters in a reasonable amount of time. While Q-learning is guaranteed to converge on an optimal policy (Sutton and Barto, 1998), it does so very slowly.

The optimal policies were derived for 7 levels of spatial swap noises. The model was run for 100,000 trials for each noise level, which was sufficient to obtain asymptotic performance. It achieved an accuracy of .99. The results are shown in Figure 4.7. The figure plots the saccadic selectivity at 7 levels of distractor ratio for each of 7 levels of

the spatial swap noise level (x-axis). The highest saccadic selectivity occurs with middle levels of noise. In contrast, if there is no, or little spatial swap noise, then there is reduced saccadic selectivity. This is because in the absence of the spatial swap noise the target can be fixated immediately. Similarly at high noise levels there is reduced saccadic selectivity as so little information is perceived that the fixation pattern is at chance. What can be seen is that the saccadic selectivity for the swap model on the 9 letter task is never greater than 8%. This result is consistent with the swap models prediction for the 48 letter task. Again, the spatial swap model offers a poor prediction of the magnitude of human saccadic selectivity.

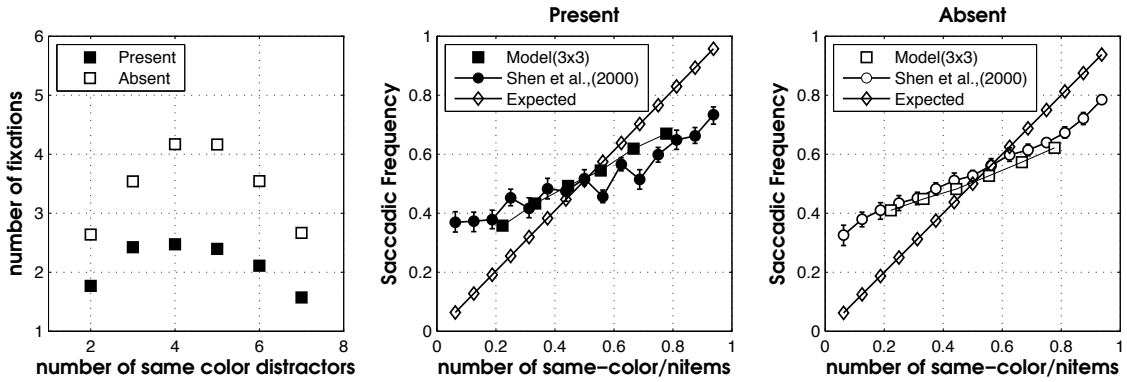



Figure 4.8: Number of fixations (left panel) and saccadic selectivity (middle and right panel) for the 9 letter distractor ratio task predicted by spatial smearing model. Human data is for the 48 letter task.

In contrast to the spatial swap model the spatial smearing model of the 9 letter task does predict a saccadic selectivity that is consistent with human performance on the 48 letter task (18%-28%). The best fitting model is shown in Figure 4.8. In the middle and right panels of Figure 4.8, the saccadic frequency to same-colour distractors for both model (square) and human performance (circle) are plotted against the ratio of the number of same-colour distractor to the number of items on the display.

## 4.7 Discussion

### 4.7.1 Spatial swap and spatial smearing

Why do the spatial swap and the spatial smearing theories lead to different optimal control policies? Over the long run the two theories have identical statistical properties; given the same level of noise, they will generate identical distributions of probability as to the location of a peripheral letter or colour. However, the percept and therefore the state that the model obtains for each sample vary differently. In the spatial swap model, the state is a consequence of stochastic changes of a feature value at one location into the value of an adjacent location (there is either a swap or there is not a swap). In the smearing model, at every instant the state is a consequence of a gradient mixture of feature information across adjacent locations. In other words, the smearing model will tend to represent the uncertainty with values, say for colour of 0.5 (half way between red and green), whereas in the swapping model the uncertainty leads to a moment-by-moment commitment to either red or green (1 or 0). One interpretation of this finding is that the swap model alternates between being confident that the state contains a red item (1) and confident that it contains a green item (0) and that therefore the observation model is not correctly updating a belief in whether the item is red or green. In the future a development of the swapping observation model needs to accept the partial observability of the setting and use a Bayesian update of a belief state given each piece of evidence.

### 4.7.2 Eye movements behaviours

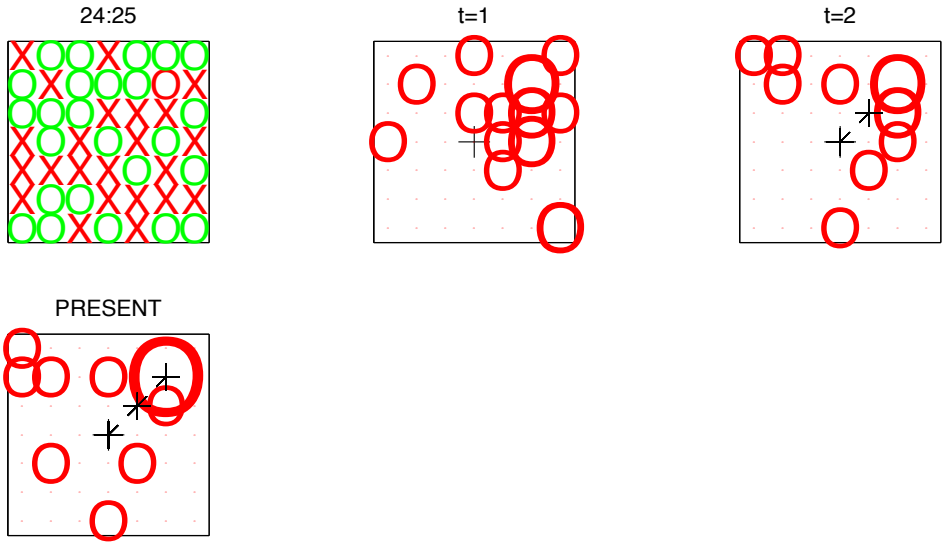
It has been reported that human eye movements behaviours in visual search revealed different types of fixations, including both MAP fixations and center-of-gravity fixations (Zelinsky, Rao, Hayhoe, & Ballard, 1997; Rajashekar, Cormack, & Bovik, 2002; Najemnik & Geisler, 2005, 2008). MAP fixations always direct to the scene location most likely to contain the target. Center-of-gravity fixations direct to a location near the centroid of a cluster of locations where have high probability of containing the target. Figure 4.9

provides some eye movements behaviours of the optimal policy of the spatial smearing model. It shows various types of eye movements that are a consequence of the optimal policy. For example, in the subfigure (a) of Figure 4.9, we see both a 'centre of gravity' like fixation and a MAP like fixation. Specifically, in the subfigure (a) of Figure 4.9, the $t = 1$ panel (the upper middle) shows that there are a cluster of target-potential locations at the upper right of the display region. In the $t = 2$ panel, it shows that the model goes to near the centroid of the cluster of locations (i.e., an example of 'centre of gravity' like fixations). After this fixation at $t = 2$, one of the locations near the current fixation becomes the mostly likely target location, and the model chooses to go to the location that most likely to contain the target (i.e., an example of MAP like fixation).

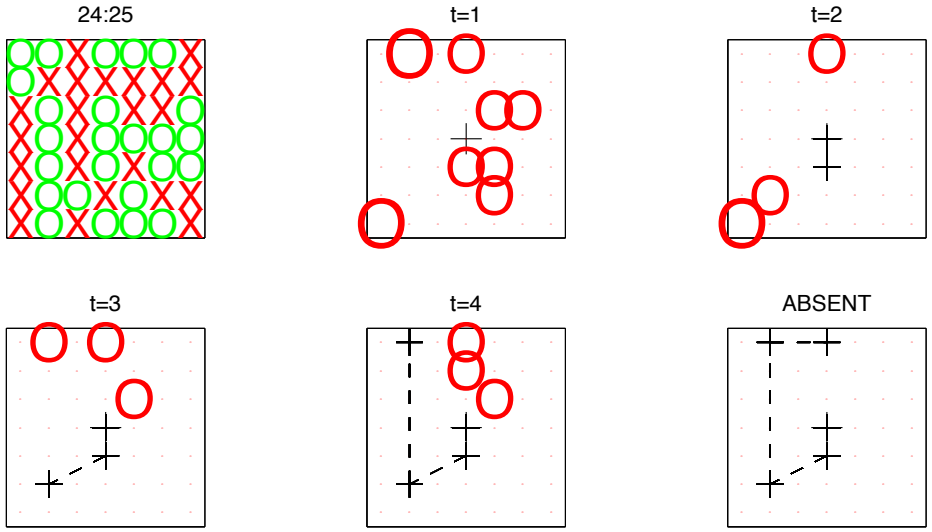
While these detailed fixation-by-fixation walkthroughs of the model provide some intuitions about what the model does, the behaviour of the model is better understood by examining the aggregate statistical behaviour in Section 4.6.

## 4.8 General discussion

The analysis presented above provides evidence that saccadic selectivity in visual search can be explained as an adaptation to spatial smearing and reward. Given a set of assumptions about low level visual information processing and given a simple reward function, behavioural predictions were derived by finding the optimal control policy. The inputs to the modeling made no assumptions about this policy, rather it was derived through a machine learning algorithm that maximised reward given the theoretical assumptions about spatial smearing in the periphery of human vision. This noise level was fitted to the distractor ratio effect and the emergent optimal policy then predicted that people should selectively saccade to minority set items. The model was tested on both the 48-item task used by Shen et al. (2000) and a 9-item task that allowed a fuller exploration of the consequences of parametric variations. In both tasks the spatial smearing theory of peripheral vision outperformed the spatial swap theory. Importantly, I did not program the model




(a) Eye movements examples of the optimal policies (target-present)




(b) Eye movements examples of the optimal policies (target-absent)

Figure 4.9: Scan path samples of the spatial smearing model

to saccade to minority set items, nor to saccade to uncertainty, nor to saccade to salient items, nor to probable target locations. Neither did I make any other heuristic assumptions about the saccadic policy or the stopping rule. Rather, both saccadic selectivity and the stopping rule policy emerged as the optimal policy given the spatial smearing and reward assumptions.

The fact that the policy is optimal allows us to establish a causal link between spatial smearing and the behaviour. The results thereby provide the answer the two questions posed in the introduction: (1) why during visual search do people choose to look where they do and (2) what determines when they act? The simple answer to (1) is because spatial smearing causes greater uncertainty about minority set items than about majority set items, and the answer to (2) is that people make present/absent decisions so as to maximise the trade off between accuracy and time cost given spatially uncertain information.

The optimal control model is an example of class of models that have been used to explain eye movements in a range of complex tasks (Hayhoe & Ballard, 2014; Trommershäuser et al., 2009). This class of model facilitates the prediction of complex behaviours from relatively simple underlying assumptions without heuristic. One important difference between the optimal control and the optimal state estimation and map-based approaches is that the optimal control approach determines a prediction of behaviour given a constrained architecture (Lewis et al., 2014; Howes et al., 2009; S. Russell & Subramanian, 1995) whereas the optimal state estimation and map-based approaches requires heuristic assumptions about control and therefore about behaviour. These heuristics mediate between the theoretical assumptions concerning state estimation and behaviour; they thereby break the causal link between the two. Two sets of heuristics are considered in the literature. One set concerns how people terminate visual search and the other concerns how people select saccades.

First, consider heuristics for how people terminate visual search. One is 'Collapsing Thresholds', which assumes that observers do search exhaustively through all items in the display (until a target is located) but that the threshold for responding target-absent

decreases as the number of items compared increases (Donkin & Shiffrin, 2011; Thornton & Gilden, 2007). There is also the 'Early Terminations' heuristic, which assumes that a probability of terminating search with a target-absent decision increases with the proportion of display items thus far compared unsuccessfully (Donkin & Shiffrin, 2011). Chun and Wolfe (1996) assumed that observers terminate a trial when the duration of the trial exceeds some duration threshold, based on the assumption that the target should have been found by that point.

Second consider heuristic approaches to saccadic programming. The most influential heuristic here is to assume that saccades are driven by saliency (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Torralba, 2003). In this approach, a winner-takes-all heuristic is used to allocate attention to locations in the scene according to combined local feature contrast maps. Approaches with these heuristic assumptions do not necessarily exhibit a poor correspondence to human behaviour but, unlike the optimal control approach, they are descriptive rather than predictive in nature. In contrast, the optimal control approach offers a systematic alternative for investigating visual search in which no a-priori assumptions are made by the theorist about human saccadic and decision strategies.

## 4.9 Conclusion

The optimal control model presented in this chapter supports the conclusion that human visual search is bounded by visual information processing constraints and reward. More specifically, in the distractor ratio task saccadic selectivity to minority set letters and colours is caused by optimal adaptation to spatial smearing in peripheral vision.

# Chapter 5

# An Optimal Control Model for Menu Search

One of the reasons that human interaction with technology is difficult to explain and model is that people are extraordinarily adaptive. The flexibility of human cognition means that it is difficult for scientists and designers to predict exactly what people will do given only a description of the interface design, the task environment and a goal. In the current chapter I test the hypothesis that users rationally adapt to (1) the ecological structure of interaction, (2) cognitive and perceptual limits, and (3) the goal to maximise the trade-off between speed and accuracy. This hypothesis is tested with a model of menu search. Unlike in previous models, no assumptions are made about the strategies available to or adopted by users, rather the menu search problem is specified as a Markov Decision Process (MDP) and behaviour emerges from a reinforcement learning algorithm. The model is tested against existing empirical findings concerning the effect of menu organisation, menu length, and whether or not the target is a known word. The model predicts the effect of these variables on task completion time and eye movements. The discussion considers the pros and cons of the modelling approach relative to other well-known modelling approaches.

## 5.1 Introduction

Over the past few years, a trend is for models of human performance to encompass the *adaptive* nature of interaction (S. J. Payne, Howes, & Reader, 2001; S. Payne & Howes, 2013; Fu & Pirolli, 2007; Vera et al., 2004; Howes, Vera, Lewis, & McCurdy, 2004b; Pirolli & Card, 1999). In these models a sequence of user actions is predicted from an analysis of what it is rational for a user to do given a device design and given known constraints on human cognition. These analyses take into account the costs and benefits of each action to the user so as to compose actions into efficient behavioral sequences. Examples include models of multitasking in which the time spent on each of two or more tasks is determined by their relative benefits and time costs (Zhang & Hornof, 2014; Janssen, Brumby, Dowell, Chater, & Howes, 2011). They also include models of the search for information on the web in which the benefits include information gain (Pirolli & Card, 1999) and even models of how to use a visual programming language (Fu & Gray, 2004). The analysis of rationality has also informed empirical studies of tasks such as driving while using a phone or ipod (D. P. Brumby, Salvucci, & Howes, 2009), studies of powerpoint use (Charman & Howes, 2003), and studies of interactive planning (O'Hara & Payne, 1998). In each case human behavior is shown to be a rational adaptation to the task environment and psychological constraints.

Consider as a detailed example a model of on-line product review reading behaviour reported by (Lelis & Howes, 2008). It is known that, prior to purchase, consumers tend to read more reviews about their favoured product and it is also known that they prefer to read negative over positive reviews. Lelis and Howes (2008)'s model explains these strategies as adaptations to the distribution of information in the environment. It turns out that on a site such as Amazon, there are many more positive reviews than there are negative reviews. This skew in the valence of available information means that negative reviews of the favoured choice are more likely to lead to a beneficial change in a decision makers preference, by reducing the expected value of the favoured choice, and are therefore more valuable.

The objective of the current chapter is to report a model of adaptive interaction that is novel in two respects. First, the model makes use of a simple but powerful machine learning framework in order to derive rational behavior given the constraints. Second, the model offers the first account of rational menu search in which predictions of search time and eye movements emerge from assumptions about the user's task environment and limitations. Menus are a widely used technique for interaction in a broad range of applications and devices. There are also many empirical studies of the use of menus and they continue to present design challenges. Menu search is important, because while users may learn the locations of frequently used menu items, many menu commands are used infrequently and users often have to search for commands that are new or that have forgotten locations. Sometimes the search will be localised to a particular menu and sometimes not. A predictive model of menu search could assist in research aimed at improving this ubiquitous and powerful interface technology, for example (Bailly, Oulasvirta, Kötzing, & Hoppe, 2013).

Evidence suggests that there is substantial scope for strategic adaptation in menu search (Byrne, 2001; Miller & Remington, 2004; D. Brumby & Howes, 2008; Hornof & Halverson, 2003; Halverson & Hornof, 2007). Despite deceptive simplicity, menus invite a flexible range of behavior. Even, for example, the seemingly mundane task of searching through a vertically arranged textual menu can be achieved by starting at the top and considering each item or, alternatively, by guessing where the desired item will be in an alphabetically ordered or semantically grouped list. Sometimes users appear to skip over some items but not others (D. Brumby & Howes, 2008). They also search differently when looking for a known-word than when looking for a semantically related item (D. P. Brumby, Cox, Chung, & Fernandes, 2014). There are many further possibilities and refinements and many subtle variations that determine, for example, where to look next, when to make a guess, and when to search in a different menu entirely. For many users, these are choices that are made quickly and implicitly.

There has been a substantial effort to model menu search and related visual search

tasks and some of this work has addressed the issue of search strategy (Halverson & Hornof, 2011; Byrne, 2001; Miller & Remington, 2004; Fu & Pirolli, 2007; Kieras & Hornof, 2014; D. P. Brumby & Howes, 2004). However, to varying degrees the existing models tend to require the modeller to program either a single or a set of strategies. For example, the EPIC model reported in (Halverson & Hornof, 2011) uses production rules to implement a visual search strategy in which fixations are constrained to fall on new items that are outside the current effective field of view. The ACT-R model reported in (Byrne, 2001) had five productions for the first saccade of a visual search and two productions for subsequent saccades. One of these two productions directs the gaze randomly and the other directs it to the next location with a feature that matches the target. In (Byrne, 2001) the productions are allowed to compete but nonetheless encode assumptions that limit the space of possible behaviors. For example, the productions restrict the menu search to consist of some mix of top-to-bottom attention with random 'look-anywhere' jumps (Byrne, 2001).

In contrast to the previous models of menu search, the model reported in the current chapter has at least two advantages (1) it does not require the modeler to hand-code production rules and, as a consequence, (2) there are no arbitrary restrictions on the range of possible behaviors. Instead control knowledge and therefore predicted effects are an emergent consequence of rational adaptation to (1) the ecological structure of interaction, (2) cognitive and perceptual limits, and (3) the goal to maximize the trade-off between speed and accuracy. In other words, the model shows how user strategy might emerge from the constraints imposed by the statistics of experienced menus and perceptual/motor and cognitive constraints. Therefore, while the proposed model builds on a number of insights concerning cognitive mechanisms that were first proposed with ACT-R and EPIC models, it does not require any hand-coded production rules to make predictions.

In what follows I first review the background to the work. In particular, I focus on contributions to understanding Human-Computer Interaction as a rational adaptation to constraints. Subsequently, I report a model of menu search which shows how visual

search behaviour is an emergent consequence of adaptation. Critically, the model makes no a priori assumptions about eye movement strategy nor about when to keep looking and when to make a choice (which are usually thought of as being stopping rule problems). The model is tested with two studies one of which was chosen to demonstrate that the model can generate predictions for real-world menus and the second of which was chosen so as to provide a comparison to an existing data set. The first study generates predictions for users of an Apple Mac who have experienced a number of applications and a number of the menus deployed by those applications. The second study tests the model's predictions against data from participants in a laboratory experiment reported by (Bailly et al., 2014). Detailed comparisons to human performance metrics (including performance time and eye movements) are provided.

The primary contributions are:

1. The further development of a framework for explaining the emergence of rational interactive behavior.
2. A novel computational model, based on machine learning, of menu search showing how behavior is an emergent consequence of environment, cognition, and utility.
3. A quantitative account of how existing empirical findings concerning menu search can be explained as rational adaptation to constraints.

## 5.2 Background

The idea that human interaction with technology can be understood as a rational adaptation has strong roots in the HCI and Human Factors literatures (Baron & Kleinman, 1969b; S. J. Payne et al., 2001; Pirolli & Card, 1999; Vera et al., 2004; Fu & Pirolli, 2007; Pirolli, 2007; Tseng & Howes, 2008; Lelis & Howes, 2011; S. Payne & Howes, 2013). One way to summarise the framework adopted in this literature is with Figure 2.1 (S. Payne & Howes, 2013), which was illustrated in Chapter 2.

The successful application of this framework requires a theory of each of the components in Figure 2.1. One key set of constraints are those imposed by the human vi-

sual system. In the vision research literature there has been much recent interest in explaining visual search as an adaptation to visual processing mechanisms and reward (Trommershäuser, Maloney, & Landy, 2009; Sprague & Ballard, 2004; Hayhoe & Ballard, 2014; Nunez-Varela & Wyatt, 2013). Key constraints imposed by the mechanisms concern saccade and fixation latencies (Rayner, 1998) and also the reduction of acuity with eccentricity from the fovea (Kieras & Hornof, 2014). It has been shown that, given these constraints, strategies can be derived through the use of a reinforcement learning algorithm (Sprague & Ballard, 2003; Hayhoe & Ballard, 2014; Chen, X, Howes, A., Lewis, R.L., Myers, C.W., Houpt, 2013), though it is possible that strategies may be acquired by other learning mechanisms, for example, by cultural transmission, through instructions, or by evolution.

The optimal control approach explored in this thesis is also influenced by ideas in optimal control and Machine Learning (Baron & Kleinman, 1969b; S. Russell & Subramanian, 1995; Sutton & Barto, 1998). A key contribution has been to provide a formal basis for learning an optimal control policy given only a definition of the reward function, the state space and the action space. Control knowledge is simply knowledge that determines what-to-do-when. In the case of menu search it concerns where to move the eyes and what/when to select an item. In this framework the expected value of an action given a state is the sum of the immediate reward plus the rewards that would accrue from subsequent actions if that action were selected. This simple assumption has provided a means of deriving human visual search strategies in well-known laboratory tasks (Chen, X, Howes, A., Lewis, R.L., Myers, C.W., Houpt, 2013). It also provides a means by which to derive rational menu search behaviour given assumptions about ecology, utility and psychological mechanisms but only if the user's menu search problem can be defined as a reinforcement learning problem. In the following section I report a model that does just that.

