APPROXIMATE OPTIMAL CONTROL MODEL FOR VISUAL SEARCH TASKS

by

ADITYA ACHARYA

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

School of Computer Science
College of Engineering and Physical Sciences
University of Birmingham
January 2019

ABSTRACT

Visual search is a cognitive process that makes use of eye movements to bring the relatively high acuity fovea to bear on areas of interest to aid in navigation or interaction within the environment. This thesis explores a novel hypothesis that human visual search behaviour emerges as an adaptation to the underlying human information processing constraint, task utility and ecology. A new computational model (Computationally Rational Visual Search (CRVS) model) for visual search is also presented that provides a mathematical formulation for the hypothesis. Through the model, we ask the question, what mechanism and strategy a rational agent would use to move gaze and when should it stop searching?

The CRVS model formulates the novel hypothesis for visual search as a Partially Observable Markov Decision Process (POMDP). The POMDP provides a mathematical framework to model visual search as a optimal adaptation to both top-down and bottom-up mechanisms. Specifically, the agent is only able to partially observe the environment due to the bounds imposed by the human visual system. The agent learns to make a decision based on the partial information it obtained and a feedback signal. The POMDP formulation is very general and it can be applied to a range of problems. However, finding an optimal solution to a POMDP is computationally expensive. In this thesis, we use machine learning to find an approximately optimal solution to the POMDP. Specifically, we use a deep reinforcement learning (Asynchronous Advantage Actor-Critic) algorithm to solve the POMDP.

The thesis answers the where to fixate next and when to stop search questions using three different visual search tasks. In Chapter 4 we investigate the computationally rational strategies for when to stop search using a real-world search task of images on a web page. In Chapter 5, we investigate computationally rational strategies for where to look

next when guided by low-level feature cues like colour, shape, size. Finally, in Chapter 6, we combine the approximately optimal strategies learned from the previous chapters for a conjunctive visual search task (Distractor-Ratio task) where the model needs to answer both when to stop and where to search question.

The results show that visual search strategies can be explained as an approximately optimal adaptation to the theory of information processing constraints, utility and ecology of the task.

This theses is dedicated to my family.

ACKNOWLEDGEMENTS

The journey that I have been on to produce this thesis would not have been possible without the support of many people around me.

First and foremost, I would like to thank my supervisor Professor Andrew Howes, without him this journey would have never even started. I thank him for this constant support, guidance and discussions throughout my PhD. Also, for being the best mentor and a friend.

I thank my research group members Prof. Russell Beale, Prof. Robert Hendley and Prof. Jeremy L Wyatt. Their support and inputs are greatly appreciated. I also would like to thank Christopher W. Myers, Richard L. Lewis, Xiuli Chen and Chris Baber for their support and guidance during our collaboration.

Also I would like to thank my friends. Ashish, Ajit, Joy, Karthik, Lakshmi, Manu, Rameez, Sarthak and Sunit they have helped me through the stressful times and always been there with me. I thank Rajiv Singh and his family for their support and hospitality.

Finally, a special thanks to my family, parents, sister and brother-in-law. Thank you all for your support and understanding you have shown me for the past 4 years. Especially, my sister for all the times you have convinced me that I can do this.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Visual search is a cognitive process that makes use of eye movements to bring the relatively high acuity fovea to bear on areas of interest to aid in navigation or interaction within the environment. It is a fundamental and ubiquitous task that we perform in our daily lives. Infact, we search or look for things all the time (Eckstein, 2011). For example, searching for family or friends in the crowd or car in a parking lot. Also, object localisation and motor actions are often preceded by a soft search (Mennie, Hayhoe and Sullivan, 2007), like, locating and fixating on a coffee mug before moving the hands to grab it.

As modern society is interwoven so strongly with technology today, we spend an enormous amount of time searching through and looking at various displays. For example, searching and launching an app on smart-phone or using histogram visualisation on Google to find the cheapest flight tickets or using e-commerce sites like Amazon or eBay to find products online to purchase. Due to such a vast diversity of displays that we encounter in our everyday life, it is not surprising that visual search plays a pivotal role in our lives.

In addition to ecological importance, visual search also contributes to understanding higher-level cognition. In particular, visual search requires making a series of eye movement decisions due to the foveated vision property of human eye (Kowler, 2011). These eye movements involve the deployment of covert attention and overt attention (Wright and Ward, 2008). Visual search is arguably one of the most prominent paradigms used to study the deployment of covert and overt attention (Eckstein, 2011). In a typical experimental scenario, search performance is evaluated by varying number of elements in a display, which is then used to make inferences about different mechanisms involved behind the covert deployment of attention (Carrasco, 2011).

Moreover, in real-world rewards and costs are often associated with search. For example, in cases when people fail in finding the target, like in cancer detection, failing to find malignant tissues in x-rays may incur a high cost due to patient death. In some cases, multiple targets might have different cost or reward associated with it, like, searching for food products in a supermarket with different nutritional value. Previous research has used visual search paradigm to study the influence of reward on eye movement behaviour (Hikosaka, Takikawa and Kawagoe, 2000, Glimcher, 2003, Najemnik and Geisler, 2005, Della Libera and Chelazzi, 2009, Stritzke, Trommershäuser and Gegenfurtner, 2009, Eckstein, Schoonveld and Zhang, 2010, Navalpakkam, Koch, Rangel and Perona, 2010, Tseng and Howes, 2015). For example, Navalpakkam et al. (2010) showed that when fixating on an item in a display that is rewarded differently, people adapted their strategy to make saccadic movements towards more rewarding locations. Stritzke, Trommershäuser and Gegenfurtner (2009), showed that people are risk evasive, such that they direct their gaze to a more rewarding region and stay away from the region that incurs a cost.

While previous research has looked at these components individually to understand an isolated phenomenon, evidence suggests that an explanation of higher-level cognitive functions requires a combination of these individual components (Lewis, Howes and Singh, 2014). In this thesis, I explore a novel hypothesis that human visual search behaviour emerges as an adaptation to the underlying human information processing constraint, task utility and ecology, and present a computational model that integrates *ecology*, *mechanism* and *reward/utility* to understand high-level behaviour. Specifically, decision making for eye movements which includes the question of where to move the gaze next, and when to stop searching. Before going any further, we begin with an introduction to the visual search.

1.1 Visual Search

Visual Search is described as a perceptual task in which the cognitive system has to scan an environment for relevant information. In an experimental paradigm, participants are asked to find a target visual stimulus amongst other visual stimuli (distractors). The target is either present or absent on trial by trial basis, and the participant has to make a target present or target absent decision as quickly and accurately as possible. For example, searching for a 45-degree tilted Gabor patch in a high contrast background Najemnik and Geisler (2008), or searching for a coloured letter (e.g., red letter O) with is surrounded by some distractor coloured letters (e.g., red X's and green O's) Shen et al. (2003). These tasks typically involve multiple stimuli that compete for attention, placed randomly across a visual field. Since the brain has a limited capacity to process information simultaneously (Desimone and Duncan, 1995), attention is drawn to a limited set of stimuli for processing. Two questions that arises here is why do we direct our attention towards a particular object and neglect others in a visual field? In other words, how do people decide where and what to look next? Second, target in some cases may or may not be present in the environment. How do people decide when to stop searching?

1.1.1 When to terminate search?

In visual search experiments, it is not always necessary that the target be present. In fact, in real-world, searches not always succeed, especially when the targets are hard to find (Wolfe, Horowitz and Kenner, 2005). For example, in an airport security check, the security officer has to look at x-ray scan to search for dangerous items in the luggage. If the officer takes too long to look at each item, this might prolong the security queue, and on the contrary, if he terminates early, there is a risk of not recognising potentially dangerous item. The question is then when do people stop searching and what mechanism they use to make that decision?

Previous literature has suggested different strategies that might explain the search termination behaviour. The focus has been on identifying the utility of searching and the acceptable threshold of the utility, below which they stop searching. For example, the exhaustive search approach (Treisman and Gelade, 1980), where people decide on a target being absent after iterating through every item in the display. Nakayama et al. (1986), Zohary and Hochstein (1989) extended the idea to conjunctive visual search and showed that people search only a subset of stimulus exhaustively to make the decision. The subset selected is dependent on the distractor ratio size (Zohary and Hochstein, 1989). For example, in a display of 36 items with 5 red circles, 30 green cross and green circle as target, people exhaustively only search the minority set of features. However, the exhaustive search has been rejected in a study done by Wolfe and Horowitz (2004) (see (Wolfe, 2012) for a review). Wolfe (2012) described search termination in terms of the number of items searched or time spent in searching. The assumption is that people keep track of a noisy estimate of either of these two sources of information which is then used to formulate a decision threshold to terminate the search. The thresholds can be conservative or liberal (terminate early or later), which is decided by the number of goal-relevant items in the display, crowding and cluttering of items in the display or the value of target item (Wolfe, 2012).

As highlighted above stopping rule has predominately been studied through utility theory and heuristic thresholds are define as rules for termination. While heuristics thresholds do explain the termination on the specific task that the utility thresholds are fitted to. However, they require significant development work on the application of the models to new tasks.

1.1.2 Where to look next?

The guidance of attention or where to look next is considered to involve a *pre-attention* process (Müller and Krummenacher, 2006). The role of *pre-attention* process is to extract information from the visual field, which is then used to direct attention and the fovea to the relevant location.

Consider a scenario where a moving object or a bright light appears in our visual field. We immediately direct our attention towards it, these behaviours have been previously explained as a *bottom-up* or *stimulus-driven* deployment of attention. The *bottom-up* or *stimulus-driven* approach uses information processing mechanisms to explain the eye movement behaviour. They define architectures and algorithms (Koch and Ullman, 1987, Itti, Koch and Niebur, 1998) to explain the underlying mechanism used by the human cognitive system. For example, in the Feature integration theory (Treisman and Gelade, 1980) the incoming visual information is first received by the visual neurons and extracts basic features (colour, shape, orientation, etc.) of the stimulus present in the display. Features are then processed in parallel over entire visual field. The processed output is represented as a feature map which consists of visual features at a given location. Attention is then deployed to a region based on this map, and then objects are re-assembled in the given region to form more complex representation. This was further extended by Koch and Ullman (1987), they introduced the concept of *saliency map* which represented how much a given location differs from its surrounding visually. The selection of where to deploy attention next was based on a heuristic called *winner-take-all* (Pomplun, Reingold and Shen, 2003). According to *bottom-up* or *stimulus-driven* approach attention is directed to most salient regions in a display. They explain 'how' the eye movement behaviour emerges what underlying mechanisms are being used. However, they fail to explain 'why' people choose the salient object and 'why' sometimes they do not?

Alternatively, consider a situation where we are searching for our car in a parking lot, certain objects (other cars from the same manufacturer, colour or model) draws our attention towards them because we are searching for them. This way of attention deployment was previously explained as *top-down* or *goal-driven* (Yarbus, 1967). The *top-down* approach is a voluntary deployment of attention on certain features or objects that are relevant to the task. For example, Yarbus (1967) in a photograph viewing experiment asked the participants to view photographs with different goals. He reported that participants adapted their eye movements to part of the scene that was most task-relevant. The *top-down* approach focuses on explaining behaviour as an adaptation to task demands or the goal to achieve. These approaches have especially been used in explaining attention in natural tasks (Ballard, Hayhoe and Pelz, 1995, Land and Hayhoe, 2001, Hayhoe, Shrivastava, Mruczek and Pelz, 2003). The *top-down* approach provides a 'why' explanation of human behaviour. It assumes that the underlying system/architecture (human mind) is a black-box, and the actions can be predicted as rational adaptation to the environment.

There has been an on-going debate over which approach better explains the eye movement behaviour. Whether basic features (colour, shape, orientation, etc) are only extracted which is followed by a *bottom-up* or *stimulus-driven* deployment of attention. For example, human performance has been reproduced in a conjunctive visual search task (Itti and Koch, 2000) using *bottom-up* attention models. Also, attention allocation in natural scene like military car detection (Itti et al., 2001), face and motion detection (Ma et al., 2005), pedestrian detection (Miau et al., 2001), free viewing Wang et al. (2011) have been shown using *bottom-up* models. Or, features that are relevant to task demands are only extracted and a *top-down* or *goal-driven* attention deployment follows? Especially when people are engaged in a real-world task, *top-down* processing is seen as an influential factor for guiding attention in comparison to *bottom-up* processing (Borji et al., 2011). For example, attention deployment in tasks like block copying (Ballard et al., 1995), tea making (Land and Hayhoe, 2001), reading (Rayner, 1998), and object search (Navalpakkam and Itti, 2005) have been explained as *top-down* processing. As shown above, previous research has looked at these approaches independently, i.e., attention deployment is either purely *bottom-up* or *top-down* driven.

From a modelling perspective, by focusing only on the *top-down* approach, the model isolates itself from considering the underlying mechanisms that was used. In doing so, it fails to explain the information processing capacity that lead to the observed behaviour. On the other hand, the *bottom-up* approach is influenced by user defined mechanisms that are based on both modellers intuition and some empirical evidence.

Evidence suggests that both *bottom-up* and *top-down* processing is involved in visual search. For example, when searching for a red car in a parking lot, attention is increased to all the cars with the colour red. This suggests that interaction between *bottom-up* and *top-down* approaches are not mutually exclusive (Awh et al., 2012), rather they more often agree on. Neurological studies also suggest plausibility of large part of attention selection includes a mixed *bottom-up* and *top-down* approach (Corbetta and Shulman, 2002). Despite the evidence, interaction between *bottom-up* and *top-down* approaches have received limited attention. Our approach in this thesis follows this train of thought. Rather than approaching the visual search as a pure *bottom-up* or *top-down* processing, we present a computational model in this thesis that works on the interaction of both *bottom-up* and *top-down* approach. Specifically, we use the theoretical framework 'Computational Rationality' (Lewis et al., 2014) as a basis to build our model. According to this framework, human behaviour can be derived using cognitive mechanisms that are rationally adapted to both the mind and the environment (Lewis et al., 2014).

In this thesis we present a new computational model *Computationally Rational Visual Search* (CRVS) model that operationalises the 'Computational Rationality' framework as a *Partially Observable Markov Decision Problem* (POMDP). The framework provides a mathematical formulation for the eye movement problem (detailed overview provided in chapter 3). In the POMDP formulation, information perceived from the environment is incomplete and noisy. An *observation function* is defined to represent the noisy and incomplete information. In the model, the information processing constraints are encoded in this function. For example, in visual search task, the observations obtained on a fixation are constrained by the noise in human vision (ability to correctly perceive decreases with increase in eccentricity). These observations are then integrated across eye movements in a task-relevant representation known as *state estimation* (*bottom-up approach*). A solution to the POMDP problem determines the question of what to do next given the current state? The solution represents the optimal strategy given the information processing constraints imposed in the observation model. It is obtained by maximising task reward where the reward is constrained by the task demands (*top-down approach*). Thus, the new computational model presented here in this thesis provides a mathematical means to apply ‘Computational Rationality’ on visual search tasks. It is to be noted, the goal is to find the overall optimal strategy and not individual actions.

1.2 Problem Description

The sections above has previewed the two questions asked in visual search literature and the approaches used. What follows in this section is the formalisation of the problem description associated in the modelling of these question that this thesis solves.

- **How strategies emerge from mechanisms:** The first problem is to show how information processing mechanisms as a consequence of the biological architecture can give rise to the strategies without any description of rule or heuristics. For the *CRVS* model, the strategies emerge as an adaptation to the underlying assumptions about the mechanisms, utility and ecology. The aim here is to go beyond traditional approaches that only focus on utility and environment to describe behaviour and do not account for control of actions. For example, Bayesian approaches to visual search explain how to represent the environment, but neglect how to use this representation to guide attention. Rather, they use heuristics like ‘maximum a posteriori (MAP), to guide attention. The use of heuristics not only require domain expertise it is also a significant development work on the application of the models to new tasks.

- **How mechanisms can be tested:** The second problem is to show how theories of information processing mechanisms can be tested and what does it imply about behaviour. Testing theories for different mechanisms and evaluating against alternatives is difficult, especially to find correlation between a behaviour and the underlying mechanism. For example, one mechanism can generate multiple strategies, and it is these strategies that determine the behaviour. More often, different mechanisms can also generate very similar strategies. Therefore, it is not sufficient to just choose a mechanism and a strategy to describe behaviour. Rather, what is require is to determine which strategies are efficient given the choice of mechanism.

1.3 Thesis Contributions

- **A new computational model:** This thesis presents a new computational theory of visual search and operationalises it using the *CRVS* model that explains the human search behaviour. The model explains the eye movement strategies as an emergent consequence of ecology, reward and critically the architecture defined for the task. In doing so it avoids any predetermined rules or heuristics, and thereby is able to easily generalise across tasks. Furthermore, through the *CRVS* model we show how to combine both the *top-down* and *bottom-up* processing of visual information.

- **Framework to test theories:** Through the *CRVS* model we show how to test for perceptual noise originating due to the biological structure of the retina and what it tells about behaviour. the *CRVS* uses an optimisation algorithm (Deep Reinforcement Learning) to find an efficient strategies. By selecting a strategy through an optimisation algorithm, it allows a causal relationship between the theoretical assumption made in the model and the resulting behaviour that emerged.

- **Scale to real world tasks:** The model further contributes to the application of Deep Reinforcement Learning in solving visual search problems. In doing so, it is able to scale to real world tasks which consist of continuous and high dimensional state space and not limited to laboratory experimental tasks.

1.4 Thesis Outline

Chapter 2: This chapter provides an introduction to the 'Computational Rationality' framework. Also, reviews the literature on previous modelling approaches to the visual search.

Chapter 3: This chapter provides an overview of the mathematical background for the thesis. In particular, it provides an overview of the CRVS model, the underlying framework used, reinforcement learning, and state representation.

Chapter 4: In this chapter the CRVS model is applied to an image search task experiment by Tseng and Howes (2015). The model solves the control problem of when to stop the search in this chapter. To preview the model shows that the stopping strategy emerges as an adaptation to the user preference defined as the reward in the model and the task ecology which is defined as the skewed distribution of the number of target features in an image.

Chapter 5: In this chapter the CRVS model is applied to William's Object search task (Kieras et al., 2015*a*). The model solves the control problem of where to search next in this chapter. To preview the model shows that the where to search next strategy emerges as a consequence of the constraints imposed by human peripheral vision and memory.

Chapter 6: In this chapter the CRVS model is applied to the Distractor-Ratio Task (DR-Task) Shen et al. (2003). The model solves the control problem of both where to search next and when to stop in this chapter. To preview the model shows that the where to search next strategy emerges as a consequence of the constraints imposed by human peripheral vision (crowding effect) and task rewards.

Chapter 7: : The final chapter provides a summary of the primary outcomes of the study done in this thesis. Also, the contribution to cognitive science is discussed. Finally, future extension of the current work is highlighted.

CHAPTER 2

BACKGROUND

2.1 Introduction

The *CRVS model* presented here in this thesis is used to explain the visual search behaviour when performing any visual search task. Our approach here is to explain the search behaviour as a bounded optimal adaptation to constraints imposed by the human physiology and the task utility. Hence, we first systematically review the *computational rationality* framework (Lewis et al., 2014), which has been proposed as a means to explain behaviour as *bounded optimality* (Russell and Subramanian, 1995).

Also, we review alternative modelling approaches that explain visual search behaviour. In this thesis we try and classify these model as to how the eye movement strategy is derived, i.e., either using some form of heuristic algorithm or using an optimal approach. Also, in the review, we highlight for each of the alternative models, how the utility, ecology of the environment and mechanism is considered for deriving the search behaviour.

2.2 Computational Rationality

The *rational analysis* framework in cognitive science was coined by John Anderson to explain the cognitive processes of the human mind (Anderson, 1991). The framework uses rationality as an empirical tool to explain how and why people adapt to their environment. The framework is described as a 6-step process, in which, a specification of three factors are required, those are: goals, task environment and computational limitations. In the first step, the modeller specifies the task goals the cognitive system needs to achieve. Following which, a formal model of the external environment is defined that will be acted upon. In the third step, bounds or limitations are imposed/defined due to the statistics of the environment. As a consequence of these three factors, optimal behaviour is derived. The optimal behaviour is then evaluated against human data to validate the model. This process is iterated until a good fit for the human data is found. A key point to be noted here is, the framework assumes that people are optimal with respect to the environment, and the emphasis here is on the constraints imposed by the environment. An apparent drawback in this framework is that it fails to account for the mechanism involved. It explains the question of '*why*' people behave the way they do, but, is unable to answer '*how*' they achieve it and what architecture is being used.

Alternatively, the *bounded optimality* framework in artificial intelligence literature was proposed by Russell and Subramanian (1995). According to this framework, an artificial agent bounded by the computational resources available to it and the task environment behaves as well as possible given these bounds. Bounded optimality describes an agent's behaviour should be similar to an optimal program running on a bounded system/architecture while interacting with its environment where the program dictates what to do on receiving observation from its environment. By framing the problem using this framework, it provides the generality and flexibility to accommodate both the benefits of rationality to explain the '*why*' behaviours and the underlying architecture used to generate those behaviours.

Computational Rationality is an application of *bounded optimality* adapted to psychology (Lewis et al., 2014). It is based on the idea that behaviour is generated as an adaptation to the environment and the mind (architecture). In other words, it extends the question asked by rational analysis, i.e., *what behaviour should a rational agent ex-*

hibit while interacting with this environment? And includes the information processing bounds, i.e., *what should a utility maximizing rational agent do, when constrained by its information processing mechanism, in this environment?* Computational Rationality focuses on testing theories of the architecture and what does it imply about the behaviour. By describing behaviour in this way, the framework provides a medium for formulating theories using optimality as a tool and which in turn generates explanations that unify both the top-down rational approaches and bottom-up mechanism approaches. Therefore, Computational Rationality is the most suitable framework for approaching the modelling of visual search behaviours.

In the section below, we describe the individual component involved in this framework.

2.2.1 How Computational Rationality Works

The Computational Rationality framework can be summarised by a description of four components as shown in figure-2.1. According to the framework the strategy/behaviour people exhibit is dependent on the utility, ecology and mechanism involved. A successful explanation about the behaviour requires a theory of all the three components. Below, we review and define each component.

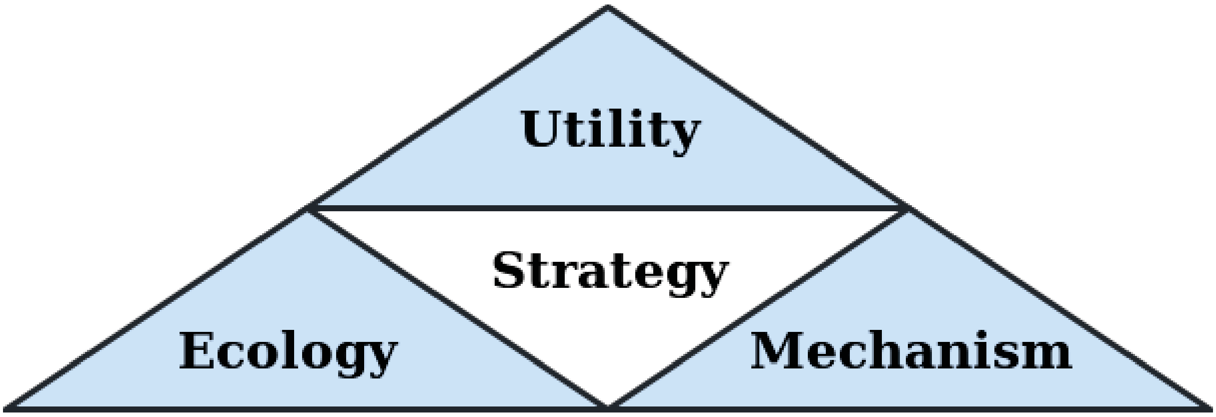


Figure 2.1: Diagram illustrates individual components of the Computational Rationality framework (Payne and Howes, 2013).

Utility. the concept of *utility* has more often found its roots in microeconomics, game theory and decision theory. Also, has been provided with different interpretation (Kahneman and Tversky, 2013, Von Neumann and Morgenstern, 2007). For example, the concept of utility as per the *expected utility theory* states that people under risk or uncertainty make decisions by comparing the expected value between choices (Von Neumann and Morgenstern, 2007) where the expected value is the statistical expectation of the value of the outcomes. Alternatively, the *prospect theory* (Kahneman and Tversky, 2013) states that people decide by taking into account potential profit or loss they may incur by making that choice rather than the end outcome.

In the Computational Rationality framework, the utility is thought as what is essential to people and makes similar assumptions about what people would value more. For example, in flight-based tasks people might value time more and may find strategies that minimise the amount of time spent. In other words, they may find strategies that trade-off speed-accuracy. The utility theory has many theoretical implications, where, defining utility as experimental goals (objective utility) through instructions can verify whether people's subjective utility was consistent with experiments objective utility. Alternatively, the subjective utility was entirely different, for example, people just wanted to finish the experiment quickly. Defining utility in this way is consistent with Singh et al. (2010) formulation of intrinsic motivation and reward. In this thesis we describe the utility in form of a reward function.

Ecology. In the framework, *ecology* is concerned with the statistical distribution of an environment that the user is currently interacting with and which he has interacted during his lifetime. Therefore, the agent that is described as computationally rational has to also adapt to the environment it has experienced thus far.

Anderson (1991) through the *rational analysis* framework showed that people do adapt their behaviour to the statistics of the environment. For example, in Schooler and Anderson (1997) they showed that the memory decay in humans was optimally adapted to a task environment where the task was to predict odds of a word encountered in the front-page headlines of the New York Times newspaper. In the experiment, the frequency of a word encountered followed a power function. This was consistent with the memory decay behaviour in humans (Schooler and Anderson, 1997).

Furthermore, as described by Tseng and Howes (2015), visual search strategies are adapted to the task ecology. For example, in a study done by Vlaskamp et al. (2005), they found that the distance between items in a display has a direct impact on the search performance, i.e., the amount of time fixated on an item, number of fixations and dwelled time. They showed that by controlling the distance between items (range between 1.5 to 7.1 visual angle), the number of fixations and fixation duration increases with an increase in distance. These experiments provide evidence that environment ecology plays an essential role in shaping human behaviour.

Mechanism. here concerns with the human information processing capacities (attention, perception, memory, motor control). In other words, *mechanism* defines the underlying architecture of the human mind. It explains the mapping of the sensory input (perceptual, auditory or smell) to how it is stored and processed and finally how that information is converted to interactive actions (eye movements, motor movements for a button press).

In cognitive psychology, the term *mechanism* or *mechanistic explanation* refers to identifying underlying *processes* involved in a particular phenomenon (Bechtel, 2008). Also, these *processes*/*mechanisms* are then further decomposed into atomic operations and an architecture is defined to explain the organisation and functioning of these *processes* (Bechtel, 2008). Over the years, many computational models have been proposed that make assumptions about the underlying information processing architectures. For example, cognitive architectures with production rule (Anderson, 1996, Kieras and Meyer, 1997), saliency models (Itti and Koch, 2001) , artificial neural networks (McClelland and Cleeremans, 2009). In each case, a set of *mechanisms* are defined assuming, that is how the human mind would process the information. Following which, a set of control programs are defined based on the task specifications to generate behaviour. By defining a theory for the underlying *mechanism*, inferences can be made about the behaviour as to how it emerged. Also, inferences can be made about what constraints were involved in explaining those behaviours.

Strategy. In the framework strategy is defined as a computational *program*. These *programs* define a sequence of actions that can be performed by an agent to interact and reach task-specific goals. There can be n number of programs with different combinations of actions taken in sequence, that achieve task goals. Computational rationality assumes that human mind is synonymous to a bounded computational unit that runs these *programs* (Lewis et al., 2014). While the behaviour that people exhibit is defined as the best/optimal *program* that is adapted to not only to the task environment but also to the bounded computational unit it is running on. Here, the best/optimal *program* is obtained using the principle of rationality. Payne and Howes (2013) state that in order to explain human behaviour theory of utility, mechanism and ecology must be provided. In case of any missing component, the strategy obtained would be significantly different from the actual behaviour (Payne and Howes, 2013).

In the framework, to find the optimal strategy a utility maximisation/minimisation approach is adopted. Where any formal optimisation approach can be used that involves utility maximisation/minimisation to find the optimal strategy. An important thing to note here is that there are no constraints or assumptions on the strategy. Instead, the best strategy emerges as a consequence of adaptation to the utility, ecology and mechanism. If the model with the best payoff strategy fits the observed human data, then it can be concluded that the model behaves correctly and can be used for further prediction or analysis of the behaviour. Otherwise, it can be concluded that the utility function or the underlying hypothesis about the human information-processing constraint that was encoded in the model is incorrect. Hence, a bad fit does not mean the rejection of bounded optimal behaviour. Instead, it supports rejection of the underlying assumptions made in the model.

The goal here in this thesis is to present a computational model that makes a testable prediction of the underlying behaviour. Also, combines the benefits of both *top-down* and *bottom-up* approaches. The Computational Rationality framework provides the framework to achieve this goal.

2.3 Models of Visual Search

In the sections below, we review some existing models of visual search. The focus here would be to highlight in each of the models how the ecology, utility and mechanisms are implemented. Also, what control strategy is used in each of the models.

2.3.1 Guided Search model

The Guided Search model by Wolfe (1994, 2007) was first presented as an alternative to the two-step Feature Integration Theory (Treisman and Gelade, 1980). In contrast to the parallel feature search and serial conjunctive search, Wolfe (1994) proposed that the feature map generated through the parallel process is used to guide attention in conjunctive visual search as well.

In the Guided Search model, the information captured for different item features channels (e.g., colour, shape, orientation) are combined as a feature map. The feature map holds a topographical distribution in terms of activation for each location in the display. The attention is then directed towards the location with the highest activation. If the location consist of all target features an immediate target present response is registered. However, due to the inherent property of the visual system, the model may wrongly encode noise to the activation function and direct attention to a distractor. The search continues by moving to the next highest activation location. The search terminates if a target is found or else the activation's drop below a threshold and followed by target absent response.

The activations are described as a weighted sum map. Where through the bottom-up process the low-level features are extracted from the display first and represented in different channels. These are then weighted according to a Top-down approach where only the target features are given higher weight, and non-target features are suppressed by lower weight values. By, weighing the activation, in conjunctive search, the target then gets higher activation from both the feature channels and distractors only get activation from a single target channel. Through this mechanism, the search is shown to be efficiently guided in conjunctive search tasks. Also, the top-down weighing is used to explain the selectivity shown by people by ignoring a specific set of items.

In the Guided Search model, the *mechanism* is described as the bottom-up and top-down neuron activation for each location which also has some inherent noise due to the properties of the human visual system. The Guided Search model does not have any explicit *utility* function. The description of *ecology* is encoded in the local task environment. Finally, the *strategy* is described as a heuristic control (MAX-rule) in which attention is drawn serially to the highest activation location.

One key drawback of Guided search model is that they do not take into account the decline in acuity. The experiments for guided models are described in such a way that the stimuli in displays are large enough to be available throughout the display. The Guided models are more focused on covert attention than the overt deployment of attention and thereby ignore the eye movement behaviour.

2.3.2 Signal Detection Theory

Signal Detection Theory (SDT) based models for visual search (Verghese, 2001, Eckstein et al., 2000) are a class of search models that emphasized on only a single parallel processing stage as compared to the traditional two stage information processing architectures like *Feature Integration Theory* (Treisman and Gelade, 1980). The SDT framework are models of covert information processing, and does not account for overt visual search.

In the SDT model, the goal is to discriminate the target stimulus in a visual representation against the background/surrounding noise or distractors. Each element in the visual display is internally represented as a noisy random variable with a mean and a variance to account for uncertainty in response. SDT assumes a filter is applied across the visual display in parallel and the output of this filter is what an observer monitor. The filter here is a detector that is sensitive to the target features. On repeated presentation of the visual stimuli, a response is generated. This response is represented as a Gaussian distribution centred around the response mean. The framework states that due to the noise in the visual system there is an overlap of the Gaussian distribution with adjacent distractors. If the distractors are similar to the target, then there is an even higher overlap in the internal representation.

Regarding human physiology, the SDT framework can be understood as a neuron that is well trained through experience to fire on detecting a target stimulus. The response is then defined as a mean neuron spike count and some variability across the mean spike count affected by the nature of distractors. The *mechanism* is described here as the transfer of visual stimuli to the neurons to generate a neuron spike through the sensory system which is bounded by some information processing constraints. The architecture defined here is through the scope of filters involved in processing the information. The *utility* is defined as a trade-off between the correct detection of a signal (Hit) and the false alarm. The description of *ecology* is encoded in the local task environment. Finally, the *strategy* is described as a heuristic control (MAX-rule) in which a threshold is defined for the spike count. If the spike count exceeds the threshold, the model then responds target is present or else it is absent. In the SDT framework, the threshold is based upon the sensitivity index d', such that it maximises hits and minimises false alarm.

To summarise, the SDT framework explains visual search behaviour as an adaptation to the constraints in the visual system (noise) and the local task environment. It is also seen as a rejection of the two-stage architectures proposed by Treisman and Gelade (1980). While it has been successful in explaining some of the visual search behaviours, for example, effects due to distractor set size and similarity. However, it fails in explaining the fixation distribution in the visual field. In other words, the SDT framework is unable to capture the intermediate behaviour that emerges before the end goal is reached.

2.3.3 Bayesian Models

The Bayesian approach to model visual search tasks (Najemnik and Geisler, 2005, 2008, Butko and Movellan, 2008, Myers et al., 2013, Nunez-Varela and Wyatt, 2013, Vincent, 2015), focuses on explaining the search phenomenon using the Bayesian probabilistic framework. These models assume that the visual information from a search task is stored as a probabilistic estimate (posterior) of the state of the world. On each fixation, the estimated state is updated optimally by integrating information (Bayes rule) from the previous state and the current observation according to the acuity function of the human eye. The accuracy of the observation is determined by using a psychophysical function that replicates the decline of acuity of the human eye from the fovea. The eye movements are then made using these states and applying a heuristic decision rule (e.g., 'Maximum A Posteriori' (MAP) (Myers et al., 2013) or information-based strategy (Najemnik and Geisler, 2008)) to navigate. This rule generates a behaviour in which attention is directed to areas which have the highest probability of target present (for MAP) or to areas that maximises information gain(Najemnik and Geisler, 2005) (for information-based strategy or Ideal searcher). For example, (Najemnik and Geisler, 2005) observed that the number of fixations and spatial distribution of fixations could be better explained by a model in which each eye movement was directed to an 'ideal' location (i.e., a location that maximises information gained). Their model took into account the decline in acuity as a function of eccentricity, i.e., the accuracy of perceiving a feature degrades with eccentricity. In contrast, (Clarke et al., 2016) showed that a stochastic model with random fixation sampled from a biased distribution is also able to replicated human eye movement behaviour.