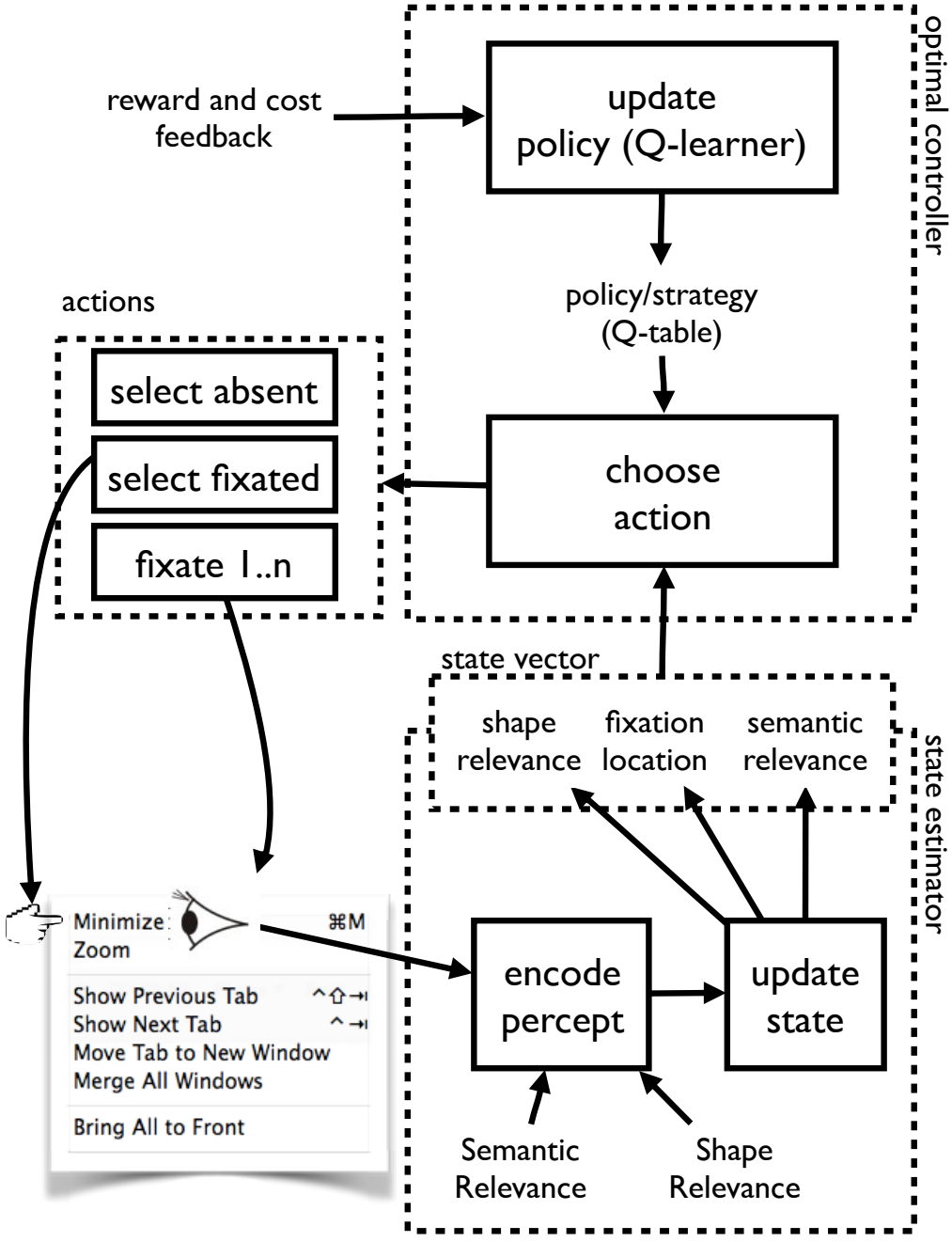



Figure 5.1: An overview of the adaptive menu search model.

## 5.3 Theory and model

Imagine that the goal for a user who is experienced with menus, but who has never used Apple's OS X Safari browser before, is to select 'Show Next Tab' from the Safari Window menu. This task and menu are illustrated to the bottom-left of Figure 5.1. A user might solve this goal by first fixating the top menu item, encoding the word 'Minimize'; rejecting it as irrelevant to the target, moving the eyes to the next group of items, that begins 'Show Previous Tab', noticing that this item is not the target but is closely related and also noticing, in peripheral vision, that the next item has a similar word shape and length to the target; then moving the eyes to 'Show Next Tab', confirming that it is the target and selecting it. The aim of the modelling is that behaviours such as this should emerge from theoretical assumptions. Importantly, the aim is not to model how people learn specific menus and the location of specific items, rather the aim is to model the menu search task in general. The requirement is that the model should learn, from experience, the best way to search for new targets in new, previously unseen, menus.

The *state estimation and optimal control* approach was used to achieve this goal we use. In Figure 5.1 an external representation of the displayed menu is fixated and the state estimator encodes a percept containing information about the relevance of word shapes ('Minimize' and 'Zoom', for example have different lengths) and semantics (word meanings). This information is used to update a state vector, which has an element for the shape relevance of every item in the menu, an element for the semantic relevance of every item in the menu, and an element for the current fixation location. The vector items are null until estimates are acquired through visual perception. Updates are made after every fixation, e.g. after fixating 'Minimize' in the above example. After having encoded new information through visual perception, the optimal controller chooses an action on the basis of the available state estimate and the *strategy* (i.e., the policy that determines a state-action value function). The chosen action might be to fixate on another item or to make a selection, or to exit the menu if the target is probably absent. State-action values are updated incrementally (learned) as reward and cost feedback is received from

the interaction. The menu search problem is thereby defined as a reinforcement learning problem (Sutton & Barto, 1998).

The paragraph above offers only a very brief overview of the theory and it leaves out many of the details. In the following subsections more detail is provided about how the state estimation and optimal controller work. Subsequently a model walkthrough is provided.

### 5.3.1 State estimation

The state estimator (the bottom right of Figure 5.1) encodes semantic, alphabetic and shape information, constrained by visual and cognitive mechanisms.

#### Semantic relevance

In common with many previous models of menu search (D. P. Brumby & Howes, 2004; Fu & Pirolli, 2007; Miller & Remington, 2004; Pirolli & Card, 1999; Pirolli, 2007), the optimal control model assumes that people have an ability to determine the semantic relevance of items by matching them to the goal specification. To implement this assumption, I used average pairwise relevance ratings gathered from human participants (which are taken from (Bailly et al., 2014)). These relevance ratings are described in detail below. For now, consider the following example: if the model sampled the goal Zoom and foveated the word Minimize then it could look-up the relevance score $0.75$ which was the mean relevance ranking given by participants. The level of this relevance score will only acquire meaning (whether it is considered a good or bad match) during learning. I also assume that people are able to maintain a short term representation of the semantic relevance of items that have been perceived. These are encoded in the state representation. No capacity limitations are assumed in this version of the theory.

Note that while the pairwise relevance ratings provide the model with a substantial database of information they do not specify the actions that should be taken on the basis of this information. How the best actions are found is described below in *The Optimal*

*Controller* section.

### Alphabetics and Shape relevance

The shape of each menu item was represented by its length in characters. No effort was made to model the shape of individual characters. The shape relevance had two levels, [0 for non-target length; 1 for target length]. The alphabetic relevance of two items was determined using the distance apart in the alphabet of their first letters. This was then standardised to a four-level scale between 0 and 1, i.e., [0, 0.3, 0.6, 1].

### Saccade duration

The saccade duration $D$ (in milliseconds) was determined with the following equation (Baloh, Sills, Kumley, & Honrubia, 1975): $D = 37 + 2.7A$, where $A$ is the amplitude (in terms of visual angle in degrees) of the saccade between two successive fixations.

### Fixation duration

It is known that the average fixation duration for reading is 200-250ms (Rayner, 1998). However, menu search involves some matching process and so some additional latency per menu item gaze is expected. In fact, in a typical menu search task the mean duration of item gazes was reported as about 400ms (D. P. Brumby et al., 2014) and this is the fixation duration assumed in our model.

### Peripheral vision

Visual acuity is known to reduce with eccentricity from the fovea. In the model, the acuity function was represented as the probability in that each visual feature of the item was recognised. The optimal control model made use of semantic features and shape features-but could easily be enhanced with other features such as colour. Semantic acuity (the semantic relevance of the item to the target) was specified as being available within $1^\circ$ of the current fixation (Inhoff & Rayner, 1980). The model predictions were compared

with the empirical data from Bailly et al. (Bailly et al., 2014). In their experiment, the height of the items in the menu was about $0.7°$. Hence, the model assumed that the semantic of item was obtained only when it was fixated. Shape acuity was specified as a quadratic psychophysical function from (Kieras & Hornof, 2014). This function was used to determine the availability, P(available) in Equation 5.1, of the shape of the item based on the eccentricity and the size of the item.

$$\begin{cases} P(available) & = P(s + X > threshold) \\ threshold & = ae^2 + be + c \\ X \sim \mathcal{N}(s, v \times s) & \end{cases} \tag{5.1}$$

where, $s$ is the item size; $e$ is eccentricity; $X$ is random noise with standard deviation $v \times s$ ($v$ is a constant). The parameters in Equation 5.1 were chosen so as to fit the materials used in Bailly et al. (Bailly et al., 2014). In their experiment, the height of an item was $0.75$ cm; the distance of the users eyes from the screen was $65$ cm. Participants should therefore have been able to simultaneously gather information about the shape of 3 items given a fovea of $2°$. Parameters for Equation 5.1 were set as follows, $v = 0.7$, $b = 0.1$, $c = 0.1$, $a = 0.3$ and $s = 0.75$. These parameter settings resulted in the following availability probabilities: 0.95 for the item fixated, 0.89 for items immediately above or below the fixated item, and 0 for items further away. On each fixation, the availability of the shape information was determined by these probabilities.

### 5.3.2 The optimal controller: Strategy/Policy Learning

The function of the optimal controller (top right in Figure 5.1) is to choose which action to do next. It does so by looking up a value for each action available in the current state and picking one. These values are called *Q-values* or state-action values and they are stored in a Q-table. The most important question to answer is how these Q-values are learnt and why they therefore implement the optimal menu search policy. Before answer this

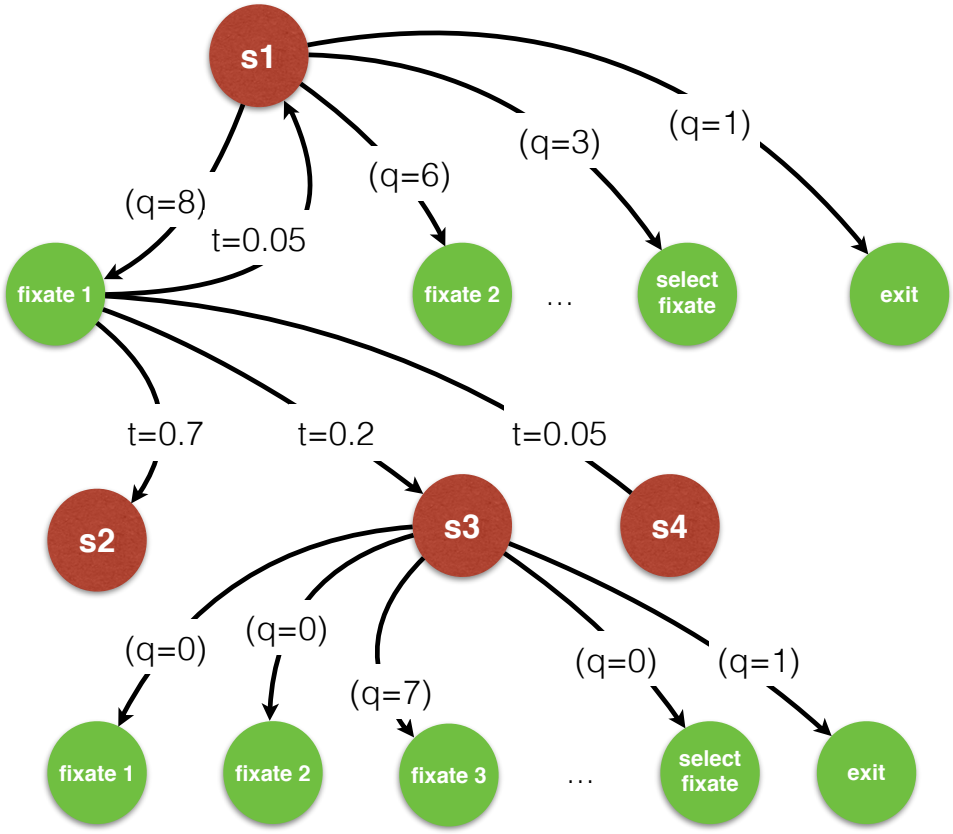



Figure 5.2: A Markov Decision Process (MDP) for searching the Safari Window menu. Red circles labelled 's' represent states. Green circles represent actions. 'q' values represent learned state-action values. 't' values represent state-action to state transition probabilities. Action 'fixate 1' is the consequence of choosing the highest value action. The model subsequently transitions to state 's3' with probability 0.2. 't' and 'q' are used in the figure for simplicity. 't' and 'q' are introduced in Chapter 3 as the transition probabilities $P^a_{s,s'}$ and state-action value function Q(s,a) respectively.

questioning, I demonstrate an example of constructing a Markov Decision Process (MDP) for a menu search task, and briefly describe how learnt Q-values are used to control search.

Imagine that the goal for a model that is experienced with menus, but which has never used Apple's OSX Safari browser before, is to select 'Show Next Tab' from the Safari Window menu. The model starts in state $s1$ (Figure 5.2), which is the *initial state*. As no visual information has been encoded the state vector [semantic relevance, fixation, word shape relevance] is null (See the top row of Table 5.1). As shown in Table 5.1, the Safari element names do not appear in the models state. The state nodes in the figures are vectors consisting of [semantic relevance, fixation, word shape relevance] in the model. The next

step is to choose an action. From each state the available actions include an action for fixating on each menu item, an action for selecting the fixated item, and an action for stopping the search because the item is believed to be absent. In the figure the state-action values (q-values) are represented on the arcs from states (red circles) to actions (green circles). (How these q-values are learnt is explained below but for now assume the values presented in the figure.) If the menu searcher is greedy then it will select the most valuable actions as represented by larger q-values. It follows that, using the MDP in the figure, the model will first fixate on the first menu item; that is it will choose the 'fixate 1' action (with value q=8) in the uppermost row of green actions in Figure 5.2 and as a consequence it would fixate the 'Minimize' item in Figure 5.1.

Having fixated 'Minimize' the model encodes the semantic and shape relevance for the fixated item. The goal is 'Show Next Tab' and so the relevance of 'Minimize' relative to this goal is likely to be low. The model also encodes shape relevance about neighbouring items in accordance with Equation 5.1. However, the exact values encoded are subject to noise because visual information processing is uncertain. As a consequence, the model might end up in either state $s2$, $s3$, or $s4$, or perhaps back in state $s1$ if no new information was encoded after the previous action. The transition probabilities to these states, $t$, are shown in the figure. These transition probabilities are a consequence of the interaction between the model of the visual system and the external environment. After any action the model is more likely to transition to a state that is (a) more likely in the environment and (b) more likely to be encoded by the perceptual model.

Subsequently, assuming that the model has transitioned to state $s3$ representing low semantic and low shape relevance, then the highest value action from this state is to fixate on item 3 (q=7). In other words, because of the low relevance of item 1, the model has skipped item 2 and focused on an item in the next semantic group. Table 5.1 gives an example of how the state vector is updated as information is gathered through eye movements and visual perception.

| *Action* | *Semantic relevance* | *fixation* | *Shape relevance* |
|---|---|---|---|
| start | $N, N, N, N$ | N | $N, N, N, N$ |
| fixate 1 | $0.1, N, N, N$ | 1 | $0.1, 0.4, N, N$ |
| fixate 3 | $0.1, N, 0.2, N$ | 3 | $0.1, 0.5, 0.9, 0.9$ |
| fixate 4 | $0.1, N, 0.2, 1.0$ | 4 | $0.1, 0.5, 0.9, 1.0$ |
| select 4 | | | |

Table 5.1: Example changes in the state representation for a 4 item menu ($N = null$).

### Learning

The Q-values, as illustrated in Figure 5.2, represent control knowledge and are learnt with a reinforcement learning algorithm. The reinforcement learning algorithm used was a standard implementation of Q-learning. The details of the algorithm are provided in Chapter 3, and further details can be found in any standard Machine learning text (e.g., Sutton and Barto (1998)).

Q-learning requires reward and cost feedback so as to learn the state-action values. A state-action value can be thought of informally as predictions of the future rewards minus costs that will accrue if the action is taken. The rewards and costs must be combined into a single utility score. A simple utility function was assumed here, in which there is a reward for success, a penalty for an error, and time is a cost. The time cost is measured in milliseconds and is determined by the visual information processing and motor assumptions described above, as well as by the number of fixations made. The reward for a correct menu item selection was $+10000$; the penalty for an incorrect selection was $-10000$. These numbers are large so as to emphasise accuracy over time taken. Indeed in the studies reported below the model achieved 99% accuracy. I have not at this stage investigated variations in this parameter, but lower values might, for example, emphasise speed over accuracy.

Before learning, an empty Q-table was assumed in which all state-action values were zero. The model therefore started with no control knowledge and action selection was entirely random. The model was then trained on 20 million samples. On each trial a menu and goal was sampled from the ecological distributions defined below. The model then

| *Assumption* | *Description* |
|---|---|
| Utility | Utility = $10000 \times correct - 10000 \times error - time$ where correct and error are Boolean variables, and *time* is in the unit of millisecond. |
| Ecology | Menus have a distribution of length and group size. Menu items have a distribution of semantic relevance and length/shape. See Figure 5.3. |
| Mechanism | People can estimate the semantic relevance of the foveated menu item. They can also estimate the shape/length of items in the periphery, although acuity decreases with eccentricity. |
| Stratetgy | A strategy (or policy) for menu search is optimised to Utility, Ecology and Mechanism assuming a state space that consists of the relevance vectors and the fixation. |

Table 5.2: Summary of assumptions for the adaptive model of menu search.

explored the action space using an $\epsilon$-greedy exploration. This means that it exploited the greedy/best action with a probability $1 - \epsilon$, and it explored other actions with probability $\epsilon$. Q-values were adjusted according to the reward and cost feedback. The optimal policy acquired through this training was then used to generate the predictions described below. To do so the optimal policy was run on a further 10,000 trials of newly sampled menus and its performance was recorded.

Although Q-learning is used here, any MDP solver that is guaranteed to converge on the optimal policy is sufficient to derive the rational adaptation. The Q-learning process is not a theoretical commitment. Its purpose is merely to find the optimal policy. It is not to model the process of learning and is therefore used to achieve methodological optimality and determine the computationally rational strategy (Gray et al., 2006; Lewis et al., 2014).

In summary, Q-learning was used to learn (or estimate) the value of each state-action pair by simulated experience of interaction with a distribution of menu tasks, where the interaction is mediated by the theory of visual perception and knowledge. The optimal policy is then the greedy policy given the Q-values. The assumptions are summarised, briefly, in Table 5.2 and explained fully below.

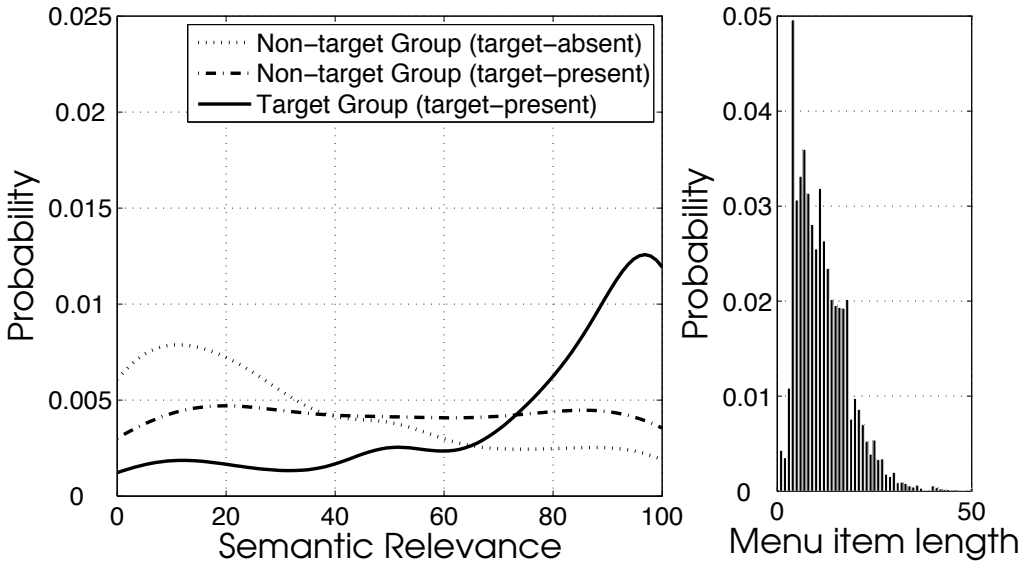



Figure 5.3: Menu ecology of a real-world menu task (Apple OSX menus). Left panel: The distribution of semantic relevance. Right panel: The distribution of menu length.

## 5.4 Study 1: Predicting real-world menu search

The purpose of this first study was to examine the model's predictions on real-world menus. In order to determine the statistical properties of a real-world task environment I used the results of a previous study (Bailly et al., 2013) in which the menus of 60 applications from Apple OSX were sampled. Together these applications used a total 1049 menus, and 7802 menu items. I used these to determine the ecological distribution of menu length, item length, semantic group size and first letter frequencies (for alphabetic search). The probability of each length (number of characters) is shown in Figure 5.3 right panel. The mean item length was 11.5, the median was 10 and the standard deviation was 6.82. The most frequent menu item length was 4 characters. As should be expected the distribution is skewed, with a long tail of low probability longer menu items.