In the Bayesian modelling approach, the *mechanism* is defined as a reliability function for the observation obtained on each fixation. This reliability encodes various limitations imposed by the architecture of the human information processing system. The definition of the *mechanism* in these models are not only limited to the reliability function, but also on the memory limitation for storing these observations Butko and Movellan (2008). The description of *ecology* is provided in the model of the external environment. The *utility* here is described as "one that minimises the number of fixations to find the target while keeping the error rate below some threshold" Najemnik and Geisler (2008). Finally, the strategy is described as a heuristic control using the 'Maximum A Posteriori' (MAP) or the ideal-search strategy to guide attention.

To summarise, the Bayes optimal state estimation approach assumes that saccades are programmed to move the foveal region of the eyes to areas with a highest posterior probability of the target being present or highest information gain and are informed by a Bayesian estimate of the world. Bayesian approach are data intensive and visual search models have complemented Bayes optimal state estimate with heuristic control.

2.3.4 Cognitive Architectures

Cognitive architecture like ACT-R (Anderson et al., 1998), EPIC (Kieras and Meyer, 1997) and EMMA Salvucci (2001) are class of models that represent a set of hypotheses about human information processing system that remains constant over time and are independent of task. These hypotheses are encoded in the form of information processing architectures that a model uses to produce different behaviour, e.g., search behaviour.

The EPIC (Kieras and Meyer, 1997) cognitive architecture provides a framework for simulating human performance in a given task environment. Here, the human is modelled as an information processing system that consists of units that process information, for example, cognitive processors, perceptual and motor processor. For the architecture to work, the simulated environment needs to be programmed by the analyst. Also, the strategy space the model can use needs to be defined as production rules (if-else conditions). The input to the perceptual processor is provided in the form of symbolic information and responses are returned either as a perceptual action or using motor movements. When the model is run, the architecture generates a sequence of perceptual-motor events that are required to perform a task, within the constraints determined by the architecture and the environment. Recently, EPIC cognitive architecture has been successfully applied to the classic William's object search task Kieras and Hornof (2014).

In these models of visual search, *mechanism* is described in the form of the architecture that is being used. The architecture consists of processors that encode a theory about the information processing constraint in play while performing a task. The description of *ecology* is provided in the local task environment. These architectures do not use an explicit model of *utility*. Instead, it is implicitly defined by the analyst in the strategy that is consistent with the task goals. The *strategy* space is described in terms of production rules (if-else conditions). These heuristic controls are encoded by the modeller taking into account the task objectives.

One criticism these cognitive architecture face, which also is a drawback is that the models are restricted to the strategy that was explicitly defined (hand-coded) by the modeller. Adding constraints on the strategy space makes these models quite brittle for understanding the complexities of human behaviour.

2.3.5 Uncertainty minimisation Models

The uncertainty minimisation models (Renninger et al., 2007, Friston et al., 2012) for visual are a class of models that tries to minimise the overall uncertainty of the possible states of the stimulus in the display. These models utilise the information theory framework that relates to the concept of uncertainty reduction and information gain. Here, uncertainty is defined by using entropy, that measures the degree to which probability of various model states (target present or absent, target location, target type) is similar. In other words, when each of the model state has same probability, then the uncertainty is high (high entropy value), or else, when fewer states have high probability and rest low probability, this represents low uncertainty (low entropy value). For example, (Renninger et al., 2007) used the entropy minimisation model for target orientation identification task and showed that people make eye movements towards region that minimise local uncertainty rather than global uncertainty. Also, (Friston et al., 2012) presented a model of visual search based on the free energy principle, called the active inference framework. The model can be described as minimising expected free energy; where the free energy is a proxy for the entropy.

In the uncertainty minimisation approach, the *mechanism* is defined as a reliability function for the observation obtained on each fixation/measurement. This reliability function is a measure of uncertainty of the sampled observation. The description of *ecology* is encoded as a prior distribution. The model does not have a explicit notion of the *utility*, but the goal here is to minimise task entropy. Finally, the strategy is described as a heuristic control that minimised uncertainty/entropy or maximises information gain.

To summarise, the Uncertainty minimisation approach assumes that saccades are programmed to move the foveal region of the eyes to areas with highest information gain or area that reduce overall task uncertainty. These approaches explain *why* people behaved the way they do but do not explain *how* information is processed.

2.3.6 Control Models

The control model approach (Sprague et al., 2007, Butko and Movellan, 2008, Rao, 2010, Nunez-Varela and Wyatt, 2013, Chen and Perona, 2014, Hayhoe and Ballard, 2014, Chen, 2015) to explaining visual search are a class of models that build upon the notion that people are rational and find optimal strategies to achieve task objectives. In other words, given some task objectives, for example, finding the target as soon as possible, the goal is to maximise the overall task utility/reward. The highest reward/utility a user can get throughout the task is constrained by people's information processing system. In contrast to heuristic approaches adopted by the models described above, in the optimal control models strategy emerges as a consequence of constraints imposed by the human information processing system (Butko and Movellan, 2008), task rewards (Nunez-Varela and Wyatt, 2013) or both (Chen, 2015).

The optimal strategy can be derived using an optimisation algorithm, such as Reinforcement Learning (Sprague and Ballard, 2004, Sprague et al., 2007, Hayhoe and Ballard, 2014). Where reinforcement learning algorithm has previously been proposed as a means of explaining human learning processes (Dayan and Daw, 2008) and also, as means of deriving rational analyses of what a person should do in particular task (Chater, 2009).

In these models, the *mechanism* is described in the form of observation function that is used to extract information from the external environment. The description of *ecology* in the local task environment. The *utility* definition is provided in the reward function used in the model. For example, a negative reward is given for information foraging and positive reward for correctly finding the target to enforce speed-accuracy trade-off. The *strategy* space is not defined rather it emerges as a optimal adaptation to the *utility*, *mechanism* and *ecology*.

To summarise, the optimal control approach assumes that the saccades are programmed to move the fovea to maximise task utility/reward. One issue associated with this approach is the ability to scale up the problem to real-world tasks.

2.3.7 Machine Learning Models

Reinforcement Learning Models

Reinforcement Learning models for visual search (Sprague and Ballard, 2004, Chen, 2015) are a class of machine learning models that frames the visual search task as a control problem (Jagacinski and Flach, 2003). They are a type of Control Model (section 2.3.6) that uses reinforcement learning to find optimal strategy. These models embrace the fact that information perceived by humans are noisy and partially observable. They define the underlying *mechanism* as the amount of information captured from the environment and saved as the state. However, to solve the control problem they convert the partially observability to a fully observable problem by keeping track of the previous information.

For example, (Sprague and Ballard, 2004, Chen, 2015) uses a kalman filter to update previously seen information. In doing so, it adds a strong constraint on the model that the noise function can only be a Gaussian function and linear.

Alternatvely, a deep learning model of visual search with reinforcement learning for optimal control (Mnih et al., 2014, Leibo et al., 2018, Xu et al., 2015) are a class of machine learning models that utilise the properties of human vision to classify the content of real-world images. These models are an extension of the Control Models (section 2.3.6) described above. They are specifically used to overcome scalability issues.

In these model's *strategy* is derived by training a neural network that acts as a classifier where the output classes are the locations in the task display. The description of *ecology* in the local task environment. The *utility* definition is provided in the reward/error function used in model. The *mechanism* is described in the form of constraints in the inputs provided to the neural network.

The visual search model by Mnih et al. (2014) uses human-like foveated images as an input to the model and learns to classify what number is present in the image (MNIST dataset). They show that a neural network based model learns to make the foveated region move before making predictions. However, the model behaviour is not validated with human behaviour performance.

End-to-End Deep Learning Models

Alternatively, Li et al. (2018) presented a deep learning model of menu search that uses the data-driven approach. According to this approach, the model is first fitted to the human data such that it can replicate or predict the human search performance. In the menu search task, Li et al. (2018) first trained a deep recurrent network on a sequence of interaction obtained from actual human data to produce similar search time performance. The trained model is then used to make inferences about the interaction behaviour by exploring network parameters. The data-driven model is useful in exploring previously unseen patterns in learning behaviour by imitation. Since the model is fitted to the data, it has not been demonstrated that it explains the adaptive behaviour that people exhibit. Nor, that it would generate human-like behaviours in the absence of a large amount of interpretable data.

One common issue with deep learning models is how to interpret what has the model learned. Also, the models are sensitive to the input diet provided, that is, a correct output can only be ensured if all the inputs are correct.

2.4 Summary

To summarise, in this chapter we reviewed different approaches to modelling visual search. In each of the techniques reviewed the behaviour was explained as an adaptation to the task demands and the model architecture. However, these techniques differed in the way strategy was defined and the mechanism used to derive those strategies. In contrast, the thesis uses the computational rationality framework that focuses on the theories of the architecture used and what it implies about the resulting behaviour. A review of the computational rationality framework is presented in this chapter.

CHAPTER 3

COMPUTATIONAL FOUNDATION OF THE NEW VISUAL SEARCH MODEL

3.1 Introduction

This thesis aims to formulate a computational model for eye movement in a visual search task. In a typical visual search task, the model is required to make two control decisions, i.e., to determine where to look next and when to stop. Also, given the fact that the human vision is imperfect, the decision-making process is affected by partial observability.

In this chapter, we present the computational foundation of the CRVS model. We first give a brief overview of the model. Second, we define the Markov Decision Processes (MDP), which is a common framework to model decision making. Third, we present Partially Observable Markov Decision Processes (POMDP), as an extension to MDP. Fourth, we highlight different techniques to keep track of the information gathered across eye movements. Fifth, an overview of reinforcement learning is presented which allows the model to learn how to perform a given task based on feedback in the form of reward signals. Finally, the Asynchronous Actor-Critic algorithm is described as a scalable extension to standard reinforcement learning.

3.2 Model Overview

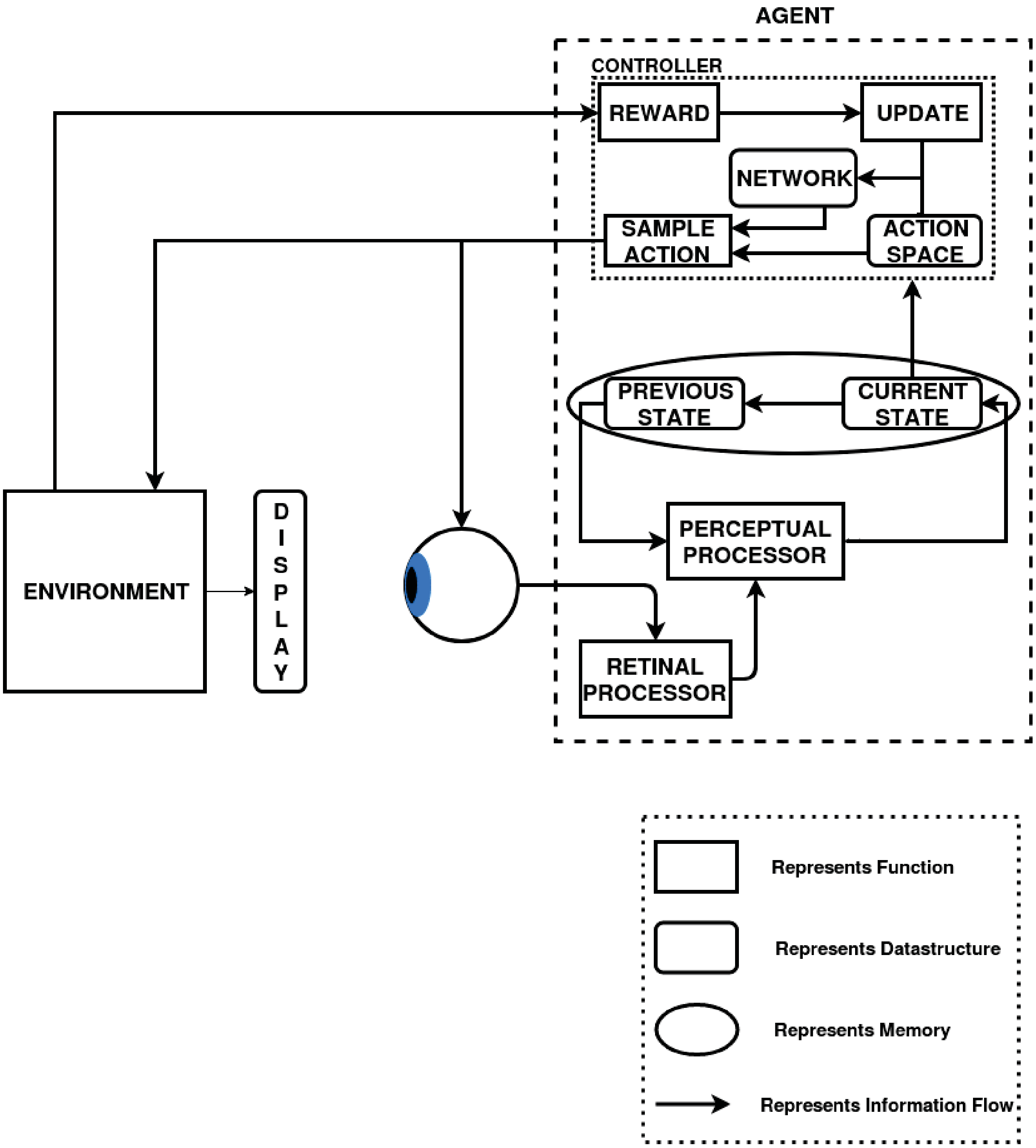
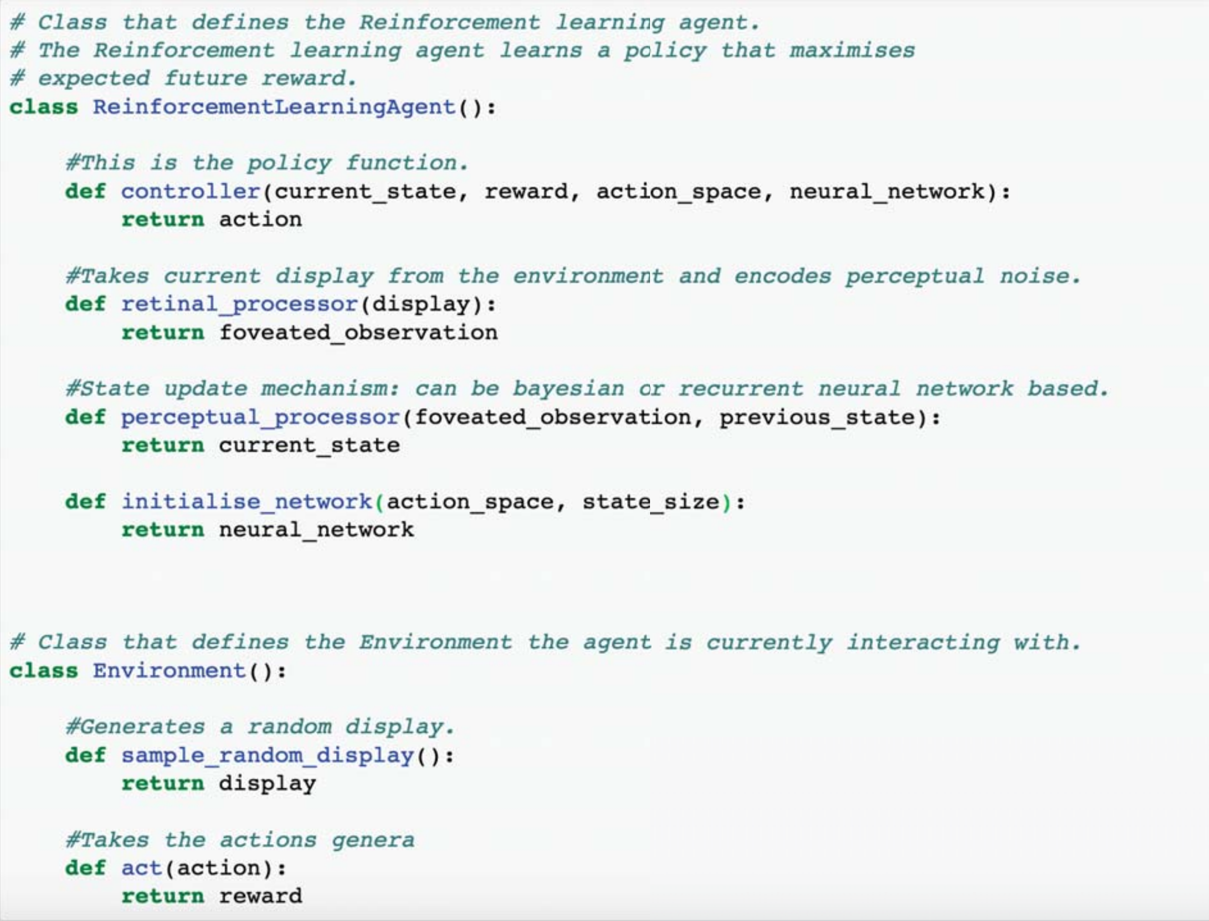


Figure 3.1: A Diagram describing the interaction between a computational agent and the external environment.



```python
# Class that defines the Reinforcement learning agent.
# The Reinforcement learning agent learns a policy that maximises
# expected future reward.
class ReinforcementLearningAgent():

    #This is the policy function.
    def controller(current_state, reward, action_space, neural_network):
        return action

    #Takes current display from the environment and encodes perceptual noise.
    def retinal_processor(display):
        return foveated_observation

    #State update mechanism: can be bayesian or recurrent neural network based.
    def perceptual_processor(foveated_observation, previous_state):
        return current_state

    def initialise_network(action_space, state_size):
        return neural_network


# Class that defines the Environment the agent is currently interacting with.
class Environment():

    #Generates a random display.
    def sample_random_display():
        return display

    #Takes the actions genera
    def act(action):
        return reward
```

Figure 3.2: High-level python code that describes different functions in the CRVS model.

We begin by giving a brief overview (as shown in Figure-3.1) of the CRVS model. The model consists of 3 processes, (1) Information Encoding process, (2) Information Updating process, and (3) Controller process.

The Information Encoding process in the CRVS model is handled by the retinal processor in figure-3.1. The retinal processor encodes the information captured by the sensor with some perceptual noise to generates a retina-like display. The perceptual noise is based on the distance (in visual angle) between the object of interest and the fovea. The human visual system is known to have an acuity that declines with eccentricity (Geisler, 2011, Findlay and Gilchrist, 2003, Kieras and Hornof, 2014). Hence, the retinal processor is described by an architecture/function that replicates the acuity function of the human visual system. In the thesis, the perceptual noise used is described in section 4.5.4.

The Information Updating process in the CRVS model is handled by the perceptual processor in figure-3.1. The perceptual processor is divided into two tasks. The first task is to extract task-relevant information, given the encoded information as an input. The second task is to update existing information with new incoming information at every time step. In the CRVS model, we describe this updated information as a state. In the model, we explore three different variations of information updating process (see section 3.6). By varying the information updating process, our intention is not to vary the theory, but, to test the scalability of each approach.

The Controller in CRVS model is responsible for the deployment of attention. Given the current state, the controller finds the best strategy for deployment of attention such that task objectives are met. The CRVS model uses a Deep Reinforcement Learning algorithm to find the approximate optimal strategy. The best strategy is informed by a reward function which is internal to the agent. The reward function is a combination of the environment task goals (e.g., rewarded on finding the target) and internal goals (e.g., minimise time spent). The learned strategy is then used to make inference about the search behaviour. An important thing to note here is that the strategy learned by the controller is an adaptation to the architecture defined in the retinal and perceptual processor and the reward function.

3.3 Agent-Environment Interaction

We assume here that an agent imitating the human behaviour has some sensors similar to the humans (e.g., human eyes), that aid in gathering information about the environment that it is interacting with. The information obtained is then processed and stored as an internal representation of the environment, which we denote in the model as a *state* (see section 3.6 for different state representations). A *state* is a set of all the parameters that are used to describe an environment. Information that can be captured in a *state* could consist of but not limited to different properties of the environment, e.g., colour, shape, size or position of an object present in the environment. Based on this *state* the model can perform some sequence of actions that may or may not change the state and the environment it is acting on. The decision that is to be made by the model is what sequence of "independent" actions should be taken based on the information obtained thus far, such that, it will lead to accomplishing of some task related goal(s). Also, every action that the model takes has an immediate and long-term effect. In other words, every action has a cost associated with it. For example, actions that lead to foraging for more information may have an immediate time cost associated with it, but, it also may result in accurate decisions later on. Hence, the model has to learn an optimal policy/strategy (what action to take and when) for some cost/reward function defined in the task. One way of guiding the performance of the model is through feedbacks, which the model receives every time it interacts with the external environment by taking action. The feedback is how the cost function is encoded in the environment and provided to the model on interaction.

The description provided above is known as *sequential decision making* (Littman, 1996). Modelling such decision-making problems is not trivial, which is further explained in sections below. Many assumptions have to be made to model these problems. In many cases, an optimal solution to these problems is not feasible for real-world problems, and hence, an approximation to the optimal solution is found instead. We begin with a simple mathematical formalisation of the sequential decision-making problem, known as the Markov Decision Process (MDP).

3.4 Markov Decision Processes

The Markov Decision Process (MDP) is a framework to mathematically formulate sequential decision-making problems. The MDP is described as a tuple $(\mathcal{S}, \mathcal{T}, \mathcal{A}, \mathcal{R})$, where,

- $\mathcal{S}$ is a set of all possible states the environment can be in.
- $\mathcal{A}$ is a set of all possible actions.

- $\mathcal{T}$ is the transition function represented as $\mathcal{T}(s, a, s')$ which denotes the probability of state s transitioning to state s' on taking action a.

- $\mathcal{R}$ is the reward function represented as $\mathcal{R}(s, a)$ is the expected reward given for taking action a in state s.

In the MDP framework, interaction with the external environment is assumed to be in discrete time steps. At any time step t, the agent perceives the current state of the environment s_t. The agent then has to decide as to what action a_t to choose that is available in state s_t. On executing the action a_t, the environment then transits to a new state s_{t+1} with a probability defined in the transition function $\mathcal{T}(s_t, a, s_{t+1})$, and a feedback in form of a reward r_t is provided to the agent. The reward r_t serves as an indicator to the agent as to how good the action a_t is when taken in state s_t.

It is important to note, that the transition to the new state s_{t+1} is only dependent on the current state s_t and action a_t and it is conditionally independent of previous states. In other words, the transition function in the MDP satisfies the *markov property*. The *markov property* states that a state is Markovian if the current state sufficiently summarises all past interactions such that the future interaction is only dependent on the current state. To define it Formally, the environment satisfies the *markov property* if the environment dynamic can be described as,

$$P(s_{t+1} = s'|s_t, a_t) = P(s_{t+1} = s'|s_t, a_t; s_{t-1}, a_{t-1}; ....; s_0, a_0)$$

The goal here for the agent is to make a sequence of independent action selection choices (until termination) such that it maximises some discounted cumulative reward over a potentially infinite horizon. To specify the goal, two functions are defined; first, the control policy, which specifies the mapping/solution of what action a_t to take when an agent is in a specific state s_t. Second, the *value function*, which contains the discounted cumulative reward as a value of a state s following the solution provided by the policy π. Equation-3.2 describes the bellman equation (Bellman, 1952) to calculate the *value* of state s.

$$V^{\pi}(s) = \mathcal{R}(s, \pi(s)) + \gamma \sum_{s \in \mathcal{S}} \mathcal{T}(s, \pi(s), s')V^{\pi}(s') \tag{3.1}$$

Where, $0 \leq \gamma < 1$ is the discount factor to emphasise the trade-off between long-term and short-term gain/reward, $V^{\pi}(s')$ is the value function for state s' and $\pi(s)$ is the deterministic policy being followed in state s. For stochastic policy,

$$\pi(a|s) = P(A_t = a|S_t = s)$$

$$V^{\pi}(s) = \sum_{a} \pi(a|s) \sum_{s \in \mathcal{S}} \mathcal{T}(s, a, s')[\mathcal{R}(s, a) + \gamma V^{\pi}(s')] \tag{3.2}$$

In general, and also in this thesis what we look for is to find the optimal solution/policy π^*. Here, the optimal solution/policy is defined by finding the maximum value for a state under all solution/policies, as defined in equation-3.3

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

$$V^*(s) = \max_{a \in \mathcal{A}} \left[\mathcal{R}(s, \pi(s)) + \gamma \sum_{s \in \mathcal{S}} \mathcal{T}(s, \pi(s), s')V^{\pi}(s')\right] \tag{3.3}$$

The MDP framework is a straightforward mathematical model that has been applied to various tasks (Littman, 1996). However, it makes one very unrealistic assumption, i.e., the agent at all times has full knowledge about the environment it is interacting with. However, in this thesis, we are trying to model the human visual system which is imperfect and is capable of only processing the environment partially. To overcome this problem, we use Partially Observable Markov Decision Process(POMDP) that has been proposed as an extension to the MDP (Littman, 1996).

3.5 Partially Observable Markov Decision Processes

As discussed in the previous section, the MDP framework assumes that the agent is always able to observe the state that the environment is currently in without any errors. In other words, the agent receives all possible information from the environment to infer the current state. This assumption is unrealistic for the class of problem being addressed here in this thesis. Since, we assume that the sensor an agent uses is constrained by the noise in the human visual system, due to which, the model is never able to observe the environment completely. For example, the noise here is based on the distance (in visual angle) between the object of interest and the fovea. The human visual system is known to have an acuity that declines with eccentricity as a consequence of the architecture of the retina (Geisler, 2011, Findlay and Gilchrist, 2003, Kieras and Hornof, 2014). A partial representation of the actual environment is available to the agent. The Partially Observable Markov Decision Process (POMDP) framework is used as a solution to model such scenarios and is, in fact, the approach that is used in this thesis.

The POMDP framework is a generalization of the MDP framework, where the agent rather than observing the current state directly, instead, receives set of observations and its corresponding probabilities (Kaelbling et al., 1998). The POMDP is described as a tuple $(\mathcal{S}, \mathcal{T}, \Omega, \mathcal{O}, \mathcal{A}, \mathcal{R})$, where, $\mathcal{S}$, $\mathcal{T}$, $\mathcal{A}$ and $\mathcal{R}$are the same as defined in Section 3.4 for MDP. Ω is the set of observations and O is the observation function represented as $O(o|s', a)$. The observation function represents the probability of making the observation o if on taking action a the environment is in state s'.

When formulating the problem as POMDP, at each time step, the agent is no longer able to perceive the true environment state. Rather, it receives an observation on interacting with the environment. Since the states are no longer observable, actions are chosen under certain uncertainty about the true underlying state of the environment. One method proposed by Littman (1996) is to maintain a probability distribution over all possible states. On receiving an observation after interacting with the environment, the agent can then update the probability distribution over all states and thereby updating its belief about the true state of the world. The probability distribution is known as *belief state* or *Information state* (Littman, 1996) and is represented as $b(s)$. In the section below we provide the details on how the belief state is updated.

3.6 Belief Representation

As described in Section 3.5, the belief state b is a probability distribution over all possible states S. Where the function $b(s)$ denotes the probability that is mapped to a state $s \in S$. In this section, we describe the process known as *state estimation* (Kaelbling et al., 1998), for updating the belief state. The *state estimation* process computes the new belief state b' based on the previous belief b, the action a an agent took, and the observation o that is received on performing that action. Below we highlight three techniques for updating the belief state.

3.6.1 Bayesian Representation

The *belief state* needs to represent the most probable state the environment is currently in taking into account agents own uncertainty from experience. The information stored should be *statistically sufficient* such that no additional information about its past interaction or observation can further inform about the current state (Kaelbling et al., 1998). In other words, rather than the observations being Markovian, the belief state now follows the Markov property.

To compute the belief state Kaelbling et al. (1998) suggested the use of *Bayes' theorem* as shown in equation 3.3.

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)} \tag{3.4}$$

where, $P(x \mid y)$ is the posterior distribution, $P(y \mid x)$ is the likelihood, $P(x)$ is the prior and $P(y)$ is the normalizing factor. The *bayes' rule* provides a probabilistic approach to updating the belief from the current observation. Here, the belief distribution is represented as posterior distribution. The equation 3.4 show the belief update using the *bayes' rule.*

$$b'(s') = P(s' \mid o, a, b)$$

$$P(s' \mid o, a, b) = \frac{P(o \mid s', a, b)P(s' \mid a, b)}{P(o \mid a, b)}$$

$$P(s' \mid o, a, b) = \frac{P(o \mid s', a,) \sum_{s \in S} P(s' \mid a, b, s)P(s \mid a, b)}{P(o \mid a, b)}$$

$$b'(s') = \frac{O(s', a, o) \sum_{s \in S} T(s, a, s')b(s)}{P(o \mid a, b)} \tag{3.5}$$

Where, $O(s', a, o)$ is the observation function, $T(s, a, s')$ is the transition function, $b(s)$ is the previous belief and $P(o \mid a, b)$ is the normalization factor.

3.6.2 Naive Bayes Representation

As described in the previous section, the *Bayes' theorem* provides a probabilistic way of recursively estimating the belief state on receiving an observation. Alternatively, the posterior can be represented as a Gaussian distribution.

The Naive Bayes or the *kalman filter* is a parametric Bayes' filter that represents the posterior distribution as a Gaussian distribution, parameterised by mean μ and variance σ^2. It assumes that the transition and observation function can be represented by a linear function (Faragher, 2012). Also, the initial belief is represented by a Gaussian distribution (Faragher, 2012).

We give here an overview of how the *kalman filter* updates the belief state. Initially, the algorithm represents the belief state b_0 as a Gaussian distribution with mean μ_0 and variance σ_0^2. On taking action a, the model receives a new observation o which is also represented as a Gaussian distribution with mean μ_1 and variance σ_1^2. The algorithm then updates the current estimate of the belief state by updating its mean μ_0 according to equation 3.6, and variance is updated according to equation 3.7.

$$k = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_1^2}$$

$$\mu' = \mu_0 + k(\mu_1 - \mu_0) \tag{3.6}$$

$$\sigma'^2 = \sigma_0^2 - k\sigma_0^2 \tag{3.7}$$

Where, μ' and σ'^2 are the updated mean and variance, and k is the *Kalman gain.* The *Kalman gain* k is a weight factor in the recursive update equations, which determines how much importance the new observation is given to estimate the current state. A higher k value weighs the new observation more as compared to the current estimate.

3.6.3 History Representation

The previous two sections (3.6.1, 3.6.2) describes two popular ways of updating the belief state and thereby estimating the current state of the environment. The third technique that is used for *state estimation* process is based on the *selective perception and hidden state* by McCallum (1996). In the *selective perception and hidden state* process it is assumed that the observations are limited by two forms of constraints. First, consist of limited sensory data due to constraints in the field of view, acuity or occlusion of objects McCallum (1996). Second, an overabundance of raw information that may or may not be relevant to the task, and limitation in computational capability for processing all this information McCallum (1996). The solution proposed was to draw attention to particular region/features and thereby assigning computational resources to process those region/features. However, this leads to the problem of generating a *hidden state*, where the agent is no longer able to determine the true underlying state of the world McCallum (1996). This is also true when information is limited by constraints in sensors.

The *hidden state* is described as a non-Markov state, since, the features of the state are partially observable and are dependent on past observations.

The non-Markovian dependencies can be overcome by introducing memory McCallum (1996). Rather than agent working with immediate observations, the belief state is represented as a sequence of past observations and actions (Equation 3.6.3).

$$b_t(s) = h(o_{t-1}, a_{t-1}; o_{t-2}, a_{t-2}; .....; o_0, a_0)$$

Here, the *history* maintained by the agent is a sufficient statistic (Heess et al., 2015). However, maintaining the entire sequence of actions and observations taken by the agent for most POMDP problems are computationally intractable. Hence, some form of approximations or summary of the past is usually learned or maintained. For example, algorithms that use a recurrent neural network to summarise interaction history is being used recently (Heess et al., 2015, Mnih et al., 2014).

3.7 Reinforcement Learning

Reinforcement Learning (RL) is a class of machine learning problem where the agent learns how to behave by interacting with the environment and receiving feedbacks in the form of numerical rewards (Sutton and Barto, 1998). Reinforcement learning is based on the MDP framework (as described in section 3.4). These class of problems differ from supervised learning in a way that there is no critic to tell the agent what the correct decision is. Instead, the agent only receives a feedback/indication on interacting (Sutton and Barto, 1998). Reinforcement learning is a technique that is defined as a learning problem rather than a learning method. Any method that solves that problem is a reinforcement learning method (Sutton and Barto, 1998). In the next subsection below, we look at one such method called Q-Learning (Sutton and Barto, 1998).

3.7.1 Q-Learning

Q-Learning is a popular model-free reinforcement learning algorithm based on based on MDP framework (Sutton and Barto, 1998). The algorithm tries to recursively find the optimal state-action value ($Q^*(s, a)$) using the following equation,

$$Q^*(s, a) = Q^*(s, a) + \alpha(r + \gamma \max_{a \in A} Q^*(s', a') - Q^*(s, a))$$

Where, $Q^*(s', a')$ is the optimal state-action value for next state s' and action a', r is the reward obtained on performing action a in state s, $0 < \alpha \leq 1$ is the learning rate and $0 < \gamma \leq 1$ is the discount factor. The learning rate α is a parameter that controls to what extent the new information overwrites the old one. When $\alpha = 0$, the agent learns no new information (exploit the prior), while $\alpha = 1$, makes the agent only consider the most recent information (ignore the prior). The discount factor γ controls the agent behaviour when the agent prefers immediate rewards than rewards that are potentially received far away in the future. By setting $\gamma = 1$, the agent tries to maximise long term reward, and with $\gamma = 0$, the agent tries to maximise immediate reward.