An empirical experiment was then conducted, in which 31 participants were asked to rate how likely two menu items were to appear close together on a menu. Each participant rated 64 pairs that were sampled from the Apple OSX Safari browser menu items. The probability of each semantic relevance rating is shown in Figure 5.3 left panel. The distribution of the menu length is shown in Figure 5.3 right panel. These ratings were

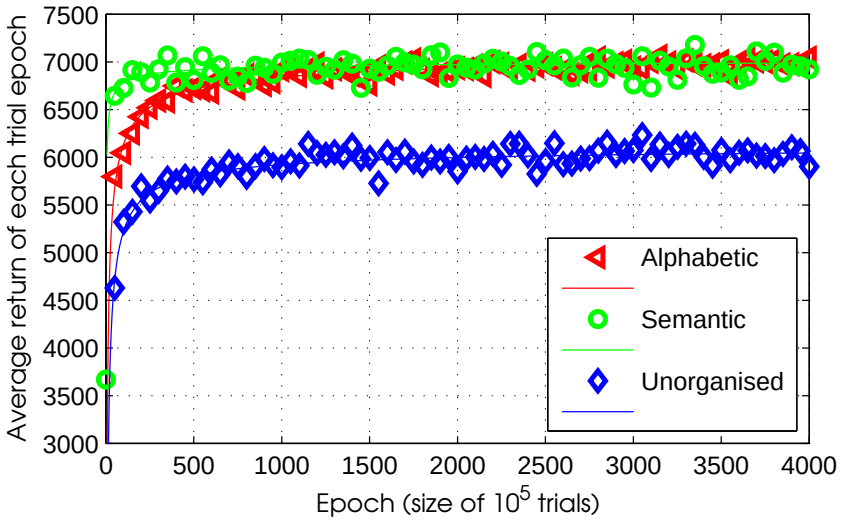



Figure 5.4: The average return of menu search against the learning trial.

used both to construct example menus and to implement the model's semantic relevance function. An example of constructing a semantically organised menu is given below. For a target-present menu, the relevance score of the target item is 1. The relevance scores of the distracted items are sampled from the distributions in Figure 5.3. There are two types of distractor items, within the target group and outside the target group. The relevance scores of the distracted items within the target group are sampled from the distribution labelled as 'Target Group (target-present)' in Figure 5.3. The relevance scores of the distracted items in other groups are sampled from the distribution labelled as 'Non-target Group (target-present)' in Figure 5.3. For a target-absent menu, all items are sample from the distribution labelled as 'Non-target Group (target-absent)' in Figure 5.3.

### 5.4.1 Results

All results are for the optimal strategy (after learning), unless stated otherwise. The optimal policy achieved 99% selection accuracy.

### Utility

Figure 5.4 shows the average return of interaction against learning trial epoch. Specifically, for each training trial, the reward is $10000 \times correct - 10000 \times error - time$, where *correct* and *error* are boolean variables and *time* is in the unit of millisecond. The average returns are over $10^5$ trials. All models reached plateau suggesting that the learned strategy was a good approximation of the optimal strategy.

### Search duration

Figure 5.5 is a plot of the duration required for the optimal policy to make a selection given four types of experience crossed with four types of test menu. Prior to training (the left most *Initial* set of three bars in the figure) the model offers the slowest performance; it is unable to take advantage of the structure in the alphabetic and semantic menus because it has no control knowledge. After training on a distribution of Unorganised menus (far right in the figure) the performance time is better than prior to training but the model is still unable to take advantage of alphabetically or semantically organised menus. After training on a distribution of semantically organised menus (middle right in the figure) the model is able to take advantage of semantic structure but this training, on its own, is detrimental in an alphabetic or unorganised menus. After training on a distribution of alphabetically organised menus (middle left in the figure) the model is able to take advantage of alphabetic ordering but misapplied this training is costly in the other menu types. The optimal policy must switch the policy depending on the menu type.

### Emergent strategies

Figure 5.6 shows typical behaviours that emerge from the models of three types of menus, i.e., alphabetic, unorganised, and semantic menus. The unadapted policy (initial policy) is random with replacement. These presented behaviours have been selected from the 10000 trials of purely exploiting the optimal policy (selecting the greedy actions on each encountered state based on the learnt policy).

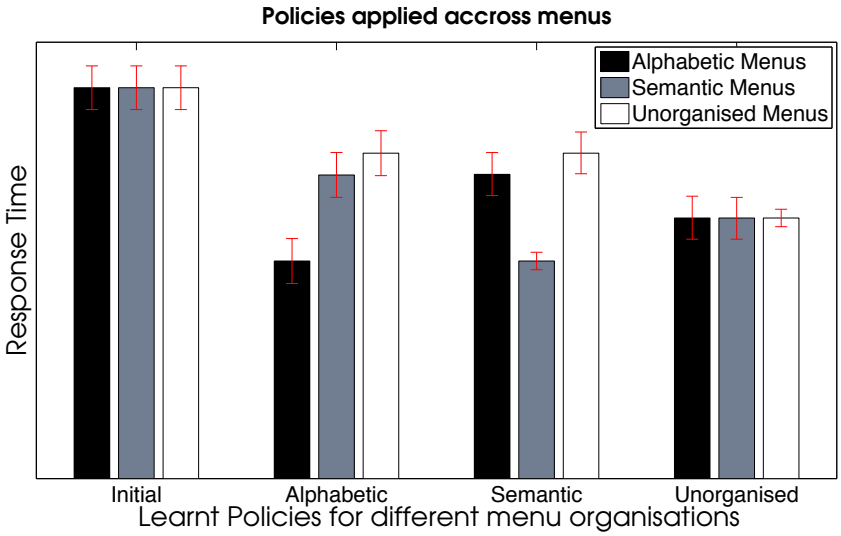



Figure 5.5: The search duration taken by the optimal strategy for each type of menu. 95% C.I.s

The policy adapted to the alphabetic menu (the top row of Figure 5.6) searches items based on the two factors below. The first factor is the starting letter of the target item. That is, the first fixation is determined by the starting letter of the target item, which can be seen from the first fixations locations in the last three panels of the first row. The second factor is the relevance between the items fixated and the target item, which drives the following fixations. The policy that is adapted to the unorganised menu (the middle row of Figure 5.6) searches item by item and skips an item if it is observed to have a low relevance in length of the items in peripheral vision. The policy adapted to the semantic menu (the bottom row of Figure 5.6) can skip whole groups after a low relevance match. Low shape relevance can also result in skipping.

### Gaze distribution

Figure 5.7 shows the effect of the menu organisation/layout on the distribution of gazes landing on target items. This is one measure of the effectiveness of each menu organisation for finding the target. A higher proportion of gazes on the target suggests a better organisation for search. The plots show the advantage of the alphabetic and semantic menus over the unorganised menus. The reason that there is higher proportion of gaze on

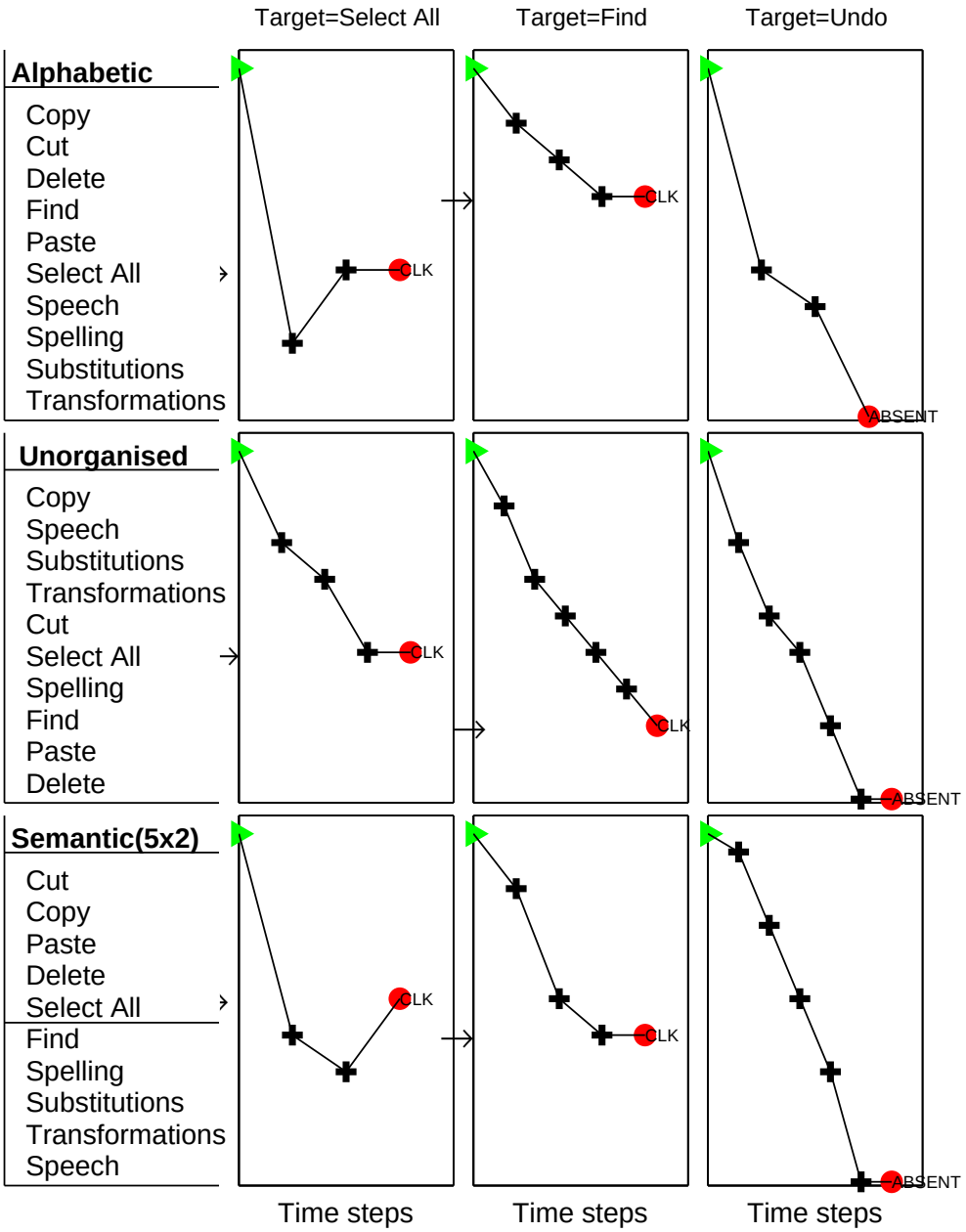



Figure 5.6: Typical behaviours that emerge from the model. Each row is for a different menu layout (alphabetic, unorganised, semantic). Examples are selected from the trials of the optimal policy.

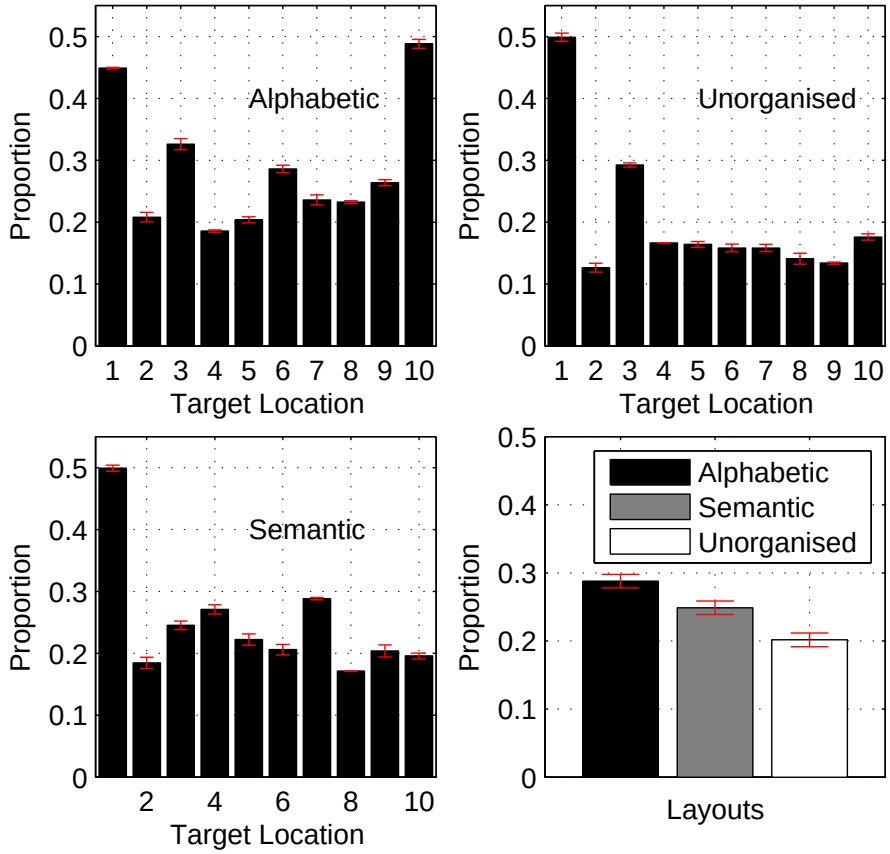



Figure 5.7: The proportion of gazes on the target location for each type of menu (95% C.I.s).

the target item in the alphabetic and semantic menus is that in the emergent policy more low relevance items could be skipped (See Figure 5.8).

### Effect of semantic grouping

Figure 5.9 shows the effect of different semantic groupings on performance time. It contrasts the performance time predictions for menus that are organised into 3 groups of 3 or into 2 groups of 5. The contrast between these kinds of design choices has been studied extensively before (See e.g., (Miller & Remington, 2004)). What has been observed is an interaction between the effect of longer menus and the effect of the number of items in each semantic group (See (Miller & Remington, 2004) Figure 8). As can be seen in Figure 5.9 while the effect of longer menus ($3 \times 3 = 9$ versus $2 \times 5 = 10$) is longer

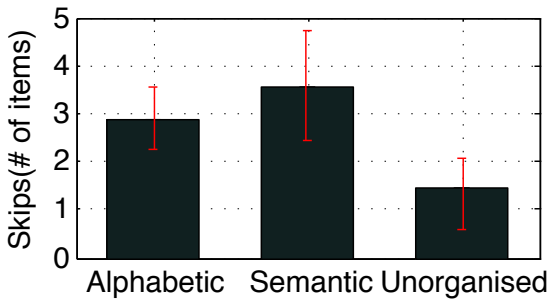



Figure 5.8: The model's prediction of skipping (mean gap between fixations).

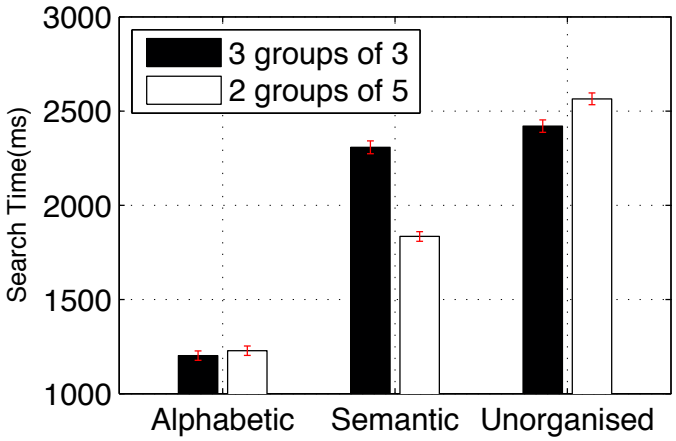



Figure 5.9: The effect of semantic group size. 95% C.I.s.

performance times in the unorganised and alphabetic menus, the effect of organisation (3 groups of 3 versus 2 of 5) gives shorter performance times in the semantic condition. This prediction corresponds closely to a number of studies (See (Miller & Remington, 2004) Figure 8).

### 5.4.2 Discussion

The results show that deriving strategies using a reinforcement learning algorithm, given plausible assumptions about the menu search problem (Table 5.2), can lead to reasonable predictions about human behaviour in ecologically valid task environments. In the following section these predictions are tested against data from a study of human participants using a small set of representative menus.

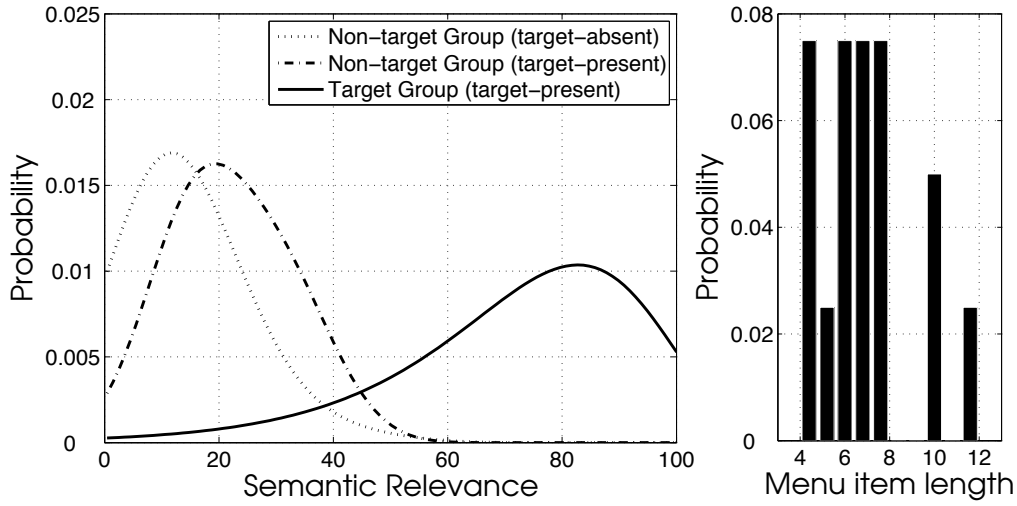



Figure 5.10: Menu ecology for the experimental menu search task. Left panel: The distribution of semantic relevance. Right panel: The length distribution.

## 5.5 Study 2: Testing the model against human data

While the previous model demonstrated the viability of the approach it did not provide an opportunity to test the predictions against human data. In this section the model is tested against a previously reported data set (Bailly et al., 2014). This data set has the advantage that it includes eye movement and task performance time data. However, it has the disadvantage that the distributions of menu length, menu item length and semantics do not correspond to the ecological distributions. Therefore, while this model shares all of the cognitive, visual, and utility assumptions with the above model, it does not share the same assumptions about the environment. Instead, for the experimental environment the distributions of semantic relevance and menu length was determined using relevance ratings reported in (Bailly et al., 2014). In addition, the experiment menus were 8-items (2 groups of 4) and 12-items (3 groups of 4). The distribution is plotted in Figure 5.10.

As with the previous model, this model was trained on distributions of menus and menu items sampled from ecological distributions. It was trained on 20 million samples. The optimal policy was then used to generate the predictions described in the following results section.

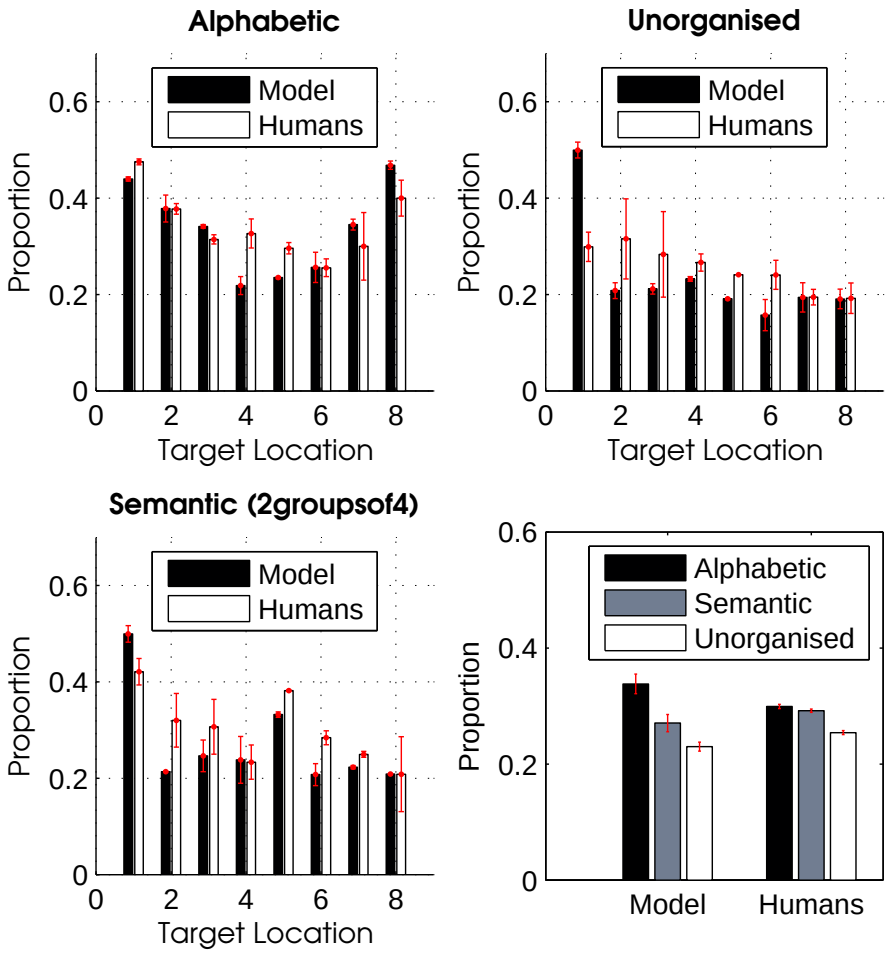



Figure 5.11: The proportion of gazes on the target location for each of the three types of menu. 95% C.I.s.

### 5.5.1 Results

A range of effects predicted by the optimal policy is reported, which is compared with the human data. As before, the optimal policy achieved 99% accuracy which corresponded with the reported participant accuracy level (Bailly et al., 2014). Effects of menu layout on performance time and also effects on gaze distribution are reported. The gaze distributions provide evidence that both model and people strategically adjust behaviour to the particular menu organisation.

### Target location effect on gaze distribution

Figure 5.11 shows the effect of target position on the distribution of item gazes for each menu organisation. X-axis represents the target location in the menu. Y-axis represents the proportion of the gazes on the target location. The model is compared to human data reported in Bailly et al. (2014). The adjusted $R^2$ for each of the three organisations (alphabetic, unorganised, semantic) are $0.84, 0.65, 0.80$. In the top left panel, the model's gaze distribution is a consequence of both alphabetic anticipation and shape relevance in peripheral vision. Interestingly, both the model and the participants selectively gazed at targets at either end of the menu more frequently than targets in the middle. This may reflect the ease with which words beginning with early and late alphabetic words can be located, e.g., 'a' and 'z'. In the top right panel, there is no organisational structure to the menu and the model's gaze distribution is a consequence of shape relevance only in peripheral vision. The model offers a poor prediction of the proportion of gazes on the target when it is in position 1, otherwise, as expected, the distribution is relatively flat in both the model and the participants. In the bottom left panel, the model's gaze distribution is a function of semantic relevance and shape relevance. Here there are spikes in the distribution at position 1 and 5. In the model, this is because the emergent policy uses the relevance of the first item of each semantic group as evidence of the content of that group. In other words, the grouping structure of the menu is evidence in the emergent gaze distributions. The aggregated data is shown in the bottom right panel; the model predicts the significant effect of organisation on gaze distribution, although it predicts a larger effect for alphabetic menus than was observed.

### Effect of length (8 v 12) on duration

It is known that people take longer to search menus with more items. Figure 5.12 shows that the model predicts the effect observed by (Bailly et al., 2014).