Q-learning defines a set of states S in an environment and a possible set of actions A in those states. At each iteration it learns the value of each of those actions for each state; this value, $Q(s, a)$, is referred to as the state-action value. So, the starting point of Q-learning is to define a Q-table that is a tabular mapping between each state and the actions, and setting all state-action values to an arbitrary value (e.g., the value 0). It then goes around and explores the state-action space. For example, a *ϵ-greedy* algorithm can be used to explore and choose actions greedily based on the highest $Q(s, a)$ value with a probability of 1-ϵ, otherwise, chooses a random action. After every action in a state is tried, an evaluation is made, i.e., what state it has led to. If the chosen action has resulted in a non-target state, the Q value is reduced for that action in that state. In doing so, the other actions will have a higher value and have a higher probability of being chosen the next time instead. Similarly, for a target state higher reward is given for taking that particular action. Importantly, when Q value is updated, it's the previous state-action combination that is being updated in the Q-table. This is also known as one-step look ahead Q-learning (Sutton and Barto, 1998).

There are two main limitations of the Q-learning algorithm,

- **Exploration vs Exploitation:** The q-learning algorithm uses a value-function to estimate what are the possible rewards for a selected action, while the policy-function decides on the best action based on the rewards. However, the chosen action is not always necessarily the best action that will lead to the maximum obtainable reward. When the policy selects the best action, it said to be exploiting the value-function; otherwise, it is exploring the environment. The problem q-learning faces is a trade-off between exploitation and exploration. If a learning algorithm (q-learning) exploits too much, it will never converge to an optimum global location which could result in a higher reward. In contrast, if a learning algorithm keeps on randomly exploring the environment, it will end up accumulating very little reward (Sutton and Barto, 1998).

- **State Space Explosion:** The Q-learning algorithm keeps a mapping of values of all the states it encountered and the actions available in a table (Q-table). Imagine the algorithm is used to learn to play an Atari game, where the state is the current game screen. Let, the image size be 84x84 pixels and is converted to a grey scale representation with 256 colours. For the Q-learning algorithm to learn a strategy is needed to maintain a Q-table with 256^{7056} rows or states. While, there may be some states that the agent may never visits, nonetheless, it would still take the Q-learner a lifetime to converge to a solution.

3.8 Deep Reinforcement Learning

As described in the previous section, Q-learning suffers from the problem of state space explosion, especially when the state space is continuous. Infact, in real world scenarios more often we encounter problems where either the state or the action space is continuous and sometimes both. A possible solution is to use a function approximation that parameterises the Q-values or the policy itself. For example, Mnih et al. (2015) introduced the *Deep Q-network* (DQN) that uses a neural network to approximates the Q-values. By using a neural network as a function approximator the model is better able to scale up to real world problems with large state space, however, it comes at a cost of no guarantees of convergence to global optimal solution. In the next subsection below, we describe a Deep Reinforcement Learning (Asynchronous Advantage Actor Critic) that parameterises the policy used in this thesis. We use the Asynchronous Advantage Actor Critic in this thesis because it is a type of policy gradient approach which has better convergence property in comparison to say the popular neural network version of the Q-Learning algorithm (DQN)Mnih et al. (2015). Furthermore, the policy gradient approach better adapts to a recurrent neural network architecture.

3.8.1 Asynchronous Advantage Actor Critic

The Asynchronous Advantage Actor-Critic (Mnih et al., 2016) is a type of policy search algorithm that solves the problem of finding the optimal policy by representing the policy space in a parametric form and then using a sampling-based technique to find the optimal policy parameters.

In policy search algorithms the goal is to find optimal parameters such that it miximises some future cumulative expected reward over a finite horizon T. Let, J(θ) be the performance function of policy π parameterised by θ.

$$\max_{\theta} J(\theta)$$

$$J(\theta) = \underset{\theta}{\overset{\pi}{\mathbb{E}}} \Big[\sum_{t=1}^{T} \mathcal{R}(s_t, a_t) \Big] \tag{3.8}$$

Where, To solve equation 3.8, *policy gradient* is the most popular approach used

(Williams, 1992, Baxter and Bartlett, 2001, Mnih et al., 2016). The idea here is to shift the parameter vector θ (e.g., θ can be the weights of a neural network) by a small amount by calculating the gradient of $\nabla J(\theta)$ such that an optimal θ is found. The gradient is defined using the likelihood-ratio theorem (Glynn, 1990). Here the expected reward in each state is the weight sum of the probability of being in the state and the reward (3.9). Due to the Markov property the current state is dependent on the previous state, hence, the $P_\theta(s_t, a_t)$ can rewritten as a product,

$$P_\theta(s_t, a_t) = P_\theta(a_0|s_0)P_\theta(s_0|a_0)...P_\theta(a_{t-1}|s_{t-1})P_\theta(s_t|a_{t-1})$$

As a consequence of multiplying probability the resulting values will be very small. Using this small value to train for example a neural network by back propagating will be very slow. To avoid this, likelihood-ratio theorem (Glynn, 1990) is used to convert the product of probabilities to a sum (3.10)

$$\mathop{\overset{\pi}{\mathbb{E}}}_{\theta}\Big[\sum_{t=1}^{T}\mathcal{R}(s_t, a_t)\Big] = \sum_{t=1}^{T}P_\theta(s_t, a_t)\mathcal{R}(s_t, a_t) \tag{3.9}$$

$$\nabla J(\theta) = \mathop{\overset{\pi}{\mathbb{E}}}_{\theta}\Big[\Big(\sum_{t=1}^{T}\mathcal{R}(s_t, a_t)\Big)\nabla\Big(\sum_{t=1}^{T}logP_\theta(s_t, a_t)\Big)\Big] \tag{3.10}$$

While equation 3.10 is an unbiased estimator of gradients for the expected cumulative reward, it still has high variance. In other words, the optimiser may take many steps in the poor direction, even though on average it will end up in the correct direction. This may lead to weak or slow convergence. Alternatively, Williams (1992) showed that any constant baseline value that is independent of the action can be subtracted from the gradient to minimise variance. Introduction of this baseline still makes the gradient update unbiased (see (Williams, 1992) for derivation).

$$\nabla J(\theta) = \mathop{\overset{\pi}{\mathbb{E}}}_{\theta}\Big[\nabla\sum_{t=1}^{T}logP_\theta(s, a)\Big(\sum_{k=t}^{T}\mathcal{R}(s_k, a_k) - b(s_t)\Big)\Big] \tag{3.11}$$

The Actor-Critic algorithm is a class of policy search method that combines the policy gradient approach with value iteration approach (Sutton et al., 2000). Here, the algorithm uses the state value function $V(s)$ as a baseline. In the Actor-Critic algorithm, the Critic is responsible for calculating the value function $V(s)$ following policy π as described in section 3.4. The Actor then uses the value function estimated by the critic and using equation 3.11 updates the policy parameters. The actor-critic variant of policy gradient approach helps in reducing variance as compared to the standard policy-gradient. Alternatively, Schulman et al. (2015) explored various other unbiased estimators. In this thesis we use the low variance Generalised Advantage Estimate (GAE) variant of actor-critic which is shown to better reduce variance than other estimators (Schulman et al., 2015) and also used by (Mnih et al., 2016) (see appendix A for details).

In the Asynchronous Advantage Actor-Critic algorithm (Mnih et al., 2016), a neural network is used to maintain the stochastic policy π and the value function V as a function of the state. The policy is defined as a probability distribution $\pi(a|s;\theta)$ over the action space which the agent can choose from in the state s and use this distribution to sample an action. The value function is represented as the expected cumulative reward $V_\pi(s;\theta)$ when starting in state s and following policy π. The parameter θ represents the weights of the neural network and is used to estimate these values. The parameters or weights of the neural network (θ) are then tuned by optimising a loss function, here, the loss function is composed of three sub losses (equation 3.12).

$$\begin{aligned} total_{loss}(\theta) = policy_{loss}(\theta) + value_{loss}(\theta) \\ + entropy_{loss}(\theta) \end{aligned} \tag{3.12}$$

In the A3C algorithm, the agent interacts with its environment to generate an entire roll-out of a single trial. The network weights are then updated using this entire roll-out. In the roll-out, each step is represented as a tuple $< s_0, a_0, r_0, s_1, v_{s_0} > \rightarrow < s_1, a_1, r_1, s_2, v_{s_1} > \rightarrow ... \rightarrow < s_{n-1}, a_{n-1}, r_{n-1}, s_n, v_{s_{n-1}} >$.

$$\begin{aligned} value_{loss}(\theta) &= (V_{target}(s_i) - V(s_i;\theta))^2, i = 0, ..., n-1 \\ where, V_{target}(s_i) &= \sum_{t=0}^{n-i-1} \gamma^t r_{t+i} \end{aligned} \tag{3.13}$$

The $value_{loss}(\theta)$ is calculated using the roll-out to estimate the target value for each state using the actual rewards obtained (equation 3.13). The value of the terminal state is set to 0. Each roll-out generates a set of n samples/steps to train the neural network on the value loss function using the estimated values.

$$\begin{aligned} policy_{loss}(\theta) &= -log(p(a_i|s_i;\theta)) * A(a_i, s_i;\theta), \\ &i = 0, ..., n-1 \\ where, A(a_t, s_t;\theta) &= \sum_{t=0}^{n-i-1} \gamma^t r_{t+i} + \gamma^{n-i} V(s_n;\theta) \\ &-V(s_i;\theta) \end{aligned} \tag{3.14}$$

The $policy_{loss}(\theta)$ loss function (equation 3.14), tunes the network parameters in order to shift the policy such that the sampled actions improves the advantage value $A(a_i, s_i; \theta)$ given the current state s_i. Here, the advantage value is defined as an additional feedback signal to the agent on taking action a_i in the state s_i over the average value of $V(s_i;\theta)$ as a baseline.

Defining the policy in such a way may bias the agent towards some actions which leads to the agent not exploring other actions in a state and converge to suboptimal policy. To prevent this, Mnih et al. (2016) suggested adding an entropy loss to improved exploration by discouraging premature convergence. The $entropy_{loss}(\theta)$ loss function, maximises the entropy of probability distribution over action space for a given state (equation 3.15).

$$\begin{aligned} entropy_{loss}(\theta) &= \sum_{a \in A} -p(a|s_i;\theta) log(p(a|s_i;\theta)), \\ &i = 0, ..., n-1 \end{aligned} \tag{3.15}$$

Also, the architecture of the A3C algorithm comprises of a global network and multiple asynchronous worker agents that interact with their copy of the environment in parallel. Each worker agent uses its local copy of the gradients using the loss function to update the global network parameters and then copies the global network parameter for training on the next trail. This is done so to avoid any correlation between the training data since each worker will explore and generate a different state space and thereby reduce any correlation. Also, using multiple agents results in faster convergence (Mnih et al., 2016).

The derivation still holds when the states are non-markovian, i.e., rather than using states model uses the entire history (see (Wierstra et al., 2010) for proof).

CHAPTER 4

IMAGE SEARCH TASK

4.1 Introduction

In this chapter, we apply the *CRVS* model to our first visual search experimental task, i.e., the image search task. The goal of the image search task is to simulate a real-world visual search. The search task required the participant to search for an image with as many target feature he/she can find and as quickly as possible. Here, there was no right or wrong target every image in the display can be a target image. The control problem that the model needs to solve here is when to terminate the search.

Previous modelling approaches have utilised heuristic thresholds (Treisman and Gelade, 1980, Wolfe, 1994, 2012, Ehinger and Wolfe, 2016) to explain the stopping behaviour during visual search. In contrast, we show that an approximately optimal control model can explain the emergent behaviour without any description of the heuristic rule. Instead, it explains the behaviour as an adaption to constraints in the human visual system, ecology and the reward. Here, we demonstrate that the stopping rule emerges as an adaptation to the reward/cost received for choosing an image and the distribution of the target features.

The *CRVS* model presented in this chapter represents the environment as a symbolic vector. Where each element in the vector represents a collection of target features (e.g., colour, shape and size). The actions than the model can take consist of 36 fixations action and a select action that selects the fixated image and terminates the trail. In this chapter, we compare two state update technique, i.e., the recurrent neural network model that maintains a summarised history of observations and a full Bayesian update that uses Bayes theorem. Our intention here is to evaluate the scalability of the two approaches. The Bayesian update acts as a benchmark which is a standard method used to maintain the state update to solve the POMDP (Littman, 1996).

4.2 The Task

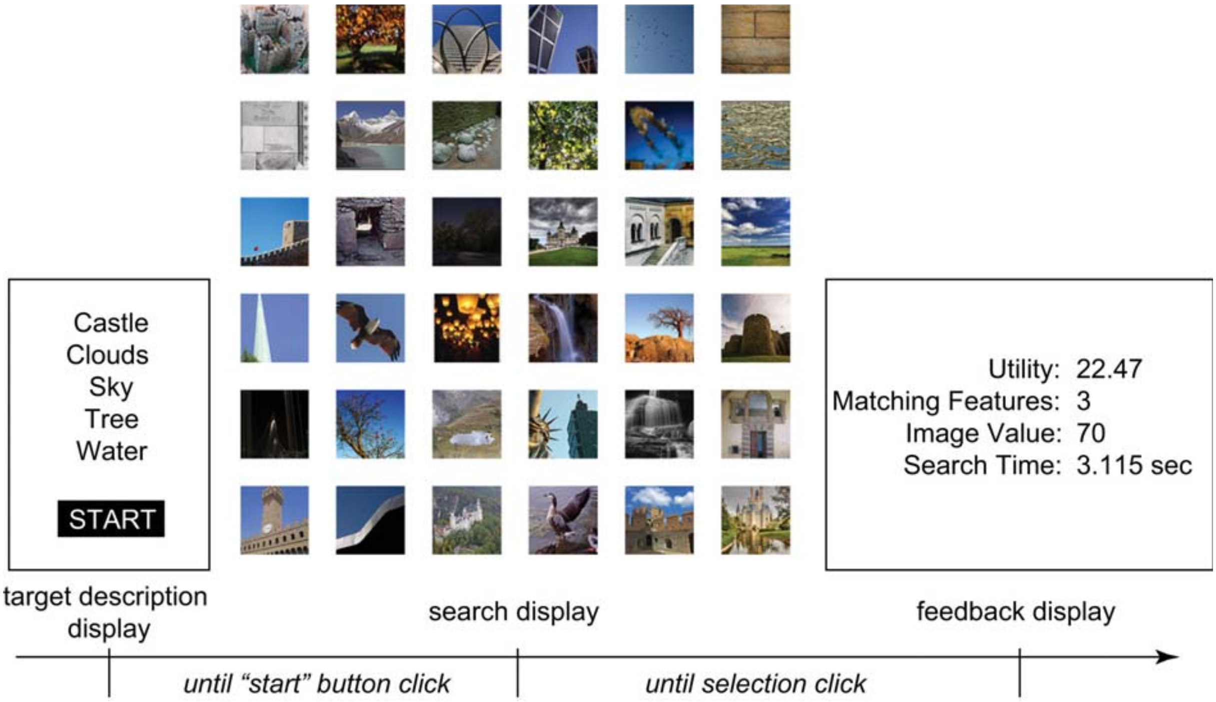


Figure 4.1: The figure illustrates a random display used in one of the search trials. The search task required the participant to search for an image with as many target feature he/she can find and as quickly as possible. In this display, the target features were Castle, Clouds, Sky, Tree, and Water. There was 1 image with 5 target features, 1 image with 4 target features, 2 images with 3 target feature, 6 images with 2 target features and 26 images with 1 target feature (Tseng and Howes, 2015).

The image search task as conducted by Tseng and Howes (2015) required participants to search for an image, in a display of 36 images, that best matches some set of target features. The target features consisted of high-level feature description like, for example [Sky, Castle, Bird, Tree, Lake] (See Figure 4.1). As in a typical visual search experiment, there are both matching and distractor features. However, this experiment differed from traditional visual search experiments by not specifying a *correct* target image. Instead, participants were free to choose whichever image they wanted, and there was no correct or wrong image. Participants were given points for selecting an image and were instructed to maximise the feedback points. The points in the experiment were a function of search time and image value, where value was a function of the number of features that matched the goal (Tseng and Howes, 2015), So, an image that matched all five features of [Sky, Castle, Bird, Tree, Lake] would receive more points than an image that matched only two of these features.

At the beginning of the trial, participants were made aware of the value of each n-feature image (where n is the number of target features in the image). Also, a high-level description of the target features that they need to search for was provided. Once the participants familiarised themselves with the target features (participants were not timed at this stage and were also made aware of it), they then started the search task, at this point they were timed. The trail terminated once they click on the selected image with the mouse. This was followed by a feedback screen of the points they scored. While the experiment doesn't measure any behavioural data about the mouse movement, however, it does add to the total search time since the trail and the timer is terminated only after an image is clicked.

In the experiment two image value functions were tested, i.e., the value of an image with n target features increased as a power function in one condition and a linear function in the other. These reward values represent the user preference and show how much people would value finding all matching features or if even finding four will do features. The density of images in the display was also manipulated: low density (edge-to-edge item spacing of 0.85 degree) and high density (edge-to-edge item spacing of 0.085 degree).

4.2.1 Observed Data

Tseng and Howes (2015) observed that people adapted their search strategy to the feed-

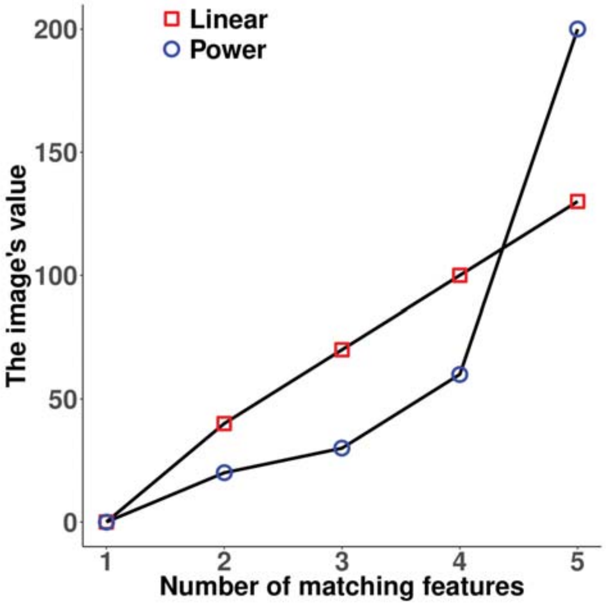


Figure 4.2: The figure illustrates the two utilities used in the experiment. (Tseng and Howes, 2015).

back reward. They reported that the images selected with the number of target feature were higher when the value of the images followed a power function distribution in contrast to the linear function distribution.

Furthermore, the reward function also affected the number of fixations people took (Tseng and Howes, 2015). Participants fixated on more images when the value of images followed a power function in comparison to the linear function. These results are consistent with the previous observations since participants found higher target images in power function scenario which had a higher return. Also, the number of fixations in the high-density display were lower than the low-density displays.

The ecological distribution of real-world images returned from a search engine that matched the search criteria was reported in (Tseng and Howes, 2015). In the study, the goal was to identify the relationship between the number of keywords in the search criteria and the number of images returned. Here, Google image search engine was used to infer the distribution. In the study, a list of keywords like 'water' and 'boat' and 'sky' keywords was entered, and the number of returned images were noted. Tseng and Howes (2015) reported that the resulting distribution followed a power distribution. In other words, as the number of keywords increased the number of target images returned decreased, and this declined followed a power-law.

4.3 Background

Search termination is a significant problem that has frequently been overlooked (Wolfe, 2012). Especially in scenarios which involve high risk, for example, searching for food, cancer symptom, a bomb in luggage, etc. Where early or delayed search termination may incur a high cost. Previous models have explained search termination by describing stopping thresholds (Ehinger and Wolfe, 2016, Treisman and Gelade, 1980, Wolfe, 1994, 2012).

Wolfe (2012) describes search termination in terms of the number of items searched or time spent in searching. The assumption is that people keep track of a noisy estimate of either of these two sources of information which is then used to formulate a decision threshold to terminate the search. The thresholds can be conservative or liberal (terminate early or later), which is decided by the number of goal-relevant items in the display, crowding and cluttering of items in the display or the value of target item (Wolfe, 2012).

Ehinger and Wolfe (2016) extended the idea by presenting a potential value version of optimal foraging model for search termination. The task was to find some target objects (gas stations) in a satellite image. The number of gas stations in an image varied for each trial. Ehinger and Wolfe (2016) showed that the search termination could be explained by a model that stops searching when the expected rate of targets found in an image is lower than the cumulative average of targets found thus far. Also, they showed that to estimate the expected rate, people combine three sources of information. They are the prior belief about the distribution of targets in images, time spent in search and the search history.

In contrast, we present a novel model of visual search that explains search termination as an optimal[1] adaptation to the task ecology, constraints involved in processing incoming

[1] we use the word optimal here for brevity. The solution found is an approximation to the optimal solution, which is asymptotically optimal.

information from sensors and task reward. Here in the model, we avoid any heuristic thresholds; instead, the termination strategy emerges as an optimal adaption to the theory defined for the ecology, reward and information processing constraints.

In the model, we utilise the *active vision* approach (Findlay and Gilchrist, 2003). According to this approach, people make a series of eye movements to gather task-relevant information. The eye movements are necessary because the high acuity (fovea has the highest density of cone density) foveated vision only covers 1-2 degrees of visual angle. With the increase in eccentricity, the acuity drops sharply (the density of cones drop off beyond 2 degrees of visual angle), and hence the vision becomes noisy and uncertain. This this further evident from Nelson and Loftus (1980) experiments on item recognition. In their experiments, they showed that the ability to correctly identify/ recognise item declined as a function of the distance between the fixated location and the object location. Also, the peripheral vision that covers a larger area, though information obtained here introduces some uncertainty, it is still useful for guiding eye movements and recognising objects (Loftus and Mackworth, 1978, Nelson and Loftus, 1980, Geisler, 2011).

Furthermore, evidence from previous literature shows strategies are adapted to the task ecology (Bertera and Rayner, 2000, Halverson and Hornof, 2004, Vlaskamp et al., 2005, Tseng and Howes, 2008, Payne and Howes, 2013, Liu et al., 2017). For example, Vlaskamp et al. (2005) found that the distance between items in a display has a direct impact on the search performance, i.e., the amount of time fixated on an item, number of fixations and dwelled time. They showed that by controlling the distance between items (range between 1.5 to 7.1 visual angle), the number of fixations and fixation duration increases with an increase in distance. Similarly, Everett and Byrne (2004) showed that small changes in spacing between icons in a display changed the search strategy which was reflected in the reaction time for finding the target icon.

Also, Chen et al. (2015) presented a cognitive model for a menu search task. In their experiment, they found that people adapted their search strategy to the ordering and grouping design of menus. Specifically, they showed that the semantically grouped menus with alphabetic ordering required the least amount of search time as compared to unorganised menus. They also controlled the menu group size (3 menu item in groups of 3 and 5 menu items in groups of 2) and showed that a semantically grouped menus with semantic ordering (highly related items are close together within the group) had a significant effect on search time, where, users skipped menu groups which were not semantically related to the target menu.

In addition, search strategies have also been shown as an adaptation to task reward (Hikosaka et al., 2000, Glimcher, 2003, Najemnik and Geisler, 2005, Della Libera and Chelazzi, 2009, Stritzke et al., 2009, Eckstein et al., 2010, Navalpakkam et al., 2010, Tseng and Howes, 2015)(see (Eckstein, 2011) for review). For example, Stritzke et al. (2009), showed that people are risk evasive, such that they direct their gaze to the more rewarding region and stay away from the region that incurs a cost. Also, Eckstein et al. (2010) showed that people also find strategies that maximise total accumulated rewards in scenarios where rewards are distributed differently across locations in a display.

In the sections below, we first define our theory for what mechanism are involved to explain the behaviour shown by people during the image search. This is followed by the description of the control model and the results of the model performance.

4.4 Theory

In this chapter, we present a control model with the theoretical assumption that the eye movement strategy emerges as a consequence of adaptation to the statistics of the environment (distance between items in a display and the target feature distribution in the display), the constraints imposed by the mechanisms of human visual system (active vision) and the task reward. For the search task, the decision-making problem can be formulated as a POMDP and solving this problem with a deep reinforcement learning algorithm to find the optimal strategy.

In the model, we assume that people's decision to stop searching is informed by the feedback in the form of rewards that are received for making a decision. Better decisions will receive higher rewards, which in turn will reinforce better eye movement strategy. Hence, our theory here is,

- When images are displayed with a smaller gap between them (high density), more information will be available in the parafoveal region. We predict that the number of fixations will reduce since more information is available. We also predict due to the availability of more information people will switch to a low-cost strategy of skipping items.

4.5 CRVS model with Recurrent Update

In the following section, we provide a detailed overview of each of the individual component used in the CRVS model with recurrent architecture.

4.5.1 External Display

On each trial, 36 images were randomly positioned in a 6 x 6 grid-like display. In the display, each image consisted of some target (represented by a scalar value 1) and non-target (represented by scalar value 0) objects. Some image has more target relevant features, and some had fewer. The distribution of target features in the display followed a power distribution. In other words, a display with a set of 36 images, there was 1 image with 5 target features, 1 image with 4 target features, 2 images with 3 target features, 6 images with 2 target features and 26 images with 1 target feature. This formulation follows the ecological study of target feature distribution done in Tseng and Howes (2015). The edge to edge gap between objects was kept at 0.85 degree for low density condition and 0.085 degree for high density condition.

Also, non-target objects were added to each image randomly by sampling the number of objects to be added between 1 to 4. It was ensured that the total number of objects (target and non-target) within an image was less than or equal to 5. Here we assume that the search engine algorithm is not efficient enough to filter images with non-target features.

4.5.2 Action Space

In the model, the action space consists of (1) fixate on an image location and (2) select the fixated image. In our study, there was a grid of 6x6 images, and there were, therefore, a total of 36 possible fixation actions. A trial was terminated by choice of the select fixated image action.

4.5.3 Reward

In the model, the value function of an image was a function of the number of target features that it contained, and there were two value functions,

- A power law condition that had image value of 200, 60, 30, 20 and 0 for images with target feature 5, 4, 3, 2 and 1 respectively.
- A linear law condition that had image value of 130, 100, 70, 40, 0 for images with target feature 5, 4, 3, 2 and 1 respectively.

These image values have been exactly adapted from Tseng and Howes (2015). The total search time used in the mode is described in equation 5.1 (Baloh et al., 1975).

$$SearchTime = SaccadicDuration + FixationDuration \tag{4.1}$$

$$SaccadicDuration = 37 + 2.7 * Amplitude \tag{4.2}$$

Where the *Amplitude* is the distance between the current and previous fixated location, *SaccadicDuration* (in milliseconds) is the time required to move fovea from previous fixated location to the current fixated location. $FixationDuration$ (in milliseconds) is the dwell time on each image, which is kept at a constant time in the model and the value used is taken from the experimental results of Tseng and Howes (2015) (figure 4(b)). The reward function was therefore defined as $(imagevalue/10)$ - $searchtime$ $(inseconds)$. Here, we reduce the image value to avoid gradient explosion (Pascanu et al., 2013) during training. By adding a time cost to the reward function, a speed-accuracy trade-off is imposed in the model. Searching longer for an image may result in higher reward but at a cost.

4.5.4 Observation Model

The human eyes ability to discriminate and perceive object features degrades with eccentricity (Strasburger et al., 2011). The rate of decline is different for colour, shape and size features (Kieras and Hornof, 2014). Here in the model, we describe the acuity function as a probability function that specifies whether the object is correctly perceived or not. A quadratic psychophysical function was used to model acuity drop as described in Kieras and Hornof (2014). The function (equation 4.3) determines the feature availability as a function of eccentricity and feature size.

$$P(detection) = P(size + X > threshold) \tag{4.3}$$

$$threshold = ae^2 + be + c$$

$$X = N(size, size * v)$$

Where, size is the size of each feature in the image (in the model we assume all objects occupy same spatial span) and represented in terms of visual angle, e is the eccentricity in terms of visual angle. In the model, the parameter values a, b, c and v were set to $0.2, 0.1, 0.1, 0.7$ respectively as described in (Kieras and Hornof, 2014).

4.5.5 State Estimation

At each time step t on which fixation is made the model receives a noisy observation for each location. The observation space consists of the number of target-features present in each location. Also, the current fixation location is also provided to the model in the form of one hot encoded vector (the fixated location is represented by value 1 and rest all cells are 0). The fixated location information is important here because when the search terminates, the fixated image is the selected image in the model. Hence, the state in the model is represented as the combined vector of observation and fixated location. The model then on each fixation updated the state using a recurrent neural network (Hausknecht and Stone, 2015). The network receives partial observation on each fixation, and by using a recurrent neural network, it maintains the summarised history of previous observations (see section 3.6.3).

4.5.6 Model Learning

Model Architecture

The model architecture as shown in Figure 4.3 provides a brief overview. The noisy observation (36 element vector) and the one hot vector (36 element vector) (see section 4.5.5) was taken as the input. The observation consists of a vector of size 36 with each cell representing the content of each image described as the number of target objects perceived with values ranging from 0-5 and another vector that informs the current fixated location a vector of size 36 with value 1 for current fixation index and rest are 0s. Images consist of a set of objects, and each object occupies equal size in an image. We add this constraint in the model as we do not have the ecological distribution of the sizes of the object used in the experiment. However, adding distractor objects to images at random will rescale the size of target objects in the image. This input was first combined to form a 72-element vector and then connected to a recurrent layer consisting of 72 hidden nodes, with tanh activation function. This is followed by a fully connected hidden layer consisting

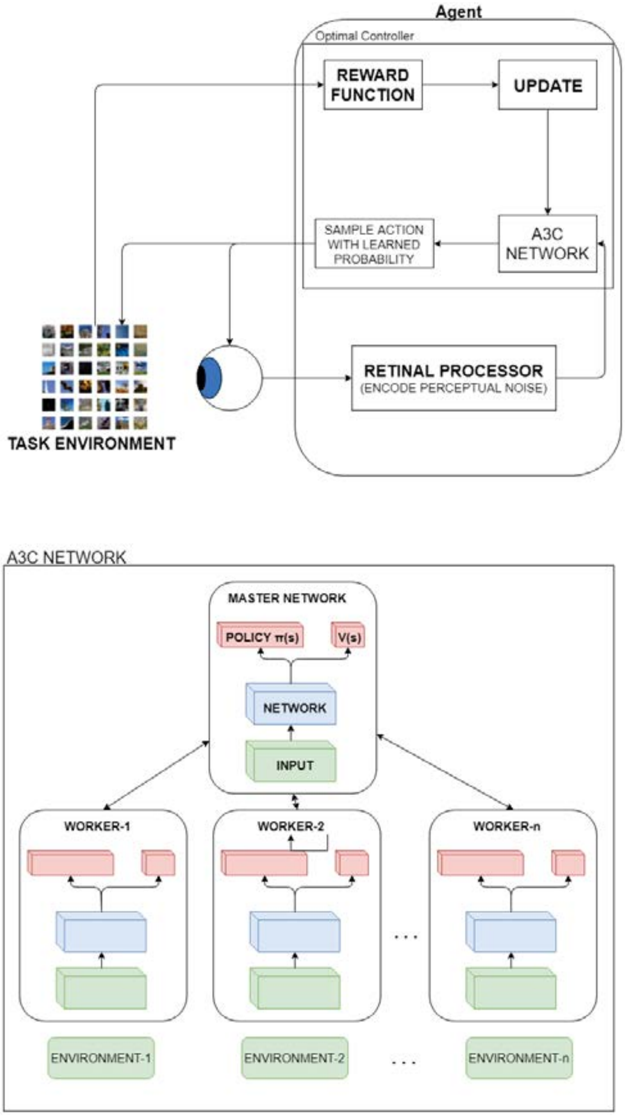


Figure 4.3: An Overview diagram of the CRVS model interacting with the image search task environment.

of 37 nodes with a single output for each action in the task, with a softmax activation function. The output of the recurrent neural network is also connected to a single node fully connected hidden layer. The two output layer provides the probability distribution over action space and the value of each state.

Policy Learning

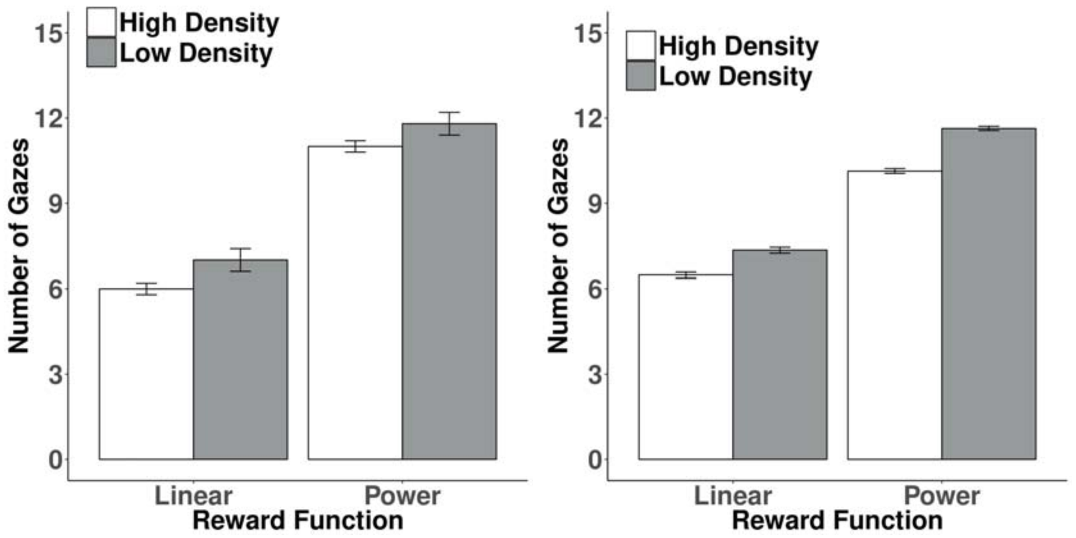
As described above, at each point in time, the model observes the external display through a noisy percept (section 6.5.4). The probability to correctly perceive feature is highest in the foveal region and declines as a function of eccentricity in the parafoveal region. The model then uses the noisy information from the foveal and parafoveal region to guide attention by taking actions (e.g., choose where to move the fovea). Since the environment is only partially observed the model needs to integrate information over time in order to determine how to act and how to make eye movements most effectively. It does this using the Recurrent network as state estimator described above. At each step, on taking action the model receives a scalar reward (section 6.5.3). Here, the goal of the agent is to learn policies/strategies that maximises the total sum of such rewards $R = E[\sum \gamma^{t-1} r^t]$ where $\gamma \in (0, 1)$ is the discount factor.