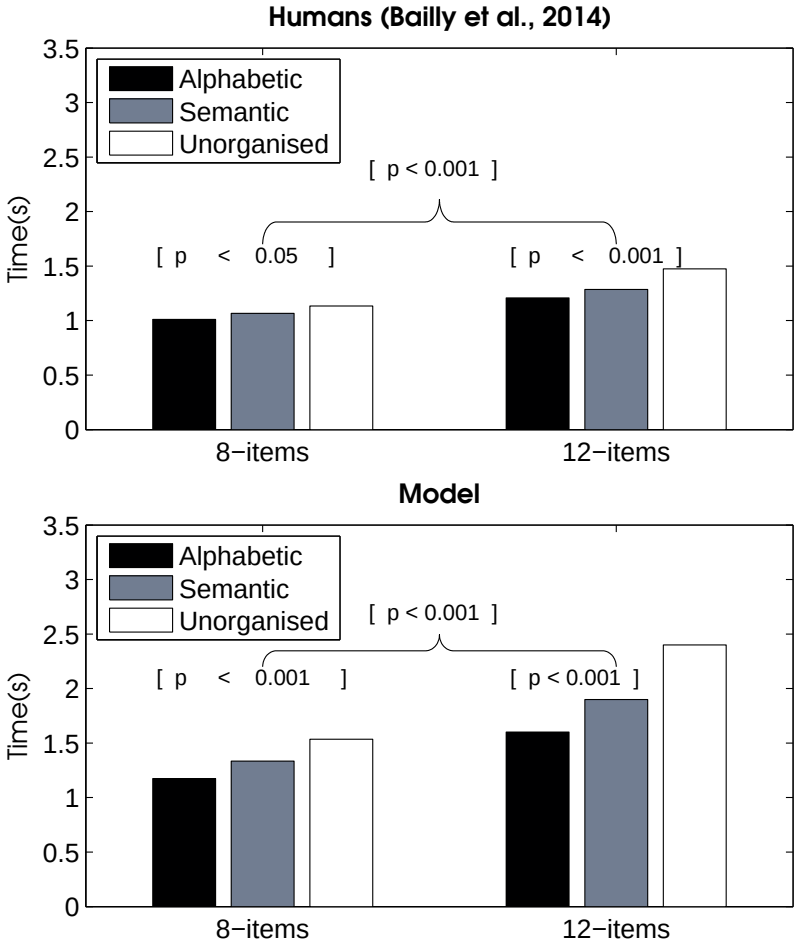



Figure 5.12: Search time for 8 and 12 item menus each organised in three different ways.

### Effect of known versus unknown word on duration

The model successfully predicted the known versus unknown word effect observed in (D. P. Brumby et al., 2014). Specifically, the effect of peripheral vision on search duration was tested. An unknown-word search model, with no peripheral vision of the word shape, was implemented. Its search was entirely driven by the semantic information. The search duration predicted by this model was ($M = 2183$, $SD = 777$), which is significantly ($p < 0.001$) slower than the search duration for a known-word search ($M = 1458$ ms, $SD = 760$ ms).

**RT distributions**

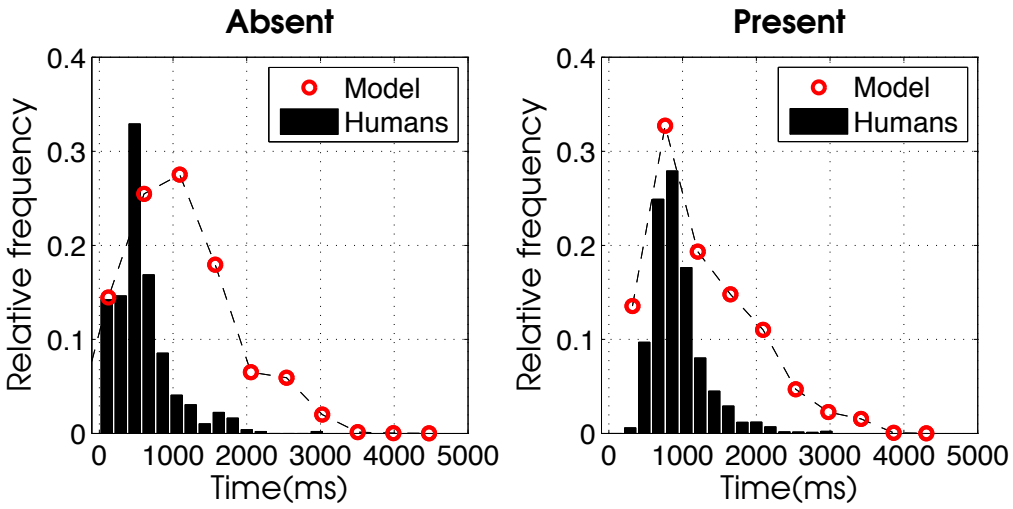



Figure 5.13: Response time distribution.

The right panels of Figure 5.13 show the response time distributions for target present responses, averaged over all target positions, for both the model and the participants. The distributions are interesting both because the mean of the model distribution corresponds closely to the mean of the participant distribution, and because of correspondence between the skewed shapes of the distribution. Long-tail distributions are a signature feature of human response times and despite the fact that the no skewed distributions are assumed in the model performance they do emerge as a consequence of adaptation to the constraints. The left panels of of Figure 5.13 suggest a less impressive correspondence between model and participant performance for target absent trials. Here the model is much slower – and perhaps less skewed – than the participants. It is possible that the reason for this discrepancy is that the participants had practice on particular menu items, which thereby made the absent decision easier; such experience was not available to the model.

## 5.6 Discussion

In the current chapter I have proposed a means of testing the hypothesis that users rationally adapt to a combination of three sources of constraint (1) the ecological structure of

interaction, (2) cognitive and perceptual limits, and (3) the goal to maximise the trade-off between speed and accuracy. This hypothesis was tested by building a model of menu search and testing it with two studies of its predictions. The first study involved applying the model to a real world distribution of menu items and in the second study the model's predictions were compared to human data from a previously reported experiment. The model was thereby tested against existing empirical findings concerning the effect of menu organisation, menu length, and whether or not the target is a known word. The predictions of the model were largely supported by the experimental evidence. The following basic effects were predicted by the model (in both real and experimental tasks) and confirmed by data: users were faster with shorter menus than longer menus; they were faster with alphabetic and semantically organised menus than with unorganised menus; and they were faster with known-words than with unknown-words. More importantly, the gaze patterns, captured with an eye tracker and reported in (Bailly et al., 2014), were predicted by the model: a higher proportion of gazes were on the target item in the alphabetic and semantic menus than in the unorganised menu. The emergent gaze strategy involved using peripheral vision for word shape to skip items, as well as using relevance estimates to skip groups of items; it also involved jumping to the end of the menu for items late in the alphabet (Figure 5.11).

Unlike in previous models, no assumptions were made about the gaze strategies available to or adopted by users, rather the menu search problem was specified as a Markov Decision Process (MDP) and behaviour emerged from solving the reinforcement learning problem. The states of the MDP were defined by a theory of foveal and peripheral vision that encoded information about semantic and word shape (length) relevance. In contrast with the production system based approaches (e.g., ACT-R and EPIC), the optimal control model has no control rules programmed by the modeller, rather the strategy is emergent from the constraints. In the extreme, the issue of tailoring the strategy space is that sometimes extra strategies are included just to produce a better fit to the data (Glöckner, Betsch, & Schindler, 2010).

While the reported models demonstrate that it is possible to predict user behaviour with both a real world and a laboratory menu system the method also has potential to predict performance with interesting design variations. It could be used, for example, to verify performance predictions with automatically designed interfaces (Bailly et al., 2013). It could also be extended to predict visual search of icons (Kieras & Hornof, 2014) or any other means of laying out options spatially on a display. It could also be used to predict rational strategies in aimed movement required for using a mouse or a touch surface (Trommershäuser et al., 2009; S. Payne & Howes, 2013).

Also, while I have reported a model of one specific task, the theory that underpins the model and much of the way in which the problem was specified as an MDP and solved using machine learning, is potentially general to many modelling problems in Human-Computer Interaction. Indeed elements of this argument have been made by a number of authors (Vera et al., 2004; Pirolli, 2007; S. Payne & Howes, 2013; Chen, X, Howes, A., Lewis, R.L., Myers, C.W., Houpt, 2013). A key property of the approach is that behavioural predictions are derived by maximising utility given a quantitative theory of the constraints on behaviour, rather than by maximising fit to the data. Although this *optimality* assumption is sometimes controversial, the claim is simply that users will do the best they can with the resources that are available to them. Further discussions of this issue can be found in (S. Payne & Howes, 2013; Howes et al., 2009).

Finally, there are many issues that need to be resolved. For example, the conversion of the Partially Observable Markov Decision Process (POMDP) into an MDP by assuming a psychologically plausible state estimation is known to fail to achieve an exact solution to the POMDP. Further work is required to understand the properties of a model in which, unlike ours, an optimal action is found for each possible belief over the states of the world (a belief state). Further work is also required to understand how this approach can be extended to capture the range of phenomena associated with skilled menu use (Cockburn, Kristensson, Alexander, & Zhai, 2007).

## 5.7 Conclusion

The chapter reported a optimal control model that given assumptions about ecology, utility and experienced ecology predicts human menu search behaviour. Unlike previous models, the optimal control model does not require the modeller to specify strategies, e.g., in the form of production rules, rather behaviour is an emergent consequence of utility maximisation given assumptions about the task environment and user psychology.

# Chapter 6

# An Optimal Control Model of Probabilistic Inference

This chapter is to test a novel theory of judgement and decision making. The theory is that decision making behaviour is an emergent consequence of constraints rather than, as demanded by the heuristic theory of decision making (Gigerenzer & Gaissmaier, 2011), a small number of predefined heuristics. The theory is tested by applying the *state estimation and optimal control* approach, or *optimal control* approach for brevity, to a judgement and decision making task called the *probabilistic inference* task. Specifically, it is a experimental task in which the participants are required to choose one of two alternatives with higher value (e.g., Gigerenzer and Goldstein (1996)). For example, consider the following problem: given information about shares in two companies, you have to choose which company's share will be the most profitable. To support this choice, you might use information gained from reading a range of different cues (e.g., positive/negative share history, or whether there has been recent investment in the company's R & D). These cues can vary in the reliability of the information provided. The *probabilistic inference* task has been used in cognitive science in efforts to discover the decision-making heuristics used by people (Gigerenzer & Goldstein, 1996; B. Newell, Weston, & Shanks, 2003; Bröder & Schiffer, 2006; Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008). I

contrast the approach proposed in this chapter to (a) the toolbox approach (Gigerenzer & Selten, 2001; Gigerenzer & Gaissmaier, 2011) and (b) the repertoire selection approach in which learning is used to choose between a small set of predefined heuristics.

## 6.1 Introduction

Following Gigerenzer and Goldstein (1996) who showed how well a simple non-compensatory heuristic, Take-The-Best (TTB), could describe human inference behaviours in probabilistic inference tasks, a number of articles offered empirical investigations into which heuristics people choose (B. Newell & Shanks, 2003; B. Newell et al., 2003; B. R. Newell, Rakow, Weston, & Shanks, 2004; Lee & Cummins, 2004; Bröder & Schiffer, 2006; Bröder, 2011; Lee & Zhang, 2012). A particular concern in this research has been whether people use a non-compensatory heuristic Take-The-Best (TTB), or a compensatory heuristic, such as Weighted-ADDitive (WADD) (Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008), Exemplar-based (Chater, Oaksford, Nakisa, & Redington, 2003), Weighted adjustment decision rule, or a loss-minimizing consistency check (Erev & Barron, 2005).

In the judgement and decision making domain, the fast-and-frugal heuristics approach (or the toolbox view) advocated by Gigerenzer and Todd (1999); Gigerenzer and Goldstein (1996) has been very influential and makes a strong commitment to the heuristic theory of decision making (Gigerenzer & Gaissmaier, 2011). However, the toolbox view is not explicit about how these heuristics are generated or selected. Mostly, the implementations of a heuristic are a description of set of rules. Human behaviours are then described by a list of predfined heuristics. For example, the heuristic Take-The-Best (TTB) is defined by set of rules as follows. (1) recognition principle: whether two alternatives are recognised? If neither of them are recognised, then guess; If one of them is recognised, then pick the recognised one; if both of them are recognised, then gather more information. (2) search rules: the cues are searched in the order of their validities. (3) stop rules:

the decision is made once a discriminating cue is found; guess if all cues do not discriminate the two alternatives. In addition, the heuristic space is not defined formally, rather a small set of heuristics is chosen by the modeller for each new data set. One common criticism is that toolbox models are difficult to falsify (Bröder, 2000; Todd & Gigerenzer, 2001; B. R. Newell, 2005; Dougherty, Franco-Watkins, & Thomas, 2008; Hilbig, 2010; Scheibehenne, Rieskamp, & Wagenmakers, 2013). This difficulty is mainly due to the lack of a criterion to determine the strategy space. In the extreme, extra strategies are sometimes included just to produce a better fit to the data (Glöckner et al., 2010).

Most of the toolbox models have been focused on the question of how individuals select among a set of predefined strategies. One of the solutions is the cost-benefit approach (Beach & Mitchell, 1978; Christensen-Szalanski, 1978; J. W. Payne et al., 1988; V. L. Smith & Walker, 1993). The cost-benefit approach claims that individuals are able to balance the trade-off between the cost and benefit of the strategies, and then choose the strategy that is the best for the problem they are facing (J. W. Payne et al., 1993). Rieskamp and Otto (2006); Rieskamp (2008) criticised this meta-strategy to the strategy selection for 'running into a recursive homunculus problem of deciding how to decide'.

Another particular proposal for how these strategy choices are made is that people learn the utilities of decision-making strategies through reinforcement learning (Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008; Fu & Gray, 2006). One of these theories, Strategy Selection Learning (SSL), casts the problem of choosing between strategies, which Rieskamp and Otto (2006), state as a reinforcement learning problem. The learning problem is to choose between a small set of predefined strategies each of which corresponds to a heuristic, including, TTB and WADD. In reinforcement learning terms, the set of available heuristics is defined as the set of actions. According to the theory a final, relatively stable selection of a strategy, is achieved via learning, that is, people learn the success and failure of the strategies through experience and select a strategy based on past success. Specifically, the hypothesis is that people build a representation of the utility of each strategy and these utilities are updated with experience. Most recently,

Scheibehenne et al. (2013) proposed a Bayesian inference based framework to test which of the heuristics are used by people. The Bayesain Inference was again developed based on a small set of predefined strategies including TTB and WADD.

However, it is known that TTB and WADD are only two examples of the range of strategies that people will adopt. For other strategies see, e.g., J. W. Payne et al. (1988, 1993). The focus on particular strategies is a natural consequence of research that is motivated by a toolbox view of bounded rational decision making (Gigerenzer & Selten, 2001) in contrast to a view in which strategies are flexibly developed in response to particular task demands. Rieskamp and Otto (2006), for example, observed that in their Study 1 participants tended to make all information available, by clicking on it, before reading it and making a decision. Similarly Lohse and Johnson (1996); Duggan and Payne (2008) have all found that participants examined very different information when the cost structure of information access changed. These behaviours do not fit easily with the toolbox model of heuristic decision making.

Unlike many approaches to modelling human information search, the optimal control approach present here requires no heuristic decision assumptions, rather the behavioural prediction is an emergent consequence of adaptation to the human information processing constraints and reward. Specifically, a state estimation and optimal control model is built for a probabilistic reference task. The state estimation is used to integrate all sources of available information. The optimal control model selects what to do next, including which piece of information to examine next and which choice to make, and when, given the state estimation. The model thereby moves towards showing how decision making strategies may emerge during an ongoing information gathering and decision making task. The results show how the model offers an explanation of the emergence of decision making heuristics, such as the non-compensatory heuristic, Take-The-Best (TTB), and the compensatory heuristic, Weighted-ADDitive (WADD) (Gigerenzer & Goldstein, 1996; B. Newell & Shanks, 2003; Bröder & Schiffer, 2006; Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008).

In particular, the model demonstrates the emergence of strategies for solving the share profitability prediction task reported in B. Newell and Shanks (2003). Crucially, the optimal control approach makes no assumptions about the set of available heuristics, rather it makes assumptions about (1) the state space defined by the cognitive architecture and the available information and (2) the available actions. Given the assumption that participants wish to maximise expected value the optimal strategies are then derived. The optimal control approach also offers novel predictions about the diversity and flexibility of decision making heuristics, thereby providing an alternative to the toolbox and repertoire selection approaches.

## 6.2 The Probabilistic Inference Task

On each trial of B. Newell and Shanks (2003)'s share profitability prediction task, two shares (Share A and Share B) from two fictional companies were present on the screen. The participants were asked to choose a share (one of two alternatives) that they thought would be the most profitable. To help the participants make this choice, they were provided with the information about the companies' financial status. Specifically, there were four aspects of information, including (a) Was the share trend positive over the last few months? (b) does the company have financial reserves? (c) does the company invest in new projects? (d) is the company an established one?

For each of the four questions, the answer could be 'yes' or 'no' for each company. For each share of one company, there were $2^4 = 16$ possible combinations of the answers. For example one vector of answers to the four questions might be [yes no no yes], represented as $[1001]$ below. On each trial, these two shares were different in at least one cue. Therefore, there were $C_{16}^2 = 120$ possible paired comparisons for two shares. In the experiment (B. Newell & Shanks, 2003), the four cues had validities of [$.80$, $.75$, $.70$, and $.69$]. The validity was reflected in how it contributed to the profit prediction. Specifically, the most profitable share on each trial was determined as follows. For each comparison,

| *Cue* | *A:B* | $P(A=1\|C_i)$ | $P(B=1\|C_i)$ | *LR(A:B)* |
|---|---|---|---|---|
| $C_1$ | 10 | .80 | .20 | $4/1$ |
| | 01 | .20 | .80 | $1/4$ |
| $C_2$ | 10 | .75 | .25 | $3/1$ |
| | 01 | .25 | .75 | $1/3$ |
| $C_3$ | 10 | .70 | .30 | $7/3$ |
| | 01 | .30 | .70 | $3/7$ |
| $C_4$ | 10 | .69 | .31 | $69/31$ |
| | 01 | .31 | .69 | $31/69$ |

Table 6.1: Probabilities that Share A or Share B is more profitable for different cue patterns in the experiment

e.g., $1011$ for share A and $0001$ for share B, the probability of share A being the best share, $P(A)$, can be computed using 'naive Bayesian rules' assuming the independence between the cues and that $P(B) = 1 - P(A)$.

The computation of these probabilities can be illustrated with Table 6.1. In the table, the first column represents the 4 cues (C1 to C4). The second column, 'A:B', lists the cue patterns for each cue. It only lists the discriminating cue patterns, '$10$' and '$01$' in the table as that non-discriminating cue patterns, '$11$' and '$00$', do not affect the probabilities computation. The third column gives the probabilities that Share A is the most profitable share given the cue patterns and validities of the cues. For example, the validity of $C_1$ is $0.80$, and given that $C_1$ is [Yes No], i.e., '$10$', then $P(A = 1|C_1 = 10) = 0.80$ and $P(B = 1|C_1 = 10) = 1 - 0.80 = 0.20$. The fourth column does the same for Share B. The far right column 'LR(A:B)' gives the ratios between the probabilities of share A and share B. Finally, to obtain the ratio of P(A):P(B), the likelihood ratios 'LR' are multiplied together for the specific cue combinations. For example, if the cue combination for share A was $1111$ (i.e., Yes Yes Yes Yes) and for Share B was $0000$ (i.e., No No No No), the ratio between P(A) and P(B) would then be $4/1 \times 3/1 \times 7/3 \times 69/31 = 62$, meaning that Share A was $62$ times more likely to be the most profitable share. If the cue combinations were $0001$ for Share A and $0000$ for Share B, then the ratio of P(A):P(B) would be $69/31$, meaning that Share A would be approximately 2 times as likely to be the most profitable than Share B.

In Experiment 1 and Experiment 2 (B. Newell & Shanks, 2003), a probabilistic environment was created by generating a random number to determine which share would be the most profitable according to this probability. In Experiment 3 (B. Newell & Shanks, 2003), a deterministic environment was created by removing the probabilistic element. Specifically, the ratio above, e.g., $69/31$, was interpreted as the share values of Share A and Share B, i.e., Share A$=$ $69$, and Share B$=$ $31$. The share with higher value was always regarded as the most profitable share on the trial. Greater details are given in the results section (Section 6.4).

## 6.3 The optimal control model

For the optimal control model, the task process was defined as a Markov Decision Process (MDP). The states were the information acquired on the task so far. The action space was not a set of heuristics (Rieskamp & Otto, 2006) instead it was the space of physical actions available to the participants for interacting with the experimental software. The reward function was defined by the validities, probabilities (as above) and the information cost. The information cost was as defined in B. Newell and Shanks (2003)'s experimental conditions. These assumptions are described in more detail below.

### 6.3.1 Define the task as a Markov Decision Process

Given the four-cue share profitability prediction experiment (B. Newell & Shanks, 2003), the model starts with an *initial state* $s_0$. The state is represented as a four-element vector. Each of the elements represents the information gained for one aspect of the company's financial status for both shares. The value of each element could be one of five possibilities, including [Yes-Yes, No-No, Yes-No, No-Yes, null-null] (where 'null' indicates that the information has not been revealed). Therefore, the initial state $s_0$ is [null-null, null-null, null-null, null-null], meaning that no information has been encoded in the state vector.

The next step is to choose an action. From each state the available actions include an action for fixating on each of the four cues, and an action of choosing Share A/Share B. For example, if the first action chosen is to look at the first cue, denoted as cue 1, and the answer for cue 1 is Yes-No, then the state transitions from *initial state* to [Yes-No, null-null, null-null, null-null]. Subsequently, if the next action chosen is to look at cue 3, and the information of cue 3 is Yes-Yes, then the new state becomes as [Yes-No, null-null, Yes-Yes, null-null]. For each trial, the model starts with the initial state and terminates when a share is selected.

The paragraph above very briefly describes the process. The key question then becomes what is the action to choose given the state in order to maximise the reward in the long run. More formally, I will subsequently introduce the *state-action value function* and how to learn the *state-action value function* by interacting with the task.

### 6.3.2 Learning

What is the action to choose given the state in order to maximise the reward in the long run? The action selection is governed by the *state-action value function*, which represents the control knowledge during the task. This is learnt with one of the reinforcement learning solutions, Q-learning[1], by interacting with the task. It does so by looking up the value of each action available in the current state and picking the one with the highest value. These values are called *Q-values* or *state-action values*, which are stored in a *Q-table*. The details of the algorithm were provided in Chapter 3, and greater details can be found in any standard Machine learning text (e.g., Sutton and Barto (1998)).

Before learning, an empty Q-table was assumed in which all state-action values were zero. The model therefore started with no control knowledge and action selection was entirely random. The model was then trained on 1 million samples (trials). Each sample

[1]While Q-learning is used here, any MDP solver that is guaranteed to converge on the optimal policy is sufficient to derive the rational adaptation. The Q-learning process is not a theoretical commitment. Its purpose is merely to find the optimal policy. It is not to model the process of learning and is therefore used to achieve methodological optimality and determine the computationally rational strategy (Gray et al., 2006; ?, ?).

was randomly sampled from all possible comparisons of the cue patterns. On each trial, the model started with an initial state and terminated when a share was chosen. During the learning, the action selection was based on an $\epsilon$-greedy policy. This means that it exploited the greedy/best action according to the Q-table with a probability $1 - \epsilon$, and it explored all the actions with probability $\epsilon$. Q-values (the state-action values) were updated according to the reward and cost feedback.

In summary, Q-learning was used to learn (or estimate) the value of each state-action pair by interacting with the simulated experience of the task, where the interaction was mediated by some theoretical assumptions below, Section 6.3.2.1 and Section 6.3.2.2.

#### 6.3.2.1 Reward

In this model, it followed the assumption that the subjective utility function exploited was consistent with the experimental pay-off regime. Hence, the reward regime used in the experimental material was used in the model, i.e., a reward for success, a penalty for an error, and a cost for each piece of information used. In addition, parametric variations in this utility function were investigated to see how the information purchase behaviours change with various utility functions.

#### 6.3.2.2 Learning cue validities

There has been some debate in the literature about whether participants can learn cue validities if they are not provided directly (Bröder, 2000; B. Newell & Shanks, 2003). The validities of the cues in B. Newell and Shanks (2003) were encoded as the relationship between the cue patterns and the most profitable share. This requires participants to be able to (1) integrate cost and benefit of the information across trials, and (2) properly assign the credit to the different cues (Brehmer, 1979). Results from the empirical testing are mixed. Bröder (2000) reported that participants were able to effectively establish the correct ranking order of the cue validities and search the cues in the order of validities. However, B. Newell and Shanks (2003) emphasised that it was difficult for the participants

to establish the correct order and use the order if the validities ranking was not provided. According to their data, when the ranking order was provided, 11 out of 12 participants searched through the information according to the ranking order. However, when the ranking order was not provided, only 1 out of 16 participants were able to establish and search through the information based on the correct ranking order.

In order to quantify the difficulty that participants had, a parameter for controlling the information validities uncertainty was explored, denoted as order-noise. The model focused on the effect of ranking reliability on information gathering behaviours. For example, B. Newell and Shanks (2003) showed that given the same information cost participants acquired more information when they were provided with the information validities ranking order. In addition, in the condition where the ranking order was provided, the participants were more likely to stop pursuing extra information when a discriminating information had been discovered. More details about the order-noise effect are given in Section 6.4.1.

### 6.3.3 Model implementation

The models were implemented in Matlab and run on an 2013 iMac with 16Gb of RAM. Q-learning converged on an approximately optimal policy with 1 million trials.

## 6.4 Testing the decision making model

The model was tested on four hypotheses and the results are presented in the following sections. Section 6.4.1 tested whether the model predicts an effect of the information cost on the amount of information acquisition and accuracy. Section 6.4.2 tested whether the model predicts an effect of the reliability of information validities ranking on information acquisition and accuracy. For example, B. Newell and Shanks (2003)'s data showed that people purchased more information when they were more certain about the information validities ranking. Section 6.4.3 tested whether the model predicts the effect of a de-

terministic environment on information acquisition and accuracy. For example, the data showed that in a deterministic environment people were more likely to stop purchasing extra information after discovering a discriminating cue . Section 6.4.4 tested the effect of the information cost on which information search strategies emerged (e.g., Guess, TTB, excess-cost, WADD).

### 6.4.1 The information cost effect

Empirical findings have shown that participants bought more information when the information cost was relatively low (Bröder, 2000; B. Newell & Shanks, 2003). For instance, when the information cost is relatively low participants are more likely to further purchase information after discovering a discriminating cue (e.g., the cue is positive for one choice and negative for the other choice). In Bröder (2000), $75\%$ and $35\%$ of participants bought extra information after a discriminating cue had been found in low and high information cost condition respectively. In B. Newell and Shanks (2003), $87\%$ and $25\%$ of participants bought extra information after a discriminating cue had been found in low and high information cost condition respectively. In this section, the information cost effect of the model is compared to the empirical data from B. Newell and Shanks (2003).

#### Experiment design

The model was built on the four-cue share profitability prediction task reported in (Bröder, 2000; B. Newell & Shanks, 2003), described in Section 6.2. In Experiment 1 (B. Newell & Shanks, 2003), (1) the information validities ranking order was not provided; (2) participants were randomly assigned to two groups. In High Relative Cost (HRC) group, when participants made a correct choice, their accounts were incremented by $5$ points minus the points spent on the information. If they make an incorrect choice, the points spent on the information were deducted from their accounts. In Low Relative Cost (LRC) group, when participants made a correct choice, their accounts were incremented by $10$ points minus the points spent on the information. If they make an incorrect choice, the points

spent on the information were deducted from their accounts.

### Parameters of the model

Two aspects of the model were adjusted to capture the experimental design. Firstly, the *reward* used in model was set based on the experimental reward regime. The utility function for High Relative Cost condition model was $[+5, 0, -1]$, i.e., $+5$ for correct choices, $0$ for incorrect choices and $-1$ for each cue used. The utility function for the Low Relative Cost condition model was $[+10, 0, -1]$.

The other is the parameter order-noise. According to B. Newell and Shanks (2003)'s analysis, when the information validities ranking was not provided, participants had difficulties establishing and searching through the information based on the ranking order of the validities. Specifically, only $1$ out of $16$ participants were able to establish and search through the information based on the correct ranking order. In order to quantify the difficulty that participants had, a parameter for controlling the information validities uncertainty was explored, denoted as order-noise. It was implemented by adding Gaussian noise to probability that was used to decide the most profitable share, mentioned in Section 6.2. Specifically, for each comparison between two shares, there was an associated probability of Share A being the most profitable $P(A)$. order-noise was then represented as a Gaussian noise to probability $p(A)$. Therefore, $P(A)' = P(A) + \mathcal{N}(0, on)$. A random number generator was then used to determine the most profitable share according to $P(A)'$. $5$ levels of order-noise, [0, 0.1, 0.2, 0.3, 0.4], were explored, as shown in Figure 6.1.

In Figure 6.1, the upper two panels are for the High Relative Cost condition where the utility function is $[+5, 0, -1]$. The bottom two panels are for the Low Relative Cost condition where the utility function is $[+10, 0, -1]$. All four panels share the same legend in the bottom right panel. In the legend 'on' is used for order-noise. Each colour represents a level of order-noise. The higher value of order-noise represents higher order uncertainty. The results show that both accuracy and information acquisition are sensitive

to the order-noise parameter, especially when the information cost is relatively high (HRC condition). For example, for HRC, when the order-noise is $0.4$, the information required is smaller than $1$. One thing needs to be noticed is that in the figure, 'order-noise=0.1' ends up being more accurate than 'order-noise=0'. This is likely to an artefact of the curve fitting, confirmed in future work.

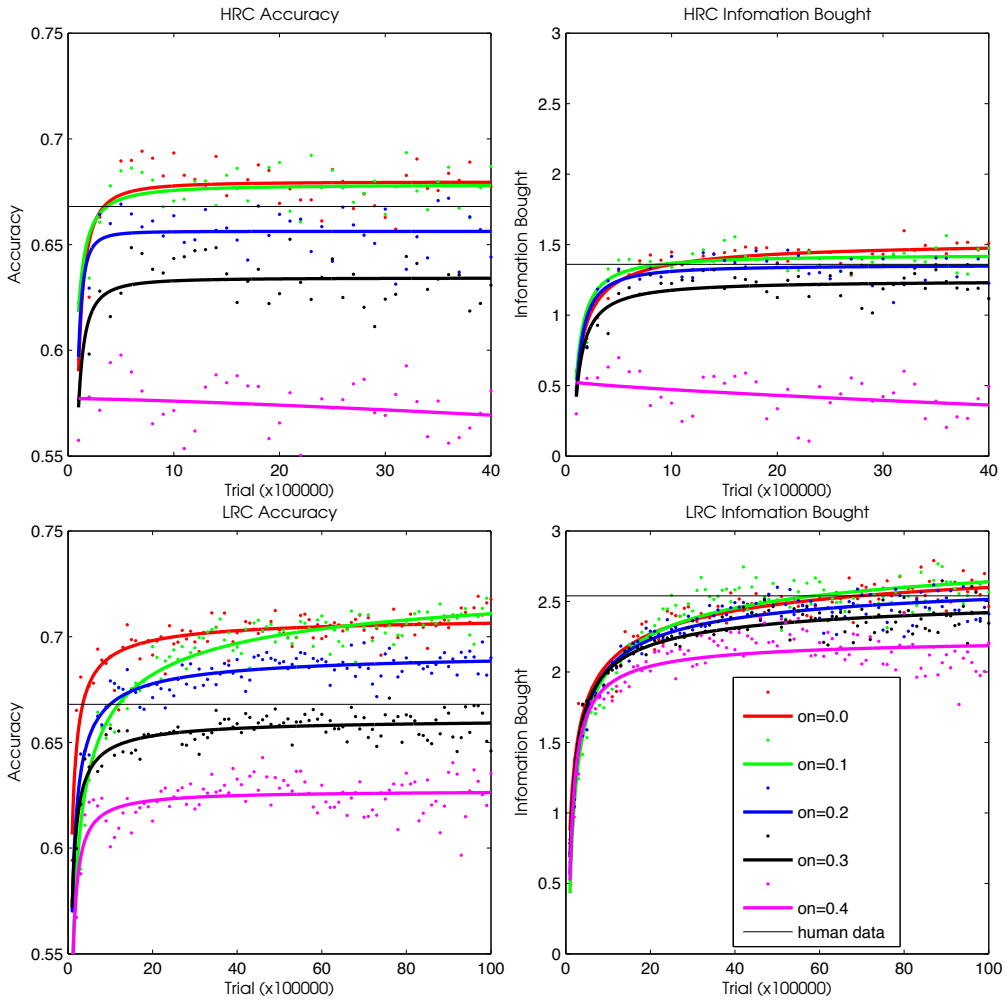



Figure 6.1: Accuracy and Information Bought for different levels of order-noise. Each colour represents a level of order-noise; the black horizontal line in each panel is the human data. For the accuracy of the human performance is the mean across both HRC and LRC conditions, as reported by Newell and Shanks, 2003

Figure 6.1 shows us the range of possible behaviours of the model given variation in the order-noise parameter. This flexibility presents a danger of over-fitting. Next I describe how the model was calibrated.

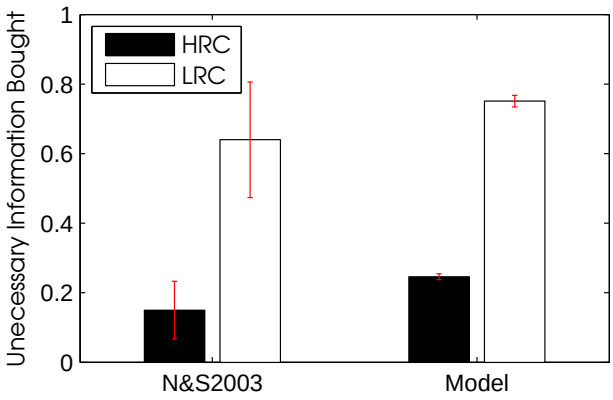



Figure 6.2: The proportion of trials where the information bought after discovering discriminating cue.

The utility function was set as the reward regimes in the experiments. The parameter order-noise was found by fitting the accuracy and information bought empirical data in HRC condition. The model with the same parameter was then used to predict (1) the accuracy and information bought in LRC condition; (2) the unnecessary-information bought in both HRC and LRC conditions. Unnecessary-information is a term borrowed from B. Newell and Shanks (2003), which is defined as the proportion of trials on which information is bought after discovering a discriminating cue.

### Results

The best fit to accuracy and information bought was found when order-noise is $0.2$ in HRC condition. Order-noise$=0.2$ was then used to predict the unnecessary-information bought in HRC condition, and all three aspects of human data in LRC condition.

Unnecessary-information is one measure of the information search stopping criterion. Figure 6.2 shows the effect of the information cost on seeking extra information after discovering discriminating cue. For both model and empirical data, there is a higher proportion of trials where extra information is required after discovering a discriminating cue in the LRC condition (where the information cost is relatively low).

Figure 6.3 shows the comparison between the empirical data and model predictions.

In each panel, the measure is plotted against learning trials. The plateau that model converged shows the performance of the optimal policy. The top right and the bottom right panels of Figure 6.3 show that the information acquisition of the model converges closely to the empirical data in both conditions (HRC:1.36 vs 1.36; LRC 2.54 vs 2.56 ). For the accuracy, the model predicted that accuracies in HRC condition and LRC condition were $64.85\%$ and $70.5\%$ respectively. A mean accuracy was reported in B. Newell and Shanks (2003), which was $66.8\%$.

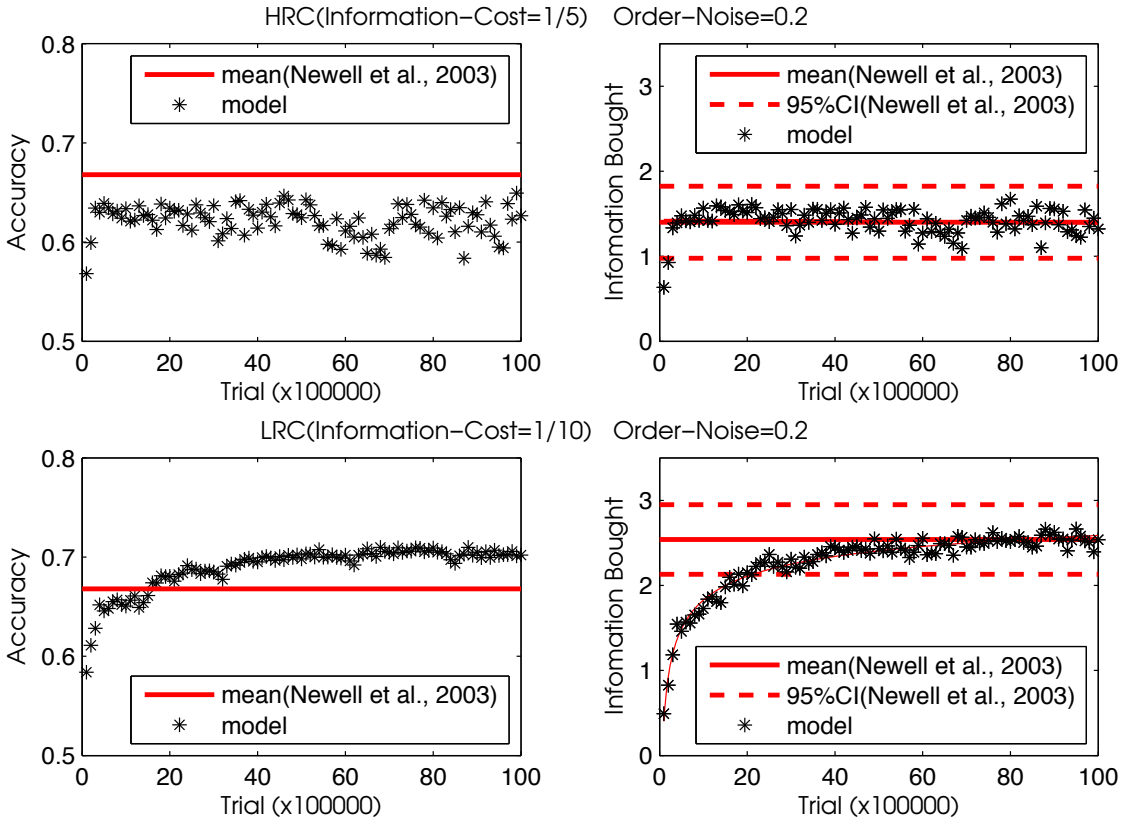



Figure 6.3: Model predictions of Accuracy (top left and bottom left) and Information Acquisition (top right and bottom right) plotted against empirical data from B. Newell and Shanks (2003).

### 6.4.2 The information ranking reliability effect

B. Newell and Shanks (2003) showed that the participants had difficulties establishing and searching through the cues in the order of validities if the validities ranking were not

provided (also see some earlier articles, Fischhoff, Slovic, & Lichtenstein, 1977; Brehmer, 1980; Einhorn & Hogarth, 1981).

This section explores whether the model could predict the effect of ranking reliability on information gathering behaviour. For example, B. Newell and Shanks (2003) showed that given the same information cost participants acquired more information when information 'usefulness' ranking order was provided, compared to a condition where the ranking order was not provided, i.e., the Experiment 1 in previous analysis (Section 6.4.1). In the following, B. Newell and Shanks (2003)'s Experiment 2 is described, followed by the model calibration to the specific experiment. Results are then reported.

### Experiment design

In Experiment 2 (B. Newell & Shanks, 2003) the ranking order of usefulness of the four cues was provided after 30 trials and after 60 trials (180 trials in total). According to the data, 11 out of 12 participants searched through the information according to the ranking order, in contrast to 1 out of 16 participants in a condition where the ranking order was not provided (Experiment 1 in Section 6.4.1). There was only a High Relative Cost condition reported in Experiment 2, so the results in Experiment 2 (this section) will be compared with results in the same condition in Experiment 1 (previous section).

### Parameters of the model

The utility function for the model was set according to the experimental pay-off regime, i.e., High Relative Cost. The ranking order was provided in the Experiment 2. According to the data, 11 out of 12 participants searched through the information according to the ranking order. Therefore, the order-noise was set to zero in the model in this experiment. The parameter order-noise could also be adjusted in this experiment. However, here I assumed that the ranking order provided in the experiment was perfectly exploited by the participants.

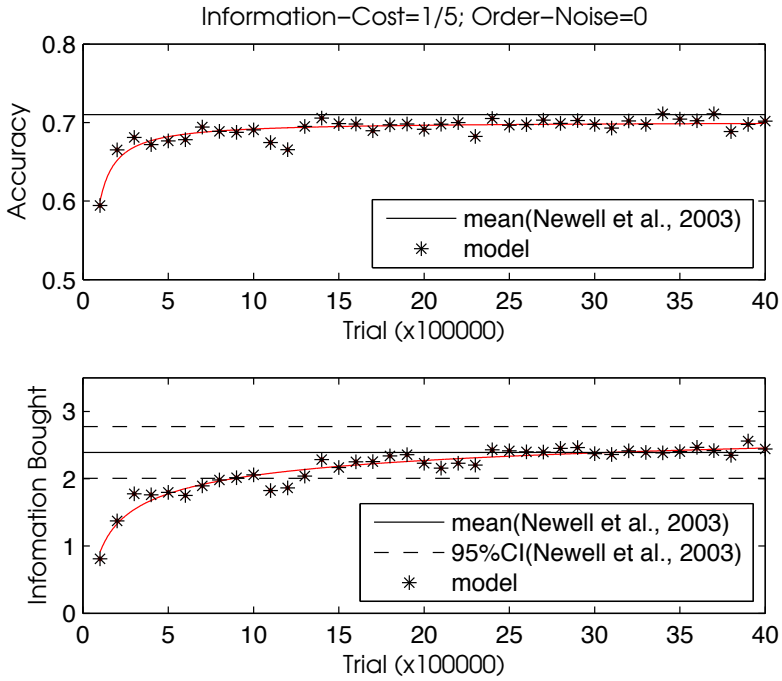



Figure 6.4: The model predictions ploted against the human data.

**Results**

As shown in Figure 6.4 below, the upper panel shows the accuracy comparison between human data ($72\%$) and the model prediction ($70.5\%$). The bottom panel shows the information bought comparison between human data ($2.39(\pm.68SD)$) and model ($2.40$). Therefore, the model performance converges closely to the empirical data of Experiment 2.

Figure 6.5 compares results of Experiment 1 (Section 6.4.1) and Experiment 2 in B. Newell and Shanks (2003). In Experiment 1, the validities ranking order was NOT provided, labelled as 'Not Provide Ranking' in the figure. In Experiment 2, the validities ranking order was provided, labelled as 'Provide Ranking' in the figure. The information cost for both was the same, $[+5, 0, -1]$.

Left panel of Figure 6.5 shows that, for both human data and model, the amount of information bought is greater in the 'Provide Ranking' condition (the white bars). Right panel of the figure shows that, as observed in human behaviour, the model in 'Provide

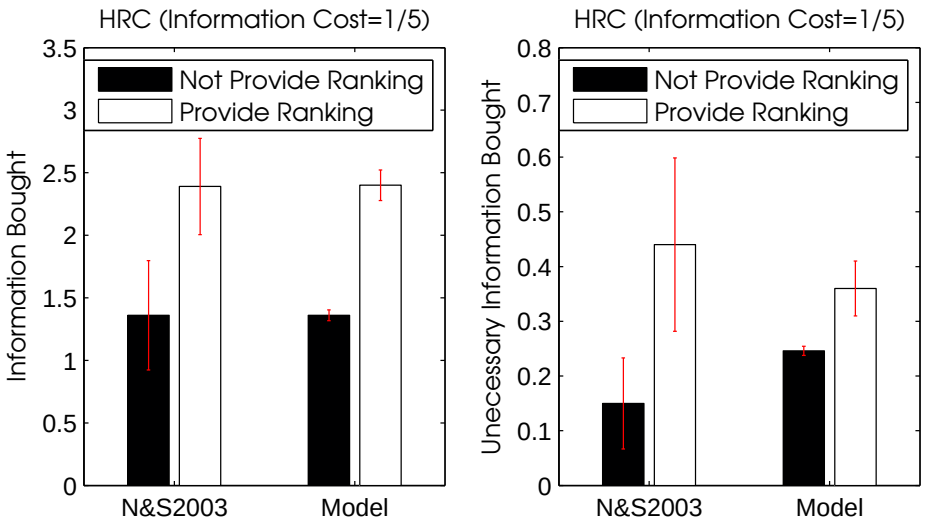



Figure 6.5: Effect of the order-noise on the information acquisition and extra information purchased after discriminating cue found

Ranking' condition is more likely to buy extra information after a discriminating cue has been found.

Therefore, the model correctly predicts the Information bought and unnecessary-information effect of ranking order reliability. Worth noting that the amount of information about (necessary and unnecessary) is an emergent feature of the models strategy.