Here, the model learns the optimal control policy by using one type of reinforcement learning algorithm, i.e., the policy gradient algorithm (for detail refer to section 3.8.1).

4.5.7 Results

In the section below, the search performance of the deep reinforcement learning model is reported and compared to the human performance for image search task. The results reported here are from a model that learns the approximately optimal policy (reward value per trial asymptoted) through training and was used to generate the behaviour (last 50,000 trials of the simulation).

The plot of the reward function versus the number of fixations is shown in Figure 4.8. The result shows that the model required fewer fixations in the linear function as compared to the power function. Also, when the image density was high, the model required fewer fixation as compared to a low-density condition. This behaviour is consistent with human performance with the goodness of fit for the model being $R^2 = 0.95$ and the Mean squared error value being $RMSE = 0.57$. Here, the model can replicate the human performance for the difference in the number of fixation behaviour between linear and power reward function. However, it takes fewer fixations for power function in comparison to human performance. We assume this behaviour is a consequence of memory decay phenomenon in the model. The LSTM network is known for capturing long-term dependencies and



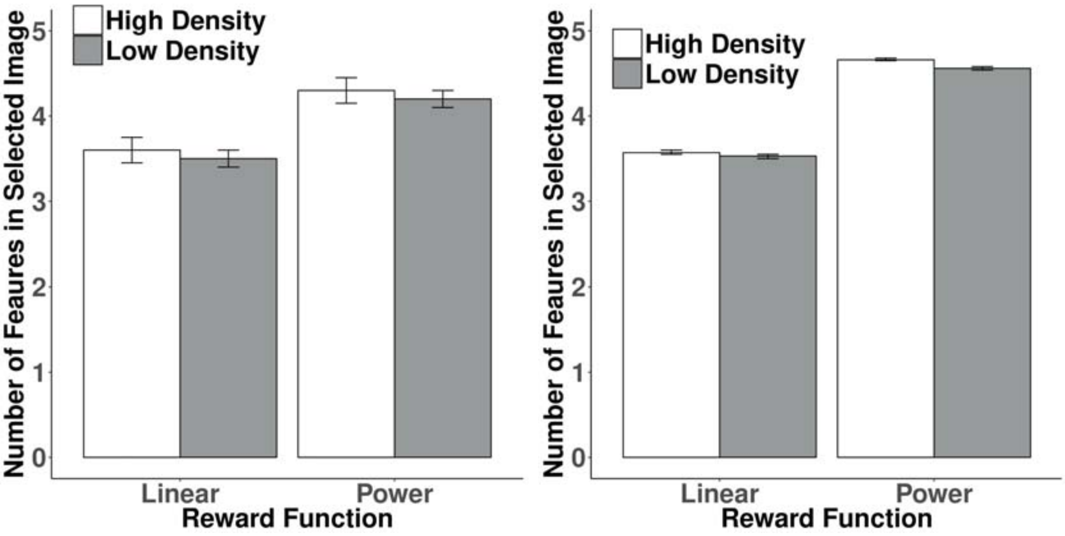
(a) Human (b) CRVS

Figure 4.4: Mean search performance as function of number of fixations for (a) Human and (b) CRVS model. Error bars represent standard error

may differ to human memory decay.

The plot of reward function versus target features selected in an image is shown in Figure 4.8. The deep reinforcement learning model found the higher number of matching target features in an image when the reward function used power-law as compared to the linear law. This behaviour is also replicated in the human performance with our model achieving goodness of fit value of $R^2 = 0.94$ and the Mean squared error value being $RMSE = 0.27$. In comparison to human performance, the model was able to achieve higher mean target features for both linear and power reward function. We assume this behaviour as a consequence of our choice of the noise model. The visual noise model parameters are not fitted to the task and hence the difference in performance.

The model behaviour shown above is consistent with the search time effect and the total reward gathered by the model. Here the model took a long time to select an image in the power law function as compared to the linear law. This was because the payoff was better with higher image value in the power function condition. The goodness of fit with human data was $R^2 = 0.93$ and the Mean squared error value being $RMSE = 1022.83$. Here, the high Mean squared error is because the model did not account for the mouse



(a) Human

(b) CRVS

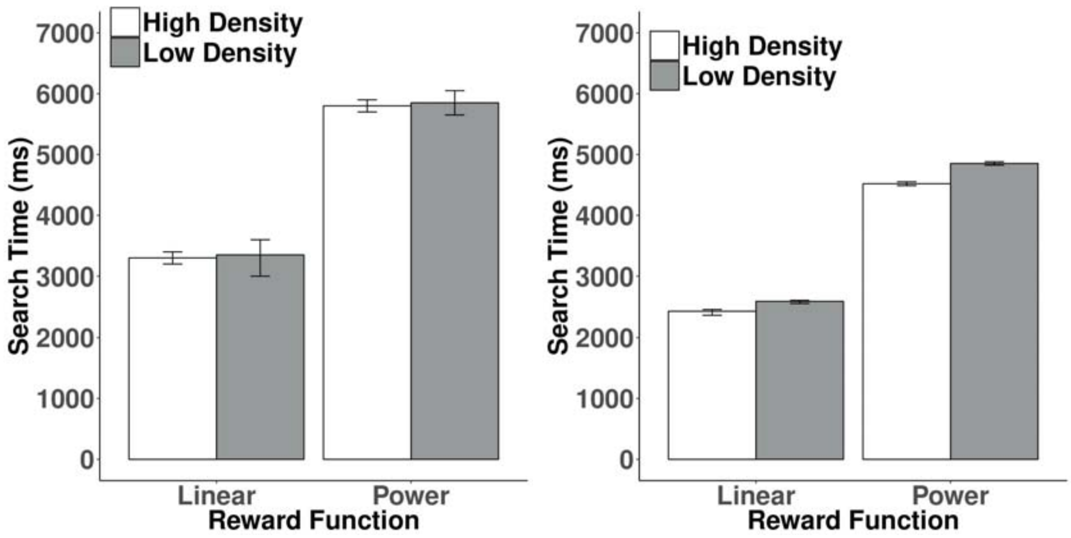
Figure 4.5: Mean search performance as function of number of target features selected in an image for (a) Human and (b) CRVS model. Error bars represent standard error

movement time. Due to the exclusion of mouse control in the model, the actual human data shows that peoples' search time was higher than the model. This is replicated in the reward plot as well, where, the model reported higher reward achieved in comparison to actual human data.

4.6 CRVS model with Bayesian Update

4.6.1 External Display

The External display is same as defined in section 4.6.1. The only difference is we scale down the display size to a 3x3 grid-like display. The distribution of target features in the display was also scaled down with 1 image with 4 target features, 1 image with 3 target features, 2 images with 2 target features and 5 images with 1 target feature. We also, double the gap between the images to reduce the availability of information.



(a) Human

(b) CRVS

Figure 4.6: Mean search performance as function of search time for (a) Human and (b) CRVS model. Error bars represent standard error

4.6.2 Action Space

The action space is the same as defined in section 4.5.4. The only difference is the number of fixation actions is reduced to match the 3x3 display size with 9 possible fixation action.

4.6.3 Reward

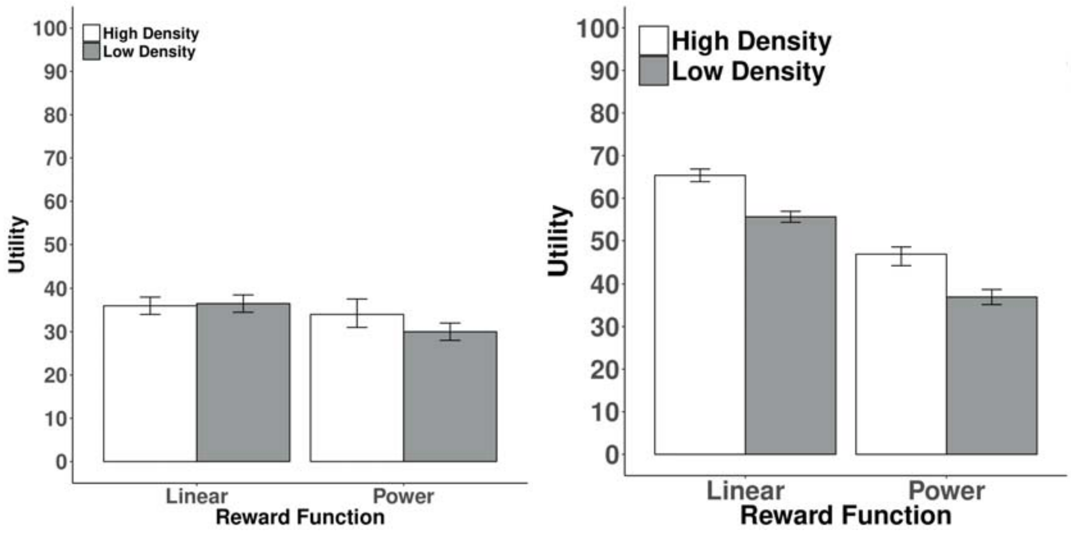
The reward function is same as defined in section 4.6.3. The only difference is the value of 4-feature image was increased for power function to 100.

4.6.4 Observation Model

The Observation model used is same as defined in section 4.5.4.

4.6.5 State Estimation

A state here in the model is represented by a hypothesis which states that the given location the hypothesis represents is the best image (best image is described as an image with highest number of matching features) and the number of fixations. For example, H_0



(a) Human

(b) CRVS

Figure 4.7: Mean search performance as function of reward (image value/search time (in seconds)) for (a) Human and (b) CRVS model. Error bars represent standard error

represents image at location 0 is the best image. In the model there were 9 hypotheses used, with each hypothesis representing the image at their respective location is the best image and are mutually exclusive. A sample state space will look like:

H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	$Fixations$
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	3

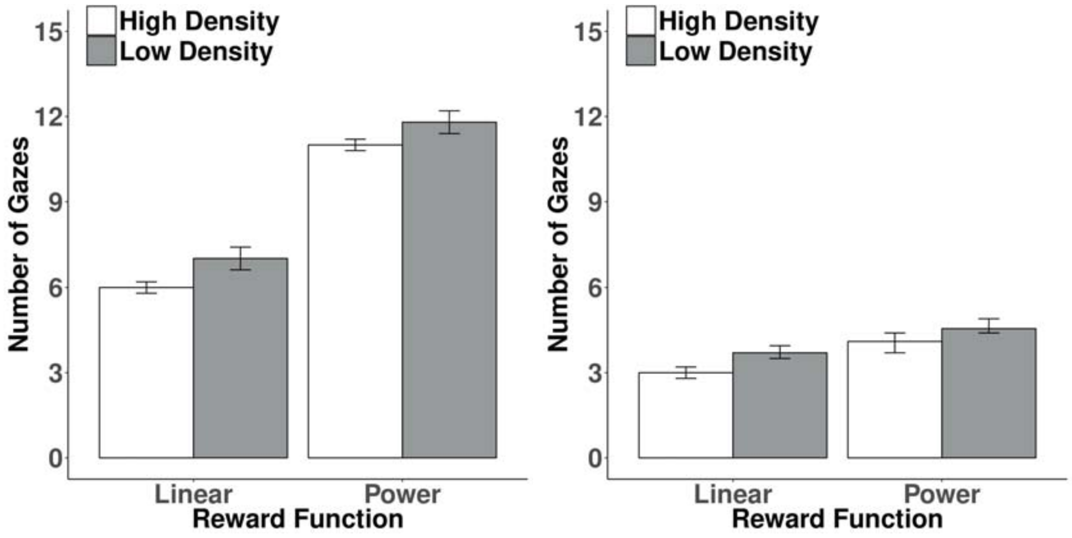
At each time step, the model is in a state which is guided by observations that are partial. The model maintains a belief distribution $b(s)$ over all states given the observations thus far. On taking action a the model observed an observation o and updated its belief according to the Bayes rule described in section 3.6.1.

At the beginning of the trial, the model assumes a uniform distribution over all states. In other words, initially, the model assumes the images at every location is equally likely to be the best image. On taking an action the model transits from state s to state s' with a probability $P(s'|s)$. In the model, the transition is assumed to be with probability 1, since the display remains static within the trail and we assume eye movements do not encode any error. The models also maintain a likelihood table which is a frequency table to calculate $P(o|s)$. Each row in the table is a mapping of an observation to the frequency of its occurrence. The observation in the table consists of the true hypothesis number, the number of features in best images, fixated location number, the observation of highest feature image seen and its location. Also, we made the hypothesis discrete by rounding the values to 5 levels [0.0, 0.25, 0.5, 0.75, 1.0]. This done so for the model learning to be scalable.

4.6.6 Model Learning

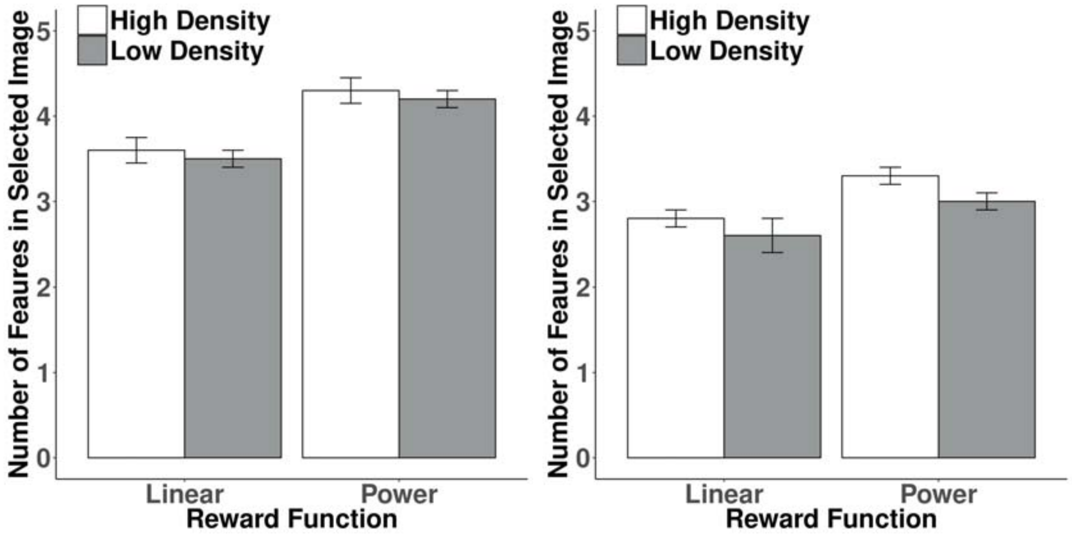
The model here tries to learn when to decide to click an image assuming this is the best image in the environment or makes an eye movement in search of better images. This learning process can be emulated with control knowledge. The control knowledge is represented as a mapping between the beliefs and actions, which is learnt with a reinforcement learning algorithm, Q-learning (see section 3.7.1). At the beginning, the values in the Q-table (i.e., Q-values) of all belief-action pairs were set to zero. The model, therefore, started with no control knowledge and action selection was entirely random. The model was then trained until performance plateaued (requiring 100,000 trials). The model 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 all the actions randomly with probability epsilon. Epsilon was set to 0.1 in our model. The idea is that these Q-values are learned (or estimated) by a simulated experience of the interaction tasks. The true Q-values are estimated by the sampled points encountered during the simulations. The optimal policy acquired through this training was then used to generate the predictions.