### 6.4.3 The deterministic environment effect

In this section I contrast the predictions of the model in a probabilistic environment with the predictions in a deterministic environment. Previous two analyses were based on a probabilistic version of environment. For the share profitability prediction task, the probabilistic version was created as follows. For example, if the value of Share A is $31$ and the value of Share B is $69$, then the probability of Share A being the best share is $31/(31+69) = 0.31$, while the probability of Share B being the best share is $0.69$ (see Table 6.1). For Experiment 3 B. Newell and Shanks (2003) created a deterministic environment by removing the probabilistic element. That is, for the case above, as the value of share B $(69)$ is greater than the value of Share A $(39)$, then Share B is always regarded as the

best share on the trial. Even in the deterministic environment, participants were not able to know exactly which share was the best share as there was still uncertainty to know the exact value of each share. A model of performance in the deterministic environment was built and compared to the performance of the model in the previously described probabilistic task environment E1, E2 in B. Newell and Shanks (2003). Predictions were then compared to empirical data from B. Newell and Shanks (2003).

### Experiment design

In a deterministic environment, the share with the higher value was always regarded as the most profitable share on the trial. The ranking order of usefulness of the four cues was provided to the participants after 30 trials and after 60 trials (180 trials in total). Therefore, the results in this experiment were compared to Experiment 2 (Section 6.4.2) where the ranking order was too provided.

### Parameters of the model

The utility function for the model was the same as HRC condition in previous two models, which was $[+5, 0, -1]$. According to data, 11 out of 12 participants searched through the information according to the ranking. Therefore, the order-noise was set to be zero in the model.

### Results

Figure 6.6 plots the accuracy and information acquisition of model performance against empirical data. According to the human data, participants were more accurate in the deterministic environment ($87\% > 0.72\%$); purchase less information in the deterministic environment ($1.98 < 2.39$). The model predicted the same trend as human data, i.e., the model was more accurate in the deterministic environment ($91.8\% > 0.70.5\%$); purchase less information in the deterministic environment ($1.95 < 2.40$) (Figure 6.6 (deterministic) versus Figure 6.5 (probabilistic)).

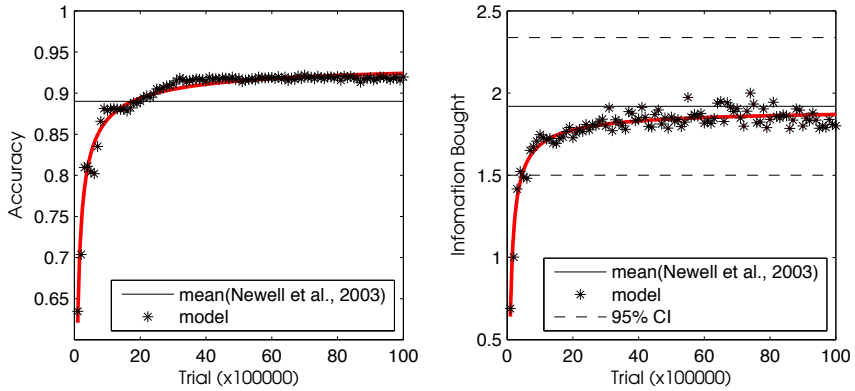



Figure 6.6: Accuracy and Information acquisition in the deterministic environment were no parameters changed from the probabilistic environment.

Furthermore, Figure 6.7 right panel shows that like human data performance, the model purchased less unnecessary-information in a deterministic environment than that in a probabilistic environment.

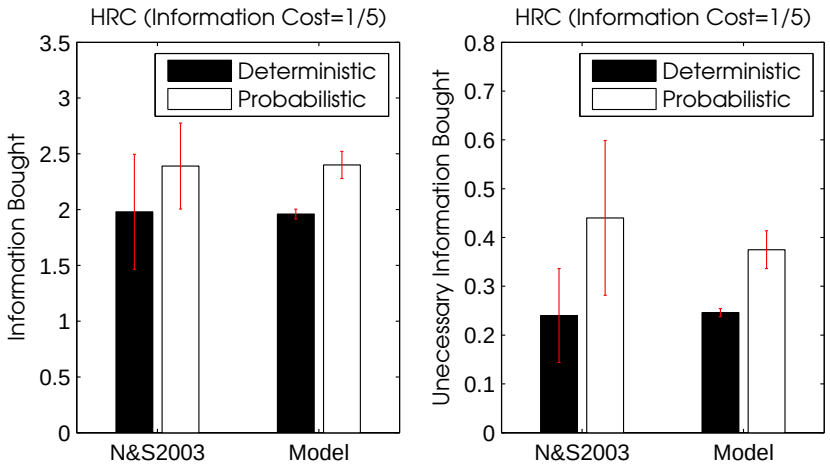



Figure 6.7: Information acquisition (left panel) and extra information discovered after discriminating information found (right panel) in both two environments: probabilistic and deterministic

### 6.4.4 Strategy sensitivity to utility function

According to the cost-benefit approach to decision makings reviewed earlier (Section 6.1), individuals trade a strategy's cost against its benefit in making their decisions. People anticipate the 'benefits and costs of the different strategies that are available and choose

the strategy that is best for the problem' (J. W. Payne et al., 1993).

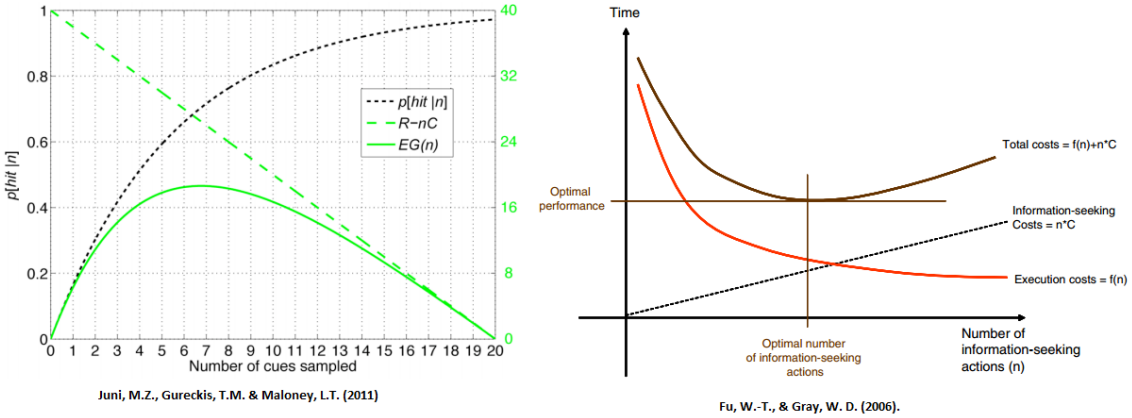



Figure 6.8: optimal information sampling

For example, in Juni, Gureckis, and Maloney (2011) a trade-off was reported between the financial benefit of accurate prediction and the financial cost of the information samples. This trade-off is illustrated in the left panel of Figure 6.8. The black dotted line represents the probability of a hit as the number of samples increases (benefit from the accuracy). The green dashed line represents the points awarded, which decreases with more samples(cost). The green solid line gives the utility of different sample sizes and thus, at its peak, gives the optimal sampling behavior.

Another specification of a cost-benefit trade-off can be found in Fu and Gray (2006). In the right panel of Figure 6.8 the red solid line represents the task execution time, which decreases as the number of information seeking actions increases (benifit from the speed). The black dotted line represents the time cost, which increases with increased information seeking. The brown line gives the utility as the number of information seeking actions increases and thus gives the optimal sampling behaviour.

In our model, the optimal strategy is a series of state-depend actions, which is adaptive to Utility, Mechanism and Environment. Given assumptions of Utility, Mechanism and Environment, the model shows how decision making strategies emerge during an ongoing information gathering and decision making task. The benefit and cost of action

shapes the emergent strategy through ongoing information processing. In this section, the utility function of the model is explored. The model offers novel behavioural predictions, suggesting that previously observed heuristics may just be points in a large space of possible human behaviours. To preview the results, the model predicted a diverse range of behaviours, that included TTB and WADD, but also included a range of other behaviours.

### Results

As shown in Figure 6.9, the strategies adopted are sensitive to the utility function, In Figure 6.9, the title of each panel gives the information cost. If the information cost is $1/4$ as in the top panel, it means that each piece of information cost $1$ point, and the reward of a correct choice is $4$ points. The y-axis represents the proportion of trials on which each strategy was used. The x-axis of the figure gives different strategies, denoted as 'Guess0', 'TTB', '$C+1$', '$C+2$', and '$C+3$'. 'Guess0' means make a random choice. For the Take-the-Best (TTB) strategy the decision is made immediately after a discriminating cue is found, i.e., when two shares differ on one piece of information. The compensatory strategy $C+1$ means that one more piece of information is obtained after a discriminating cue has been discovered. For the compensatory strategy $C+2$ two more pieces of information are gathered after discriminating formation being found. Weight ADDitve (WADD) is defined as a strategy that calculates a score for each alternative; the score is the weighted sum according to the validity of each cue (J. W. Payne et al., 1988, 1993). The heuristic Weight ADDitve (WADD) is not shown in the figure.

On the top panel of Figure 6.9, titled as 'Information Cost$=1/4$', when the information cost is relatively high, the model predicted that it is not worth buying any information at all. When the information is $1/5$, about $70\%$ of trials use TTB, i.e., make decisions as soon as a discriminating cue has been found. About $25\%$ of trials use $C+1$. When the information gets really cheap, e.g., $1/10$, the information purchase behaviours become more 'generous'. Now there is a spread of use of all of the behavioural strategies.

The model does not have five strategies built into it. It does not even give rise to

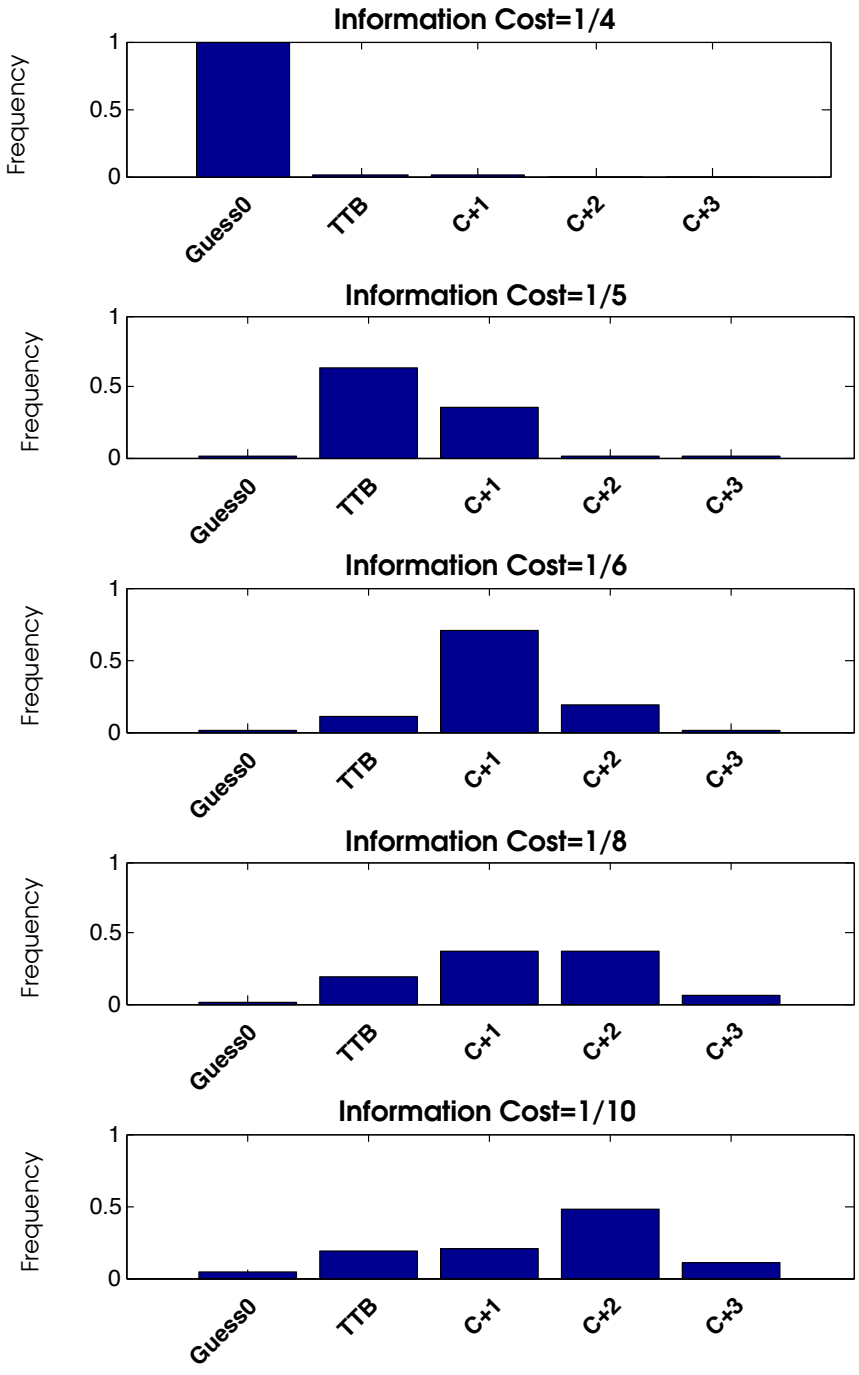



Figure 6.9: The frequency of cue use for five different strategies.

five emergent strategies. The five strategies in the Figure 6.9 are convenient behavioural markers that relate to the existing assumptions about heuristics (TTB). The five strategies can be thought of as behavioural products of a single underlying policy. This policy gathers less information (guesses) when it is more efficient to do so and it gathers more information ($C + 1$ etc.) when it is more efficient to do that.

## 6.5 Discussion

This chapter shows that human information gathering and decision making behaviour in a probabilistic inference task, which has previously been explained using a list of heuristics, can be explained as the consequences of an emergent optimal control policy given limitations on known validity of cue rankings and the deterministic or probabilistic nature of the task.

Much previous work has focused on how to make selections among a small sets of strategies. The model has included reinforcement learning among the set of defined strategies (Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008; Fu & Gray, 2006), Bayesian inference among the a set of defined strategies (Scheibehenne et al., 2013), and description based empirical tests using the process recording, or eye tracking technique (B. Newell & Shanks, 2003; B. Newell et al., 2003; B. R. Newell et al., 2004; Lee & Cummins, 2004; Bröder & Schiffer, 2006; Bröder, 2011; Lee & Zhang, 2012).

Instead of predefining a small set of strategies, such as WADD, TTB, the optimal control model showed how decision making strategies can be flexibly developed in response to particular task demands. The model predicted a diverse range of behaviours, that included TTB and WADD, but also included a range of other behaviours. In addition, the model predicted the empirical observations reported in (B. Newell & Shanks, 2003). It did so without the modeller describing the strategies and therefore offers a deeper explanation of human behaviour on this task than previous models. The optimal control approach enables a fuller exploration of the policy space that allowed by the information

processing constraints.

As mentioned (Section 2.1), for the optimal control approach, the optimal policy is sensitive/adaptive to three elements: Utility, Ecology and information processing mechanisms. In the Distractor Ratio model (Chapter 4), two information processing mechanisms were tested. The menu search model (Chapter 4) emphasised the importance of ecological environment. For example, in a alphabetic organised menu, the model's search behaviour was also shaped by the statistics of the beginning letter of menu items. In the model present in this chapter, the effect of Utility function is demonstrated. Utility function concerns the benefit and cost of the strategies. The optimal policy derived is a policy that maximises the balance between the benefit and cost. For the specific task present here, the benefit is in terms of accuracy and points. As for the cost, due to the assumption that the utility function used in the model is consistent with the experimental pay-off regime, the cost of strategy is represented as the experimental points. However, in future work, some further questions about Utility function need to be investigated.

Sometime, the experimental pay-off designs may not well reflect people's desire of goal. For example, time is precious resource for human beings. However, some studies that control the financial cost of information sampling fail to control the temporal cost (Juni et al., 2011; B. Newell & Shanks, 2003). Individuals vary either in how much time they require to make decisions or in how much time they allocate to the task or both. Greater experimental control over the time cost experienced by each individual participant might result in a different understanding of optimal information search behaviour. Another way in which individuals vary is in the weights that they give to the trade-off between time and accuracy. One response to this problem is to cast the cost and the benefit of information within the same currency (Fu & Gray, 2006). e.g., Fu and Gray (2006) expressed both the cost of information seeking and the benefit of information in terms of time. Another possibility is to cast all costs as financial and to fix the time that the participant must allocate to the experiment (see also Howes et al., 2009; S. J. Payne, Duggan, & Neth, 2007). In addition, instead of assuming that the subjective utility exploited is

consistent with the objective utility, it may requires a theory of 'internal subjective utility' (related to notions of intrinsic motivation and reward (Singh et al., 2010)).

Also, while I have modelled the aggregate data, further work is required to model individual differences. Systematic individual differences in human behaviour have been found in numerous studies, e.g., individual difference in strategies selection can be found in Lee and Cummins (2004); Bergert and Nosofsky (2007); Mata, Schooler, and Rieskamp (2007). Therefore, the aggregate data can sometimes misleading. One way in which individual experience varies, for example, is in the time cost of making decisions (Chen & Howes, 2012). Another way in which individuals vary is in the weights that they give to the trade-off between time and accuracy. Lastly, individuals vary in their preference for exploration over exploitation. Some people will choose to thoroughly explore the space of possible sampling strategies before settling on an individual strategy and others will relatively quickly decide on a strategy that they subsequently stick with.

## 6.6 Conclusion

In contrast to a small set of predefined heuristics, the optimal control model present in this chapter showed that the strategies space for the probabilistic inference tasks is shaped by the ecological environment, utility and information processing mechanism. And strategy selection problem is then represented as an adaptation to these three elements.

# Chapter 7

# General Discussion

In summary, I have explored the *state estimation and optimal control* approach, or *optimal control* for brevity, as a means of explaining human behaviours. The approach was illustrated with three models. Each of the three models yielded predictions of a different level of human behaviour, from visual perception (Chapter 4: the Distractor Ratio task) through immediate behaviour (Chapter 5: a Menu Search task), to probabilistic inference (Chapter 6: an Information gathering and decision making task). The primary contribution of the work is that in all three models behavioural strategies, rather than being programmed into the model, emerge as a consequence of utility maximisation given a theory of the information processing constraints. In all three models I explored the optimal control approach by implementing computational models that simulated psychological information processing mechanisms and then derived optimal policies using Q-learning. In each case the optimal policy predicted sequences of actions, including eye movements and decisions, from which behavioural predictions were generated and compared to data.

The aim of the first example was to explain eye movements and decisions in a visual search task. Here the optimal control model was influenced by the active vision approach for visual search (Trommershäuser et al., 2009; Sprague et al., 2007; Hayhoe & Ballard, 2014; Nunez-Varela & Wyatt, 2013). I started with the assumption that the purpose of vision is not to form the best estimate of the world but rather determine the best choice of

action. Given a set of assumptions about low level visual information processing, e.g., uncertainty in peripheral vision, and given a simple reward function, eye movements and decision making behaviours were derived by utility maximisation. Unlike many approaches to modelling visual search, the optimal control approach required no heuristic decision assumptions, rather behaviour was an emergent consequence of adaptation to the visual system and reward. In addition, two models of uncertainty in peripheral vision, 'spatial smearing' and 'spatial swapping', were tested. The results showed that in the distractor ratio task saccadic selectivity to minority set letters and colour (i.e., observed human behaviour) could be explained by optimal adaptation to 'spatial smearing' in peripheral vision, but not by optimal adaptation to 'spatial swapping'.

The aim of the second example was to explain human menu search behaviours. User behaviours were explained by a model that rationally adapted to a combination of three sources of constraint (1) the ecological structure of interaction, (2) cognitive and perceptual limits, and (3) the goal to maximise the trade-off between speed and accuracy. The predictions of the model were largely supported by the experimental evidence. Some basic effects were predicted by the model and confirmed by data. For example, both model and users were faster with shorter menus than longer menus; they were faster with alphabetic and semantically organised menus than with unorganised menus; and they were faster with known-words than with unknown-words. In contrast with previous models of menu search, there were no rules programmed by the modeller of the optimal control approach, rather the strategy was emergent from the constraints. While previous models have used a probabilistic measure of pre-programmed strategies such as serial, parallel and random search (e.g., Hornof, 1999; D. P. Brumby & Howes, 2004; Miller & Remington, 2004), in our approach the strategies were emergent from the visual information processing and task constraints. Another highlight of this analysis was the importance of the ecological environment. The model adapted to a description of the distribution of menu items in the environment. For example, in an alphabetic menu, the model's search behaviour was also shaped by the statistics of the beginning letter of menu items.

The aim of the third example was to explain probabilistic inference. The probabilistic inference task has been used in cognitive science in efforts to discover the decision-making heuristics used by people (Gigerenzer & Goldstein, 1996; B. Newell & Shanks, 2003; Bröder & Schiffer, 2006; Rieskamp & Otto, 2006; Rieskamp, 2008; Rieskamp & Hoffrage, 2008). For example, a particular concern in this research has been whether people use a non-compensatory heuristic 'Take-The-Best' (TTB), or a compensatory heuristic, such as 'Weighted ADDitive' (WADD). In this thesis, I contrasted the *state estimate and optimal control* approach to (a) the toolbox approach (Gigerenzer & Selten, 2001) and (b) the repertoire selection approach in which learning is used to choose between a small set of predefined heuristics. Instead of predefining a small set of strategies, such as WADD, TTB, the optimal control model showed how decision making strategies can be flexibly developed in response to particular task demands. The model predicted a diverse range of behaviours, that included TTB and WADD, but also included a range of other behaviours. In addition, the model predicted the empirical observations reported in (B. Newell & Shanks, 2003). It did so, without the modeller describing the strategies and therefore offers a deeper explanation of human behaviour on this task than previous models.

I now consider how these three models meet the three challenges set out in Chapter 1? To recall, the three challenges were: (1) the challenge of finding emergent control strategies, rather than programming them in the model (2) the challenge of breadth of application and, (3) the challenge of testing different information processing theories by deriving optimal strategies (Chapter 1, page 3).

The first challenge was met by all three examples. Rather than being embedded in the model by the modeller, strategies were an emergent consequence of constraints and reward. In the model of the distractor ratio task, the minority set strategy emerged as a consequence of spatial smearing in the periphery. In the model of menu search, various item skipping strategies emerged as a consequence of knowledge, environment and peripheral constraints. For example, the search was faster with alphabetic and semantically

organised menus than with unorganised menus because the emergent strategies took appropriate advantage of the structure; the search was faster with known-words than with unknown-words partly because words with the wrong length could be skipped. In the probabilistic decision task model, WADD, TTB, guessing and other variants emerged as a consequence of uncertainty of information validities and information cost.

The second challenge was met by the fact that the approach was applied to three different levels of task. The Distractor Ratio task (Chapter 4) is a basic visual perception task where the vision plays the main role for completing the task. For example, the way that colour and shape was perceived by foveated vision and peripheral vision guides the search behaviours. For the menu search task (Chapter 5), the visual information still played an important role to guide the behaviour. For example, the letter recognition and the shape information from peripheral vision. However, higher level knowledge was also required for this task. In particular, semantic relevance and/or alphabetic knowledge was crucial for this task. For the probabilistic inference task (Chapter 6), the validities of the information, and the information cost and benefit were numbers. Therefore, the task required different sources of cognitive abilities, e.g., arithmetic skill and/or frequency-encoding. In summary, diversity was achieved through the variation in sources of information across the models. Consistency is achieved through the shared approach to deriving the optimal policy.

The third challenge was met by the model of visual search (Chapter 4). Two models of uncertainty in peripheral vision, 'spatial smearing' and 'spatial swapping', were tested. Only one of these models predicted the human data. The results showed that in the distractor ratio task saccadic selectivity to minority set letters and colour could be explained by optimal adaptation to spatial smearing in peripheral vision, but not by optimal adaptation to spatial swapping. This example showed that theories of information processing mechanisms can be tested by deriving optimal strategies. In contrast with only parametric adaptation for a fixed cognitive architecture assumption, the optimal control approach could be used to test and identify alternative information processing mechanisms.

## 7.1 Lesson Learned

The work reported in this thesis makes a contribution to cognitive science by providing to the extent that the results provide insights and information on how to apply the state estimation and optimal control (SEOC) approach to modelling human behaviour. The lessons learned include lessons about how to model the three key components of the approach (utility, ecology and mechanism) so that strategies and, therefore, behavioural predictions emerge. In the following sections, I explore the contribution to each of these three components.

### 7.1.1 Utilities

The utility function provides a measure of the goodness of the strategies depending on the goal. The optimal strategy is a strategy that maximise the expected utility. Therefore, the choice of the utility function for the model is critical for understanding the optimal strategy derived. To understand human control and decision making behaviours, ideally we need the subjective utility function that is exploited by the human beings. However, the subjective utility used by humans is not clearly known. One of the assumptions that could be made is that the subjective utility exploited by the participants is consistent with the objective utility (e.g., the laboratory pay-off regime). For the information-purchase task experiment in Chapter 6, the experimenters used a quantitative function. For example, in one of the conditions, each piece of information cost 1 pence. The reward was 7 for the correct choice and 0 pence for the incorrect choice. The participants were asked to maximise the financial reward during the experiment. The model adopted the same utility function as the pay-off regime used in the experiment. For the other two models, it is more difficult to determine the 'objective' utility function because in these cases experimenters used the standard instruction of asking people to be fast and accurate. There was no quantitative utility function given to the participants in these experiments. However, in these cases it was possible to fit the model's performance to speed (number of fixations)

and accuracy (errors) by adjusting the utility function. Importantly, this did not involve fitting to the outcome data (the DR effect and the effects of organisation in the menu search task). Specifically, for the Distractor Ratio task (Chapter 4), the participants were asked to respond with a target-present or target-absent decision as quickly as accurately as possible. There was no quantitative utility function used in the experiment. In the model, the reward function was $Reward = 10 \times correct - 10 \times error - n$, where $n$ is the number of fixations. In general this means that an extra step will not be 'necessary' if it does not increase the accuracy by 10%. The model achieved 96% accuracy, compared with 98% of the participants. Also it can be seen in Figure 4.4 that the number of fixations predicted by the model corresponds well with the human data. The choice of the reward function was guided by the fact that the number of fixation that people used was from 2 to 8. In the menu search task (Chapter 5), the participants were again asked to perform the task as quickly as accurately as possible. The results showed that the accuracy achieved by the participants was 99.5%. In the model, the utility function was Reward $= 10000 \times correct - 10000 \times error -$ time(ms). This reward function massively favoured correct responses. The model achieved 99% accuracy.

Obviously, the assumption that the subjective utility is consistent with the objective utility could be problematic for various reasons. For example, some studies that control the financial cost of information sampling fail to control the temporal/effort cost, which is inevitably considered by the human beings. Also individuals might vary in the weights that they give to the trade-off between time and accuracy. This leads questions that are related to the pay-off regime design for the experiments. While not empirically investigated with controlled studies, some insights obtained during this work is discussed below.

The pay-off regime of the experiments should not be difficult for the participant to interpret, e.g. ask participants to maximise information gain or provide points feedback on the basis of complex combinations and weightings of parameters. Simple and straightforward pay-off schemes are more likely to be followed by the participants during the limited experience of the experiment. One of the most popular instructions that has been

given to participants is 'performing the task as quickly as accurately as possible'. Despite our success with quantifying this instruction in Chapters 4 and 5, we argue that a better solution to a speed/accuracy tradeoff instruction is, for example, casting the accuracy and speed (the benefit and cost) of the strategy into fewer or even one dimension. For example, in Fu & Gray (2006), they cast the cost information sampling and the benefit of more information into the same utility measure, i.e.. time. Another example is that instead of instructing the participants to perform the task as quickly as accurately as possible, a better solution is asking them to finish a fixed number of correct trials for the experiment. In the latter version, the error resulted in more time spent in the laboratory. Therefore, the major point is that a carefully designed reward regime for the experiment is critical as it promotes the consistency between the subjective utility and objective utility.

### 7.1.2 Ecologies

Briefly, ecology concerns the constraints imposed by a person's experience of the task environment, including the immediate local task environment and environment experienced through a lifetime. For the DR task (Chapter 4) and the information sampling task (Chapter 6), I assumed that the ecology was the local task environment. For the menu search task (Chapter 5), besides the local task environment, I assumed that participants in the experiment possessed the (statistical) knowledge the alphabetic tablet from their life experience.

The importance/necessity of ecology for explaining human behaviours has been mentioned several times in this thesis. For example, Hahn and Warren (2009) argue that a consideration of the nature of peoples actual experience of a fair coin toss can help explain seemingly biased perceptions of randomness. However, similar to the utility, sometimes it could be tricky for the modellers to decide the ecological environment for the human beings. Examples are discussed below to show the difficulty of controlling this aspect of the modelling and the lessons learned.

One example is the menu ecology experienced by the individuals. It is easier to mea-

sure the statistical ecology of a type of menu, e.g., menu items lengths, menu group sizes, semantical relevance between items of all Mac menus. However, it is harder to capture the frequency of all menus used by an individual. Better models could be built with knowledge of an individual's experience of the task.

The difficulty also comes from the generalisation between the local task and people's skills acquired from life experience. For example, there are two typical types of tasks in the visual search research, target-detection and target-localisation. In a target detection task, the decision response is whether there is a target present on the display. In a target-localisation task, there is always a designate target on the display. The task is to find the target. Sometimes, there may be multiple targets on the display. From the modelling perspective, the major difference between target-localisation task and target-detection task is in the hypothesis space. For a target-localisation task, each hypothesis concerns that the target is at one of the possible location. For a target-detection task, the fact that the target might not be on the display makes the hypothesis space much more complicated than that of target-localisation task. Specifically, the fact that the target might not be on the display makes the number of the possible displays doubles, which massively increases the modelling effort. However, there question is whether people really separate a target-detection task from a target-localisation task. The target search skill is one of the vital skills that human beings develop along their lives. They more or less have target-search experience in their life time. The certainty of whether the target is present is not always clear, and it is the point of search. For example, when finding the key, after a while searching on the desk, you might continue searching in your bag. Therefore, it is important to understand the extent to which the local task environment is representative of the general ecology. Do people really separate a target-detection task from a target-localisation task in the experiment or they are adaptive to a hybrid ecology of target-detection and target allocation?

### 7.1.3 Mechanisms

The study of psychological mechanisms, whether memory, attention, motion perception, or movement control is of course the core of what academic psychology is about. In the three tasks investigated in this thesis, I have focused on the mechanisms that are related to active vision. Active vision concerns questions such as what can be perceived in a fixation, how long is the saccadic duration between two fixations. By applying the optimal control analysis on three different tasks, the major lesson concerns how theories of information processing mechanisms can be tested by deriving optimal strategies. In particular, the model report in Chapter 4 is used to test two different theories of the uncertainty in human peripheral vision. Previously, the only example of where multiple theories of mechanism were tested by deriving optimal control solutions was Howes et al. (2009). Howes et al. tested theories of response selection bottlenecks. In this thesis I have extended the same method to test theories of signal processing in peripheral vision. Also I learnt that, the state representation and information integration technique carry some important theoretic roles. The availability of the information to the agent is constrained by both the tested theories of human visual and cognitive mechanisms and the task. For example, the visited item could be represented as its colour and shape, which means that the colour and shape information are encoded in the state representation. Alternatively, the visited item could be represented as one of the two values, 1 for target item and 0 for non-target item. This depends on the tested observation theories (for more details see section 3.2). Also, for example, in the menu search task (Chapter 5) the state elements are a generic state variable, representing relevance, rather than rather than specific information such as menu item names.

In the remainder of this chapter I discuss limitations, outstanding questions and, opportunities for future work.

## 7.2 Future Work

Although the optimal control model met the three challenges, there are some issues outstanding.

One of issues is that the learning technique used in the model is computationally intractable for large problems. This intractability may partly be because of insufficiently sophisticated learning techniques (Q-learning). Q-learning is known for its simplicity. However, since that it uses tables to store the state-action values, it easily loses its viability for large problems. Therefore, a more efficient learning technique will be pursued in the future. The optimal control approach would greatly benefit from an efficient learning technique. For example, in the distractor ratio task model, a binary state representation was chosen, e.g., 1 for target item; 0 for non-target item. However, with more powerful learning techniques, the optimal control approach could actually be used to test the theory by controlling how fine or coarse the state representation. With Q-learning, it was not possible to deal with the 48-item displays with a fine representation of each item.

Another issue is how the model could be used to understand the behaviours before fully adapting to the task environment. I have emphasised that Q-learning used here is not a theoretical commitment. Its purpose is merely to find the optimal policy, it is not to model the process of learning and is therefore used to achieve methodological optimality (Oaksford & Chater, 1994; Gray et al., 2006; Chater, 2009). Therefore, in the three examples, the model focused on the behaviours after fully adaptation to the task. But how does the model cope with behaviours before full adaptation? According to Howes et al. (2009), one of the solutions is to consider *experiential limitation*. That is, instead of adapting to the characterisation of the task problem, the behaviours would adapt to the limited experience. In the Computational Rationality Framework terminology, while keeping Utility and Mechanism the same, the ecological environment that the mind adapts to is changed. Here the model then makes contact with research on experiential limitation (Hahn & Warren, 2009; Hahn, 2014).

Also, primary contributions of this thesis is the contrast between strategies that emerge

from the optimal control model and predefined heuristics. However, some of the heuristics that people exhibit may reflect culturally determined strategies that are communicated through language and education. They may not be discoverable through experiential learning. Production system architectures (e.g., Kieras & Meyer, 1997; Anderson et al., 1997; Laird, Newell, & Rosenbloom, 1987) and heuristic selection models (e.g., Rieskamp & Otto, 2006; Rieskamp, 2008) may be better models of the deployment of heuristics that are transmitted through communication/education. In the remainder of this section I discuss outstanding questions and opportunities for future work.

### 7.2.1 Optimal control versus optimal state estimation

One potential avenue for future work concerns the further exploitation of optimal control models for explaining human vision. Many recent advances in vision research have been based on optimal state estimation approaches (e.g., Najemnik & Geisler, 2005, 2008; Myers et al., 2013). But, it has become increasing clear that the saccadic eye movements and decision making are sensitive to the reward, which links fixation patterns to task demands.

The reward sensitivity of the eye movement circuitry provides the neural underpinnings for reinforcement learning models of behaviour (Montague, Hyman, & Cohen, 2004; Schultz, 2000). Given a set of possible states, and actions that might be associated with those states, reinforcement learning algorithms are able to obtain a policy for selecting actions that will ultimately maximise reward. One hypothesis is that part of deciding what to do next involves state estimation. However, our analysis shows that optimal state estimation is not required to explain some aspects of visual search behaviour. While Bayesian optimal state estimation perspective has been very useful but, arguably, to explain behaviour what is more crucial is a theory of control. Optimal state estimation approaches focus on how to represent the environment but neglect how to use this representation to guide the selection of actions, e.g., Bayesian observer/analysis of the environment, which offers an option to represent the environment and thus estimate the

state of the world, but the action selection is not done by maximising local information gain rather rather actions are selected according to an optimal policy that has been trained through reward maximisation in the adaptation environment.

In the future, for example, for an extended optimal control model of Distractor Ratio task, it will need a fuller comparison to the Bayesian model for the same task reported by Myers et al. (2013). Comparison to the performance of the Bayesian model will help expose the relevant properties of both.

Another natural goal of this work is to integrate and extend the menu search and visual search models. This would be a potentially useful cognitive science approach to HCI. Perhaps it could be extended to modelling visual search in more applied tasks such as icon search on a smart phone (Fleetwood & Byrne, 2002; Kieras & Hornof, 2014), or image search in a browser (Sclaroff, Taycher, & La Cascia, 1997; Cai, He, Li, Ma, & Wen, 2004; Yan, Kumar, & Ganesan, 2010).

### 7.2.2 Comparison with other cognitive architecture approaches

The study of psychological mechanisms, whether memory, attention, motion perception, or movement control is of course the core of what academic psychology is about. Computational approaches to models of psychological mechanisms have given rise to a number of architectures. These include production system architectures (using if-then rules and procedures that operates on them to explain thinking (e.g., Rosenbloom et al., 1993; Kieras, 1997; Anderson et al., 1997, 2004) and connectionist type architectures (using artificial neural networks (e.g., McClelland et al., 1986, 2010)) amongst others.

EPIC (Executive-Process/Interactive Control) (Kieras & Meyer, 1997) is one of the most representative production system based cognitive architectures. Based on the architecture, perceptual, motor, and cognitive processing constraints are integrated. Then the architecture, which is defined as a set of mechanisms, is then able to process information and generate behaviours. Predictions are derived by programming the architecture with task knowledge and executing the program to generate simulated behaviour. For exam-

ple, EPIC consists of a production-rule cognitive processor and perceptual-motor peripherals. Information from the environment enters the model through simulated eyes, ears and hands; moves into corresponding visual, auditory, and tactical perceptual processors. Information from perceptual processors is deposited into working memory.

A benefit of cognitive architectures for the modeler is that a rich set of constraints have been built into them. However, for example, for the EPIC cognitive architecture, the cognitive processor is based on a production system. The cognitive strategies are represented by a list of production rules, which needs to be hand-crafted and pre-programmed into the model. This is much the same way that Cognitive Complexity Theory (Bovair, Kieras, & Polson, 1990), ACT-R (Anderson et al., 1997), and SOAR (Laird et al., 1987) represents procedural knowledge. Therefore, a thorough search of the space of strategies heavily depends on the knowledge of the programmer, rather than a formal problem specification. Therefore, the cognitive architectures are hard to test because of this flexibility of the embedded production rules (Howes et al., 2009).

Although the optimal control approach says less about what the particular theory of constraints are it does, as demonstrated, provide a means of *testing* constraints. And to emphasis again, instead of pre-programming the strategies, the optimal control approach allows the strategies to emerge as adaptation to the information processing constraints and reward. Therefore, the optimal control approach complements the cognitive architecture approach. It may, for example, be interesting in the future to build a version of a cognitive architecture that simulates perceptual, motor and memory processes but which instead of a production system has mechanisms for deriving optimal policies.

### 7.2.3 Cognitive constraints

The optimal control approach explains human behaviours based on utility maximisation given some low level information processing limits. Therefore, an extensive understanding of the cognitive constraints imposed by mechanism and utility is crucial. In the following I list some aspects for potential future directions.

### Active vision

Findlay and Gilchrist (2003); Halverson and Hornof (2011) presented some key questions to be investigated for active vision: (a) What information in the environment can be perceived at each fixation? (b) What information is integrated between eye movements? (c) Where do we move our eyes? (d) When do we move our eyes?

***What can be perceived in a fixation*?**

What information is available at one fixation? In the Distractor Ratio task model, the availability of the colour and shape at each fixation degraded as a function of eccentricity. In the menu Search task, the availability of 'text' and 'word shape' were different. Text can be perceived up to 1 degree of visual angle from the centre of gaze. The shape (length) information of the items degraded continuously as a function of eccentricity.

However, there are various findings about the availability functions of different features that were not integrated into the menu search model. Halverson and Hornof (2011) suggested that the defaulted availability of text could be replaced as a continuous function in which the availability of text degrades as a function of eccentricity. Kieras and Hornof (2014) gave an sketch of the probability that different features, e.g., colour, shape, size of the object can be perceived given the eccentricity (see Figure 2 in Kieras and Hornof (2014)). The optimal control approach could be used to test/identify alternative availability functions in the future. For example, the model reported in Chapter 4 could be extend by testing assumptions about relative availability of different features, i.e., colour and shape. It could also be extended to a broader range of tasks including *icon search* tasks (Fleetwood & Byrne, 2002; Kieras & Hornof, 2014). Each icon on smart phones has its colour, pattern, and/or text. The search for a certain icon would be affected by the notion of 'the ratio' in the Distractor Ratio task, e.g., the ratio between the same-colour distractors to other-colour distractors. In order to extend our model for the icon search, the availability function for different features need to be further investigated and integrated.

***What information is integrated between eye movements*?**

This question is related to how working memory affects visual search and learning.

For the optimal control approach, the answer to this question is in two parts: (1) the state representation at one fixation; (2) how these states are updated/integrated across fixations.

How fine or coarse does the state representation need to be? For example, in the Distractor Ratio task, if the theory is that people are able to remember the colour and shape of the objects they have seen, the state representation needs to be fine enough to distinguish the colour and shape of the objects. If the theory is that people are not able to remember the colour and shape of the objects they visited, but only remember whether there is the target in the visited locations, then the state representation only needs a binary representation for each location (e.g., 1 for target object and 0 for non-target object). Research has shown that visual search processes use spatial working memory[1] (Oh & Kim, 2004) but not other memory systems (e.g., verbal working memory (Logan, 1979), semantic working memory (Altarriba, Kambe, Pollatsek, & Rayner, 2001), nor visual working memory (Vogel, Woodman, & Luck, 2001)). The spatial working memory used during visual search has been shown to be somewhat coarse (Irwin, 1996). A possible use of a coarse spatial memory is to help select saccade destinations that are away from previous fixation locations (Klein & MacInnes, 1999). Halverson and Hornof (2011) also showed that the model that only maintain a high overview of what has been seen and has not yet been fixated works well to explain human data. This supports the fact that optimal control model for Distractor Ratio Task used a binary state representation, i.e., 1 for target and 0 for non-target.

However, *environmental context-dependent memory* effects show that human beings remember some other things other than the target relevance (for a review see S. M. Smith & Vela, 2001). Therefore some future work in this aspect is required.

The work present here investigated several ways to integrate information across fixations. For example, in Chapter 4, the information across fixations was integrated according to the error variances (the measurement noise), and feature information for each

[1] Spatial memory is the part of memory responsible for recording information about one's environment and its spatial orientation. For example, a person's spatial memory is required in order to navigate around a familiar city.

location was updated independently. In Chapter 5, 6, the information was simply integrated by summing up what have seen so far. In the future, alternative ways could be tested, for example, a Bayesian state estimation could be tested.

***Where do we move our eyes*?**

Many factors affecting the decision of where to fixate next have been proposed (for a review see e.g., Wolfe & Horowitz, 2004). In contrast to the top-down and bottom-up dichotomous perspectives, the *state estimate and optimal control* approach proposed an integrated perspective to answer this question. Where to fixate next is determined by the optimal solution to the specified problem. The control problem is shaped by three factor (1) Ecology: ecological environment as human beings experienced, (2) the information processing mechanism and, (3) Utility function: the desire of what to achieve. This question has been intensively discussed earlier through the thesis.

For the future work, the optimal control approach could be applied to model additional phenomena. For example, centre of gravity effects have been observed in human visual search (Najemnik & Geisler, 2005, 2008). Centre-of-gravity refers to the fixations on/near the centroid of multiple locations that are likely to have the target. For the optimal control approach, centre-of-gravity effects make much sense as the fixation to the centroid location benefits for the long term reward. Centre-of-gravity effects will emerge from control optimisation to the extent that focusing between likely target locations leads to increased reward.

***When do we move our eyes?***

When do we move our eyes or how long should the eyes linger on elements in a display? The three models presented here did not directly explore this question. In the menu search model, the fixation duration was set to be $400$ ms. It is known that the average fixation duration is $200-250$ ms (Wolfe & Gancarz, 1997; Rayner, 1998). However, menu search involves some matching process and so some additional latency per menu item gaze is expected. Hence, fixation duration in the menu search model was set according to the mean duration of item gazes in a typical menu search task (D. P. Brumby et al., 2014).

Future work should resolve this discrepancy with a deeper account of fixation duration in the menu search.

In fact, there are various findings about fixation durations. Some claim that the fixation durations are fairly constant, e.g., $200 - 250$ ms (Wolfe & Gancarz, 1997; Rayner, 1998). Others claim that varying time fixation durations. For example, the fixation durations are based on the number, proximity, and similarity of objects near the point of gaze (Lohse, 1993). Also, in a icon search model, Halverson and Hornof (2011) claimed that the fixation durations were longer for dense groups of icons than sparse groups. It is possible that the optimal control model could be extended to explain these phenomena.

In Chapter 5, Figure 5.13 showed the response time distributions for the task. Long-tail distributions are a signature feature of human response times and despite the fact that the no skewed distributions are assumed in the model performance they do emerge as a consequence of adaptation to the constraints. However, the model was slower than the participants. This is one motivation for us to do further work on this problem.

## 7.3 Conclusion

This thesis aimed to demonstrate that an optimal control approach to explaining human behaviour could give a single answer to both 'why' questions (why do people exhibit behaviour X), and 'how' questions (how does the mind/brain generate behaviour X). A *state estimate and optimal control* approach was investigated, in which human information processing mechanisms/constraints (how explanations) were embedded in the state estimation, and a rational analysis (why explanations) based on the state estimation was then provided by the optimal control policy that governs human behaviours. Three examples supported the hypothesis that human behaviours can be understood as the consequence of the optimal control policy that emerged from adaptation to information processing mechanisms and reward.

# References


Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent spanish-english bilinguals. *Perception & Psychophysics*, *63*(5), 875–890. 134

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press. 1, 2

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–60. doi: 10.1037/0033-295X.111.4.1036 2, 131

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462. 2, 9, 130, 131, 132

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, *2*(6), 396–408. 10, 16

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14*(7), 589–592. 49

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 948. 5, 37, 39, 40

Bailly, G., Oulasvirta, A., Brumby, D. P., & Howes, A. (2014). Model of visual search and selection time in linear menus. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 3865–3874). 5, 68, 72, 74, 86, 87, 88, 91

Bailly, G., Oulasvirta, A., Kötzing, T., & Hoppe, S. (2013). Menuoptimizer: Interactive optimization of menu systems. In *Proceedings of the 26th annual acm symposium on user interface software and technology* (pp. 331–342). 66, 79, 92

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Cognitive Neuroscience, Journal of*, *7*(1), 66–80. 15

Ballard, D., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(04), 723–742. 15

Baloh, R. W., Sills, A. W., Kumley, W. E., & Honrubia, V. (1975). Quantitative measure-

ment of saccade amplitude, duration, and velocity. *Neurology*, *25*(11), 1065–1065. 73

Baron, S., Kleinman, D., & Levison, W. (1970). An optimal control model of human response part ii: prediction of human performance in a complex task. *Automatica*, *6*(3), 371–383. 12, 18

Baron, S., & Kleinman, D. L. (1969a). The human as an optimal controller and information processor. *Man-Machine Systems, IEEE Transactions on*, *10*(1), 9–17. 3, 7, 12, 18, 38

Baron, S., & Kleinman, D. L. (1969b). The human as an optimal controller and information processor. *Man-Machine Systems, IEEE Transactions on*, *10*(1), 9–17. 68, 69

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*(3), 439–449. 8, 9, 96

Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-the-best models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 107. 119

Borah, J., Young, L. R., & Curry, R. E. (1988). Optimal estimator model for human spatial orientationa. *Annals of the New York Academy of Sciences*, *545*(1), 51–73. 18

Bovair, S., Kieras, D. E., & Polson, P. G. (1990). The acquisition and performance of text-editing skill: A cognitive complexity analysis. *Human-Computer Interaction*, *5*(1), 1–48. 132

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, *138*(3), 389. 2

Brehmer, B. (1979). Note on hypothesis testing in probabilistic inference tasks. *Scandinavian Journal of Psychology*, *20*(1), 155–158. 102

Brehmer, B. (1980). In one word: Not from experience. *Acta psychologica*, *45*(1), 223–241. 109

Bröder, A. (2000). Assessing the empirical validity of the" take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1332. 96, 102, 104

Bröder, A. (2011). The quest for take the best—insights and outlooks from experimental research. *Heuristics: The foundations of adaptive behavior*, 364–382. 95, 117

Bröder, A., & Schiffer, S. (2006, July). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of experimental psychology. Learning, memory, and cognition*, *32*(4), 904–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16822156 doi: 10.1037/

0278-7393.32.4.904 94, 95, 97, 117, 122

Brumby, D., & Howes, A. (2008, January). Strategies for Guiding Interactive Search: An Empirical Investigation Into the Consequences of Label Relevance for Assessment and Selection. *Human-Computer Interaction*, *23*(1), 1–46. doi: 10.1080/07370020701851078 66

Brumby, D. P., Cox, A. L., Chung, J., & Fernandes, B. (2014). How does knowing what you are looking for change visual search behavior? In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 3895–3898). 66, 73, 89, 135

Brumby, D. P., & Howes, A. (2004). Good enough but I'll just check: Web-page search as attentional refocusing. *Proceedings of the International Conference on Cognitive Modeling*, 46–51. 67, 72, 121

Brumby, D. P., Salvucci, D. D., & Howes, A. (2009). Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proceedings of the 27th international conference on human factors in computing systems* (pp. 1629–1638). 65

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Univ of California Press. 1

Busoniu, L., Babuska, R., De Schutter, B., & Ernst, D. (2010). *Reinforcement learning and dynamic programming using function approximators* (Vol. 39). CRC press. 25

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, *55*(1), 41–84. 66, 67

Cai, D., He, X., Li, Z., Ma, W.-Y., & Wen, J.-R. (2004). Hierarchical clustering of www image search results using visual, textual and link information. In (pp. 952–959). ACM. 131

Charman, S. C., & Howes, A. (2003). The adaptive user: an investigation into the cognitive and task constraints on the generation of new methods. *Journal of experimental psychology. Applied*, *9*(4), 236–48. doi: 10.1037/1076-898X.9.4.236 65

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition*, *113*(3), 350–64. doi: 10.1016/j.cognition.2008.06.014 23, 37, 129

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. 1

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational behavior and human decision processes*, *90*(1), 63–86. 95

Chen, X., & Howes, A. (2012). A Reinforcement Learning Model of Bounded Optimal Strategy Learning. In *International conference on cognitive modeling*. Berlin. 119

Chen, X, Howes, A., Lewis, R.L., Myers, C.W., Houpt. (2013). Discovering computationally rational eye movements in the distractor ratio task. In *Reinforcement learning and decision making*. Princeton. 69, 92

Christensen-Szalanski, J. J. (1978). Problem solving strategies: A selection mechanism, some implications, and some data. *Organizational Behavior and Human Performance*, *22*(2), 307–323. 9, 96

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive psychology*, *30*(1), 39–78. 63

Cockburn, A., Kristensson, P. O., Alexander, J., & Zhai, S. (2007). Hard lessons: effort-inducing interfaces benefit spatial learning. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1571–1580). 92

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 429–453. 37

Donkin, C., & Shiffrin, R. (2011). Visual search as a combination of automatic and attentive processes. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 2830–2835. 63

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, *115*(1), 199. 96

Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 39–48). ACM. 97

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32. 39

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Journal of Accounting Research*, 1–31. 109

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, *112*(4), 912. 95

Erev, I., & Gopher, D. (1999). A cognitive game-theoretic analysis of attention strategies, ability, and incentives. *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*, 343–371. 8

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American economic review*, 848–881. 8

Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*(2-3), 101–118.

43

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing.* Oxford University Press. 133

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, *3*(4), 552. 109

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381. 16

Fleetwood, M. D., & Byrne, M. D. (2002). Modeling icon search in act-r/pm. *Cognitive Systems Research*, *3*(1), 25–33. 131, 133

Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user : stable suboptimal performance in interactive tasks. *Cognitive Science*, *28*, 901–935. doi: 10.1016/j.cogsci.2004.03.005 65

Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology*, *52*(3), 195–242. 96, 114, 117, 118

Fu, W.-T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human–Computer Interaction*, *22*(4), 355–412. 65, 67, 68, 72

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision research*, *51*(7), 771–781. 13, 14, 15, 35

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451–482. 94, 95

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, *103*(4), 650. 6, 94, 95, 97, 122

Gigerenzer, G., & Selten, R. (2001). Rethinking rationality. *Bounded rationality: The adaptive toolbox*, 1–12. 95, 97, 122

Gigerenzer, G., & Todd, P. M. (1999). *Fast and frugal heuristics: The adaptive toolbox.* Oxford University Press. 95

Glöckner, A., Betsch, T., & Schindler, N. (2010). Coherence shifts in probabilistic inference tasks. *Journal of Behavioral Decision Making*, *23*(5), 439–462. 91, 96

Golomb, J. D., L'Heureux, Z. E., & Kanwisher, N. (2014). Feature-binding errors after eye movements and shifts of attention. *Psychological Science*, *25*(5), 1067–1078. 43, 44

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: a rational analysis approach to resource allocation for interactive behavior. *Psychological review*, *113*(3), 461–82. doi: 10.1037/0033-295X.113.3.461 7, 15, 16, 23, 78, 101, 129

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Proba-

bilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. 1, 2

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the bayesians got their beliefs (and what those beliefs actually are): comment on bowers and davis (2012). *Psychological Bulletin*. 1

Hahn, U. (2014). Experiential limitation in judgment and decision. *Topics in cognitive science*, *6*(2), 229–244. 129

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review*, *116*(2), 454. 10, 129

Halverson, T., & Hornof, A. J. (2007). A minimal model for predicting visual search in human-computer interaction. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 431–434). 66

Halverson, T., & Hornof, A. J. (2011). A computational model of "active vision"for visual search in human–computer interaction. *Human–Computer Interaction*, *26*(4), 285–314. 67, 133, 134, 136

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in cognitive sciences*, *9*(4), 188–194. 35

Hayhoe, M., & Ballard, D. (2014). Modeling task control of eye movements. *Current Biology*, *24*(13), R622–R628. 35, 37, 62, 69, 120

Henderson, C. M., & McClelland, J. L. (2012). Common spatial characteristics of illusory conjunctions and crowding. *Journal of Vision*, *12*(9), 340–340. 45

Hilbig, B. E. (2010). Reconsidering "evidence"for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, *17*(6), 923–930. 96

Hornof, A. J. (1999). *Computational models of the perceptual, cognitive, and motor processes involved in the visual search of pull-down menus and computer screens*. Unpublished doctoral dissertation. 121

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 249–256). 66

Howes, A., Lewis, R. L., & Singh, S. (2014). Utility maximization and bounds on human information processing. *Topics in cognitive science*, *6*(2), 198–203. 37

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751. doi: 10.1037/a0017187 3, 7, 12, 17, 37, 62, 92, 118, 129, 132

Howes, A., Vera, A., Lewis, R. L., & McCurdy, M. (2004a). Cognitive constraint modeling: A formal approach to supporting reasoning about behavior. *Proc. Cognitive*

*Science Society*, 595–600. 17

Howes, A., Vera, A., Lewis, R. L., & McCurdy, M. (2004b). Cognitive constraint modeling: A formal approach to supporting reasoning about behavior. *Proc. Cognitive Science Society*, 595–600. 65

Inhoff, A. W., & Rayner, K. (1980). Parafoveal word perception: A case against semantic preprocessing. *Perception & Psychophysics*, *27*(5), 457–464. 73

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, 94–100. 134

Itti, L. (2007). Visual salience. *Scholarpedia*, *2*(9), 3327. 35

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10), 1489–1506. 35, 63

Janssen, C. P., Brumby, D. P., Dowell, J., Chater, N., & Howes, A. (2011). Identifying optimum performance trade-offs using a cognitively bounded rational analysis model of discretionary task interleaving. *Topics in Cognitive Science*, *3*(1), 123–139. 65

Johnson, M. (1986). Color vision in the peripheral retina. *American journal of optometry and physiological optics*, *63*(2), 97–103. 49

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2011). Don't stop 'til you get enough: adaptive information sampling in a visuomotor estimation task. In (pp. 2854–2859). 114, 118

Kaptein, N. A., Theeuwes, J., & Van der Heijden, A. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1053. 39

Kieras, D. E. (1997). Building Cognitive Models with the EPIC Architecture for Human Cognition and Performance The Presenters Acknowledgments. *Electrical Engineering*, *12*, 48109–48109. 2, 9, 131

Kieras, D. E., & Hornof, A. J. (2014). Towards accurate and practical predictive models of active-vision-based visual search. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 3875–3884). ACM. 49, 67, 69, 74, 92, 131, 133

Kieras, D. E., & Meyer, D. E. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, *12*(4), 391–438. 130, 131

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological science*, *10*(4), 346–352. 134

Kleinman, D., Baron, S., & Levison, W. (1970). An optimal control model of human response part i: Theory and validation. *Automatica*, *6*(3), 357–369. 12, 18, 19, 20

Kowler, E. (2011). Eye movements: The past 25years. *Vision research*, *51*(13), 1457–1483. 35

Kuo, A. D. (1995). An optimal control model for analyzing human postural balance. *Biomedical Engineering, IEEE Transactions on*, *42*(1), 87–101. 18

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence*, *33*(1), 1–64. 130, 132

Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best"and the "rational"models. *Psychonomic Bulletin & Review*, *11*(2), 343–352. 95, 117, 119

Lee, M. D., & Zhang, S. (2012). Evaluating the coherence of take-the-best in structured environments. *Judgment and Decision Making*, *7*(4), 360–372. 95, 117

Lelis, S., & Howes, A. (2008). A Bayesian Model of How People Search Online Consumer Reviews. *Proc. CogSci 2008*, 553–558. 65

Lelis, S., & Howes, A. (2011). Informing decisions: how people use online rating information to make choices. In *Proceedings of the 2011 annual conference on human factors in computing systems* (pp. 2285–2294). 68

Lewis, R., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in cognitive science*, *6*(2), 279–311. 1, 3, 7, 10, 12, 37, 62, 78

Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(2), 189. 134

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. *Human-Computer Interaction*, *8*(4), 353–388. 136

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *System Sciences, 1996., Proceedings of the Twenty-Ninth Hawaii International Conference on,*, *4*, 86–97. 97

MacKay, D. J. (1995). Probable networks and plausible predictions-a review of practical bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, *6*(3), 469–505. 2

Marr, D. (1982). *Vision*. San Francisco: WH, Freeman and Company. 1, 2

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: cognitive aging and the adaptive selection of decision strategies. *Psychology and aging*, *22*(4), 796. 119

McClelland, J. L. (1988). Connectionist models and psychological evidence. *Journal of Memory and Language*, *27*(2), 107–123. 2

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Sei-

denberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, *14*, 348–356. doi: 10.1016/j.tics.2010.06.002 1, 2, 131

McClelland, J. L., Rumelhart, D. E., & Group, P. R. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, *2*. 2, 9, 131

McRuer, D. (1980). Human dynamics in man-machine systems. *Automatica*, *16*(3), 237–253. 12, 18

McRuer, D. T., & Jex, H. R. (1967). A review of quasi-linear pilot models. *Human Factors in Electronics, IEEE Transactions on*(3), 231–249. 19

Meyer, D., & Kieras, D. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*(1), 3. 2

Miller, C. S., & Remington, R. W. (2004). Modeling information navigation: Implications for information architecture. *Human-computer interaction*, *19*(3), 225–271. 66, 67, 72, 84, 85, 121

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767. 130

Myers, C. W., Gray, W. D., & Sims, C. R. (2011). The insistence of vision: Why do people look at a salient stimulus when it signals target absence? *Visual Cognition*, *19*(9), 1122–1157. 35, 36

Myers, C. W., Lewis, R. L., & Howes, A. (2013). Bounded optimal state estimation and control in visual search: Explaining distractor ratio effects.. 4, 41, 130, 131

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–391. 4, 7, 14, 15, 19, 36, 59, 130, 135

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3), 4. 4, 13, 14, 15, 35, 36, 41, 59, 130, 135

Neal, R. M. (1995). Bayesian learning for neural networks.
2

Newell, A. (1982). The knowledge level. *Artificial intelligence*, *18*(1), 87–127. 1, 2

Newell, B., & Shanks, D. (2003). Take the best or look at the rest? factors influencing" one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 53. x, 95, 97, 98, 100, 102, 103, 104, 105, 107, 108, 109, 110, 111, 112, 117, 118, 122

Newell, B., Weston, N., & Shanks, D. (2003, May). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, *91*(1), 82–96. Retrieved from http://linkinghub

.elsevier.com/retrieve/pii/S0749597802005253 doi: 10.1016/S0749-5978(02)00525-3 94, 95, 117

Newell, B. R. (2005). Re-visions of rationality? *Trends in cognitive sciences*, *9*(1), 11–15. 96

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making*, *17*(2), 117–137. 95, 117

Nunez-Varela, J., & Wyatt, J. L. (2013). Models of gaze control for manipulation tasks. *ACM Transactions on Applied Perception (TAP)*, *10*(4), 20. 37, 69, 120

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608. 23, 129

Oh, S.-H., & Kim, M.-S. (2004). The role of spatial working memory in visual search efficiency. *Psychonomic Bulletin & Review*, *11*(2), 275–281. 134

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive psychology*, *35*(1), 34–70. doi: 10.1006/cogp.1997.0676 65

Osterberg, G. (1935). *Topography of the layer of rods and cones in the human retina*. Nyt Nordisk Forlag. 49

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, *42*(1), 107–123. 63

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*(3), 534–552. 9, 96, 97, 115

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press. 9, 96, 97, 114, 115

Payne, S., & Howes, A. (2013). Adaptive interaction: A utility maximization approach to understanding human interaction with technology. *Synthesis Lectures on Human-Centered Informatics*, *6*(1), 1–111. viii, 8, 11, 65, 68, 92

Payne, S. J., Duggan, G., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY GENERAL*, *136*(3), 370. 118

Payne, S. J., Howes, A., & Reader, W. R. (2001). Adaptively distributing cognition: a decision-making perspective on human-computer interaction. *Behaviour & Information Technology*, *20*(5), 339–346. 65, 68

Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Information Theory, IRE Professional Group on*, *4*(4), 171–212. 13

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*

(Vol. 2). Oxford University Press, USA. 68, 72, 92

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*, 643–675. 65, 68, 72

Poisson, M. E., & Wilkinson, F. (1992). Distractor ratio and grouping processes in visual conjunction search. *Perception*, *21*(1), 21–38. 39

Pomplun, M., Reingold, E. M., & Shen, J. (2003). Area activation: A computational model of saccadic selectivity in visual search. *Cognitive Science*, *27*(2), 299–312. 35

Rajashekar, U., Cormack, L. K., & Bovik, A. C. (2002). Visual search: Structure from noise. In (pp. 119–123). ACM. 59

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372. 69, 73, 135, 136

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual cognition*, *7*(1-3), 17–42. 43

Rieskamp, J. (2008). The importance of learning when making inferences. *Judgment and Decision Making*, *3*(3), 261–277. 94, 95, 96, 97, 117, 122, 130

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta psychologica*, *127*(2), 258–276. 94, 95, 96, 97, 117, 122

Rieskamp, J., & Otto, P. E. (2006). Ssl: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207. 94, 95, 96, 97, 100, 117, 122, 130

Rosenbloom, P. S., Laird, J. E., & Newell, A. E. (1993). *The soar papers: Research on integrated intelligence, vols. 1 & 2.* The MIT Press. 9, 131

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and economic behavior*, *8*(1), 164–212. 8

Russell, S., & Subramanian, D. (1995). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, *2*, 575–609. 62, 69

Russell, S. J., & Subramanian, D. (1995). Provably bounded-optimal agents. *arXiv preprint cs/9505103*. 18

Scheibehenne, B., Rieskamp, J., & Wagenmakers, E.-J. (2013). Testing adaptive toolbox models: A bayesian hierarchical approach. *Psychological review*, *120*(1), 39. 96, 97, 117

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, *32*(3), 219–250. 10

Schultz, W. (2000). Multiple reward signals in the brain. *Nature reviews neuroscience*,

*1*(3), 199–207. 130

Sclaroff, S., Taycher, L., & La Cascia, M. (1997). Imagerover: A content-based image browser for the world wide web. In (pp. 2–9). IEEE. 131

Shen, J., Reingold, E. M., & Pomplun, M. (2000). Distractor ratio influences patterns of eye movements during visual search. *PERCEPTION-LONDON-*, *29*(2), 241–250. 5, 35, 37, 39, 41, 53, 54, 60

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *Autonomous Mental Development, IEEE Transactions on*, *2*(2), 70–82. 9, 119

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203–220. 134

Smith, V. L., & Walker, J. M. (1993). Rewards, experience and decision costs in first price auctions. *Economic Inquiry*, *31*(2), 237–244. 9, 96

Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *Journal of Vision*, *14*(5), 3. 43

Sprague, N., & Ballard, D. (2003). Eye movements for reward maximization. *Advances in neural information processing systems*, *16*. 69

Sprague, N., & Ballard, D. (2004). Eye movements for reward maximization. *Advances in neural information processing systems*, *16*, 1419–1433. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.2277&rep=rep1&type=pdf 69

Sprague, N., Ballard, D., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Transactions on Applied Perception (TAP)*, *4*(2), 11. 37, 120

Stengel, R. (1994). Optimal control and estimation",. *Dover, New York, USA*. 38

Stephens, D., & Krebs, J. (1986). Foraging theory, 1986. *Princeton: Princeton University Press*, *1*(10), 100. 1, 8

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13. 38, 49

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: an introduction*. MIT Press. 25, 26, 27, 28, 29, 69, 72, 77, 101

Swets, J., Tanner Jr, W., & Birdsall, T. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301–340. 38

Tanner Jr, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological review*, *61*(6), 401. 7, 12, 13

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, *50*(2), 184–193. 41

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, *114*(1), 71. 63

Todd, P. M., & Gigerenzer, G. (2001). Shepard's mirrors or simon's scissors? *Behavioral and Brain Sciences*, *24*(04), 704–705. 96

Torralba, A. (2003). Modeling global scene factors in attention. *JOSA A*, *20*(7), 1407–1418. 63

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. 44, 45

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. 43

Trommershäuser, J., Glimcher, P. W., & Gegenfurtner, K. R. (2009). Visual processing, learning and feedback in the primate eye movement system. *Trends in Neurosciences*, *32*(11), 583–590. 36, 37, 62, 120

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2009). The Expected Utility of Movement. *Brain*, 95–111. 69, 92

Tseng, Y.-C., & Howes, A. (2008). The adaptation of visual search strategy to expected information gain. In *Proceeding of the twenty-sixth annual chi conference on human factors in computing systems - chi '08* (pp. 1075–1084). New York, New York, USA: ACM Press. doi: 10.1145/1357054.1357221 68

Vera, A., Howes, A., McCurdy, M., & Lewis, R. L. (2004). A constraint satisfaction approach to predicting skilled interactive cognition. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 121–128). 17, 65, 68, 92

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92. 134

Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior: 3d ed*. Princeton University Press. 8

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual review of psychology*, *60*, 53–85. 1

Wolfe, J. M. (2007). Guided search 4.0. *Integrated models of cognitive systems*, 99–119. 35

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. *The Oxford Handbook of Attention*, 11. 35, 44

Wolfe, J. M., & Gancarz, G. (1997). Guided search 3.0. In *Basic and clinical applications of vision science* (pp. 189–192). Springer. 135, 136

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. 135

Wolfe, J. M., Reinecke, A., & Brawn, P. (2006). Why don't we see changes? the role of attentional bottlenecks and limited visual memory. *Visual Cognition*, *14*(4-8), 749–780. 45

Yan, T., Kumar, V., & Ganesan, D. (2010). Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In (pp. 77–90). ACM. 131

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: toward a rational bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 700. 38, 42, 43

Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., & Ballard, D. H. (1997). Eye movements reveal the spatiotemporal dynamics of visual search. *Psychological Science*, 448–453. 59

Zhang, Y., & Hornof, A. J. (2014). Understanding multitasking through parallelized strategy exploration and individualized cognitive modeling. In *Proceedings of the 32nd annual acm conference on human factors in computing systems* (pp. 3885–3894). 65

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision. *Perception*, *18*(2), 191–200. 37, 39