4.6.7 Results



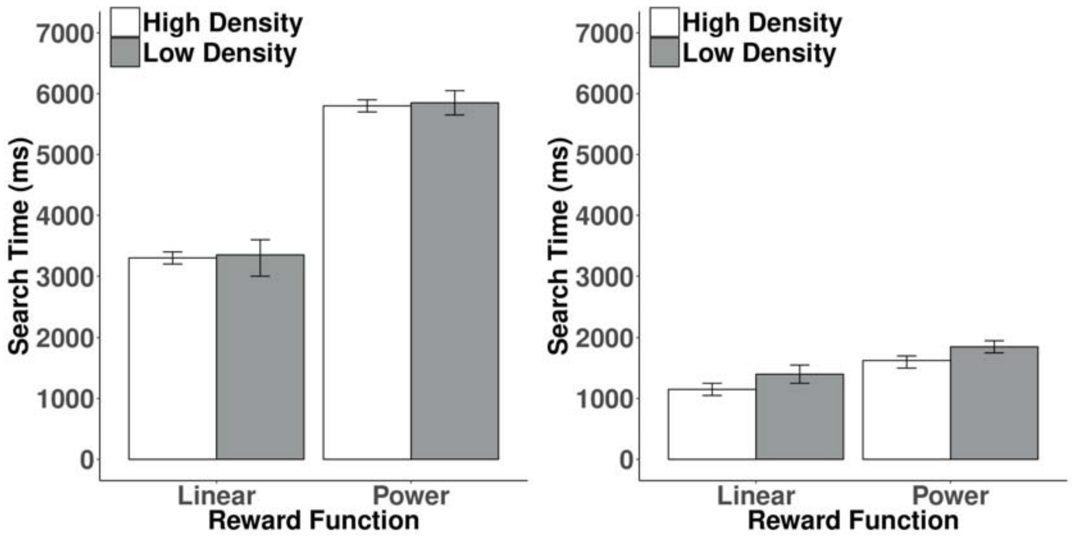
(a) (b)

Figure 4.8: Mean search performance as function of number of fixations for (a) Human and (b) Bayes model. Error bars represent standard error



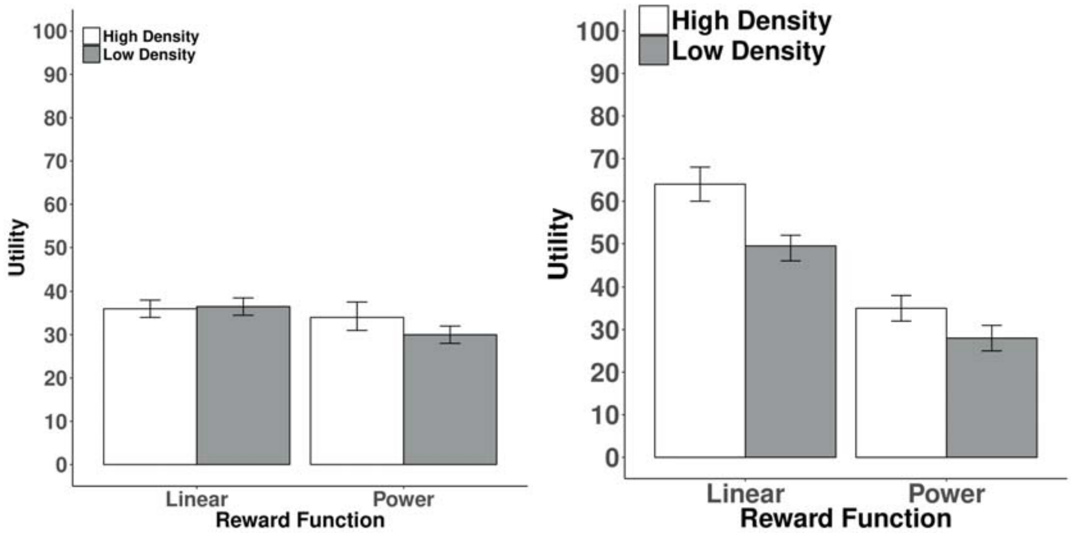
(a) (b)

Figure 4.9: Mean search performance as function of number of target features selected in an image for (a) Human and (b) Bayes model. Error bars represent standard error



(a) (b)

Figure 4.10: Mean search performance as function of search time for (a) Human and (b) Bayes model. Error bars represent standard error



(a) (b)

Figure 4.11: Mean search performance as function of reward (image value/search time (in seconds)) for (a) Human and (b) Bayes model. Error bars represent standard error

In the section below, the search performance of the optimal control model with Bayes update is reported and compared to the human performance for image search task. The results reported here are from a model that learns the optimal policy (reward value per trial converges) through training and was used to generate the behaviour (last 10,000 trials of the simulation).

The plot of the reward function versus the number of fixations is shown in Figure 4.8. The result shows that the model required fewer fixations in the linear function as compared to the power function. Also, when the image density was high, the model required fewer fixation as compared to a low-density condition. This behaviour is consistent with human performance with the goodness of fit for the model being $R^2 = 0.84$ and the Mean squared error value being $RMSE = 5.35$. Here, the model can replicate the human performance for the difference in the number of fixation behaviour between linear and power reward function. However, the magnitude of fixation was lower due to the scaled down the display.

The plot of reward function versus target features selected in an image is shown in Figure 4.8. The deep reinforcement learning model found the higher number of matching target features in an image when the reward function used power-law as compared to the linear law. This behaviour is also replicated in the human performance with our model achieving goodness of fit value of $R^2 = 0.71$ and the Mean squared error value being $RMSE = 1.03$. Again, the magnitude was lower because we scaled down the problem size.

The model behaviour shown above is consistent with the search time effect and the total reward gathered by the model. Here the model took a long time to select an image in the power law function as compared to the linear law. This was because the payoff was better with higher image value in the power function condition. The goodness of fit with human data was $R^2 = 0.80$ and the Mean squared error value being $RMSE = 3298.98$.

4.7 Discussion

4.7.1 Model Comparison

As shown above, both the CRVS and the Bayes model were able to predict the human performance. Our intention here was not to vary the theory of how information is updated, but, to show the scalability of the approaches. The theory assumes that people maintain an optimal state estimate on each fixation. To achieve this, two approaches were tested: 1. using Bayes theorem 2. using a recurrent neural network. While the CRVS model was able to scale up to the task (using the recurrent network as state estimator) and accurately predict human-like visual search strategies, in contrast, the Bayes model failed to scale up to the actual task. To maintain the likelihood table for just 3x3 display required 85 million rows of observations encountered. However, the Bayes model was able to predict the change in visual search strategies with the change in reward. Both the models were able to learn the feature threshold for when to stop searching that minimises reward defined as a user preference. While interpreting the Q-table states is easy, what internal representation the recurrent model learns is difficult to decode. The recurrent model can scale up to the actual experiment display. However, it is difficult to interpret what features it has learned from the input.

4.7.2 General Discussion

In this chapter, we proposed a new model for explaining and predicting users' visual search in complex interactive tasks. Unlike any other model, ours uses deep reinforcement learning to *generate* human-like visual search strategies. Moreover, we showed that our model accurately predicted when people would terminate visual search; and that this stopping rule was a function of payoffs, the design of the interface (item spacing) and the ecological structure of the task (the distribution of matching goal features in each display). Also, the strategy choice of the number of fixations and number of target features selected can be explained as an approximately optimal adaptation to these constraints.

The model presented here can generate human-like strategies is a departure from the traditional heuristic models that use hand-coded rules (Kieras and Hornof, 2014, Kieras et al., 2015*b*, Everett and Byrne, 2004) and approaches that learn from human examples (Li et al., 2018). Instead, strategies are derived by defining theories of constraints. Specifically, the model shows that a small change in the visual angle between items and changes in the reward structure that reflects user preference causes qualitative changes in strategy. In the model, observations were constrained by a theory of foveal and parafoveal vision that encode information about high-level features that are available. A uniform distribution of distractors in the image was assumed due to the low accuracy of the search engine. It, therefore, represents a progression to models that require fewer inputs from analysts in order to model human cognition.

Also, the reported model adds to the growing body of research that the computational models of adaptive behaviour offer mean of explaining the behaviour as an adaptation to constraints (Brumby et al., 2009, Howes et al., 2009, Chen et al., 2017). The model findings presented here is not only restricted to explaining visual search behaviour but also is crucial to human-computer interaction literature. Specifically, model findings show that the decision people make is affected by small changes in distance (in visual angle) between items and changes in reward structure. These findings support the view that people are *computationally rational*, and they adapt their strategies to known information processing constraints, the statistical distribution of task and the cost of taking any action (Lewis et al., 2014, Howes et al., 2009). It thereby helps make the crucial link between the cognitive mechanism and rationality that supports more in-depth explanations of behaviour.

While the model does predict the visual strategy people may use during the image search task, it does not take into account the mouse control movement. This is specifically seen in the search time results where we see a large difference in actual search time and the predicted. An extension to this work will be to model the mouse control movement as well. Also, we assume that the distractor distribution in each image is normally distributed. This assumption about the distribution may not be true. Further analysis of the image data set needs to be done to find out the distractor distribution.

In conclusion, we have reported and evaluated a model of how people make a decision when interacting with a web page of images. Here, the strategies are derived by formulating the problem as a POMDP and solving the problem to find the optimal solution that is adapted to the human perceptual constraints, the statistical distribution of images on the web page, and the speed/accuracy trade-off function.

CHAPTER 5

WILLIAMS OBJECT SEARCH TASK

5.1 Introduction

In this chapter, we apply the *CRVS* model to our second visual search experimental task, i.e., the Williams object search task. The goal of this search task is to explore the effect of low-level visual features on search strategy. The task required the participant to search for a target object (with either known colour, shape or size feature information) among other objects. The control problem that the model needs to solve here is where to fixate next.

Previous cognitive models like EPIC have utilised hand-coded rules to explain the search behaviour in the Williams object search task. In contrast, we show that an approximately optimal control model can explain the emergent behaviour without any description of the hand-coded rules. Instead, it explains the behaviour as an adaption to constraints in human visual system and memory. Here, we demonstrate that the observed effects like saccadic selectivity (proportion of fixation landing on target features), the number of fixations and saccadic distance can be explained as an adaptation to the difference in decline in acuity for different low-level features in human vision and the decay in information retention.

The *CRVS* model presented in this chapter represents the environment as a display of 75 objects. Where each object in the model is represented as presence or absence of the target feature (e.g., colour, shape and size). The actions than the model can take consist of 75 fixations action and a select action that selects the fixated image and terminates the trail. In this chapter, we use a history based state update with time based decay of information. Our intention here is to emulate a time based memory decay seen in human behaviour (McCarley et al., 2003). In the model we constraint the information maintained by the model using a decay parameter. In this chapter, we fit the decay parameter that best produces human like behaviour. It is to be notes that the model was not fitted to human data to best produce the behaviour result. Rather, only the constraint parameter was fitted.

5.2 The Task

In the visual search task published by Williams (1967) and re-done by Kieras et al. (2015*a*) the goal is to locate a target object in a field of 75 distractor objects as soon as possible. Each object in the display is described by a unique two-digit number and a unique combination of colour, size and shape. At the start of a trail, the participants were provided with a unique number that identified the target and one of the low-level features of the target, i.e., either colour, shape or size.

The 75 unique objects were randomly distributed across a search field of visual angle 39 by 30. figure-5.1 shows a re-creation of the display used in the experiment. Each object was sampled with a unique colour, shape and size. Where, the colours were blue, green, yellow, red and purple. The Sizes were small (0.8), medium (1.6), and large (2.8). The shapes were circles, semi-circles, squares, equilateral triangles and crosses. At the centre of each object, a unique two-digit number was presented from 01 to 75 with a size of 0.26. Between each object, a distance of at least 1 was maintained.

In the experiment, the participants were first pre-cued with the target features. This was followed by the disappearance of the target features and appearance of the search field. When the participants found the target, they were asked to click the target object

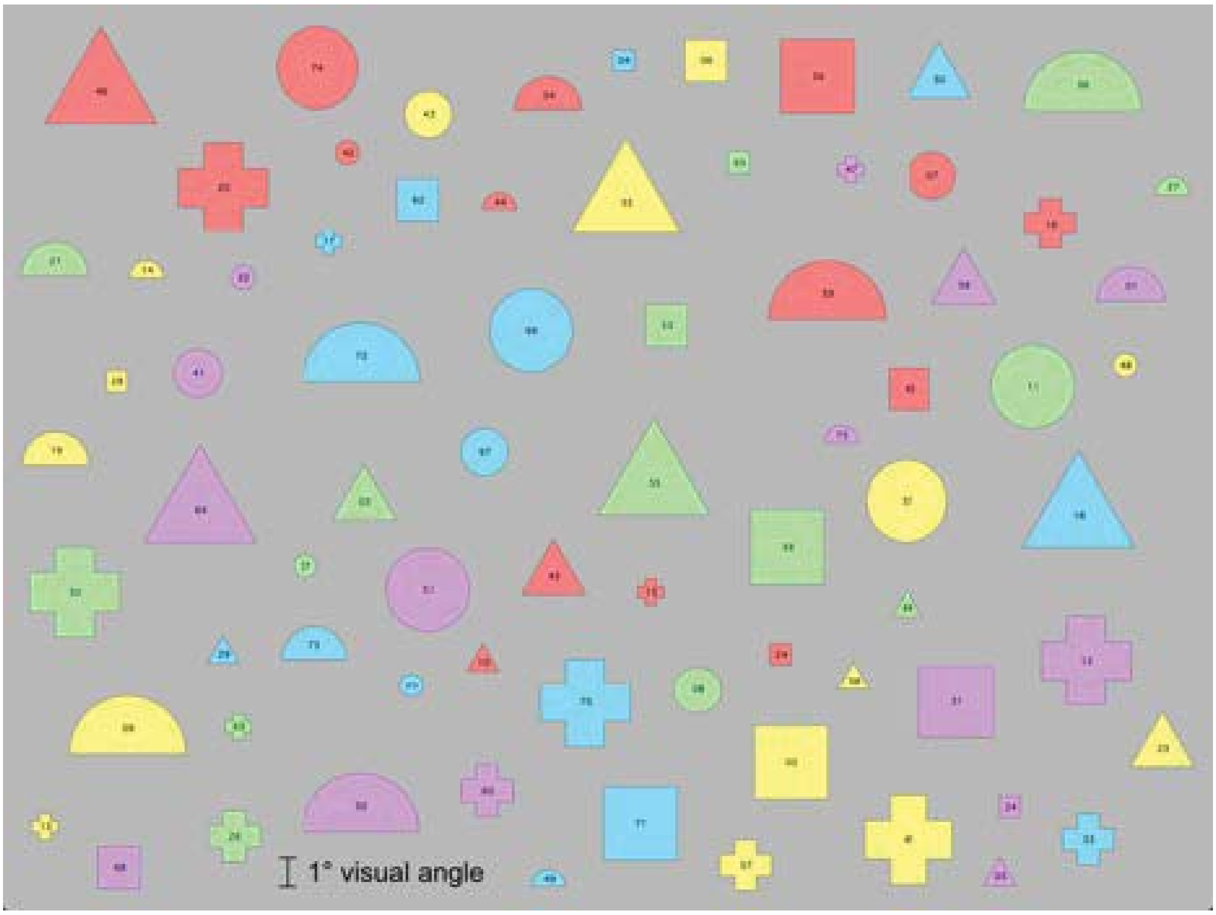


*Figure 5.1: A sample display used in the experiment. Task display used from (Kieras et al., 2015*a*)*

with the mouse. Here, participants were constrained not to move the mouse until they found the target. On successful completion of a trial, the participants were financially rewarded. Each trial started with a bonus of five, twelve or twenty-one cents, depending on the difficulty of the search condition. The bonus diminished at a rate of 0.4, 0.3, 0.15 cents per second respectively until the participants clicked on the target. Errors resulted in a penalty of five cents.

5.2.1 Observed Data

Human performance for the proportion of fixations that landed on the objects that shared the same cued features is presented in figure-5.2. Here, the participants found the colour and size cues to be more effective in finding the target object since colour and size cue

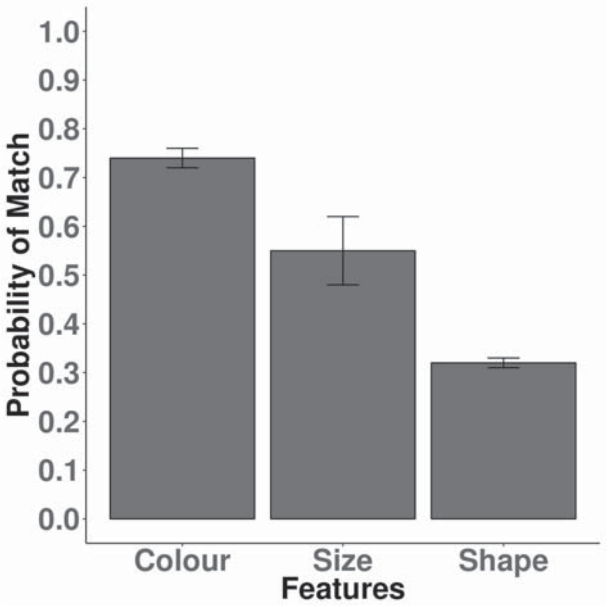


*Figure 5.2: figure re-plotted from (Kieras et al., 2015*a*) showing proportion of fixations that match the target cue.*

can be recognised over wider eccentricity range (Gordon and Abramov, 1977). Hence, the proportion of fixation that landed on these objects was higher. In comparison, the shape feature was the least useful cue between the three cues since recognising the shape required detailed scrutiny of features like edges and this ability declines rapidly with eccentricity (Kieras and Hornof, 2014).

figure-5.3 shows the mean number of fixations across trials participants took for each cue. The colour cue took the least amount of fixation followed by size since the colour information is available over wider eccentricity range (Gordon and Abramov, 1977). In comparison, the shape cue required the highest number of fixations.

Another exciting result reported by Kieras et al. (2015*a*) was that on an average across cues, 25% of the fixations were on a previously fixated object. Out of which, 14% of the fixations were immediate repeats with little variation between different cues. Peterson et al. (2001) in his findings had reported that repeated fixation are rare and are roughly around 5.7%. Kieras et al. (2015*a*) in their findings suggested that these immediate fixations could be accounted for by encoding error in peripheral vision.

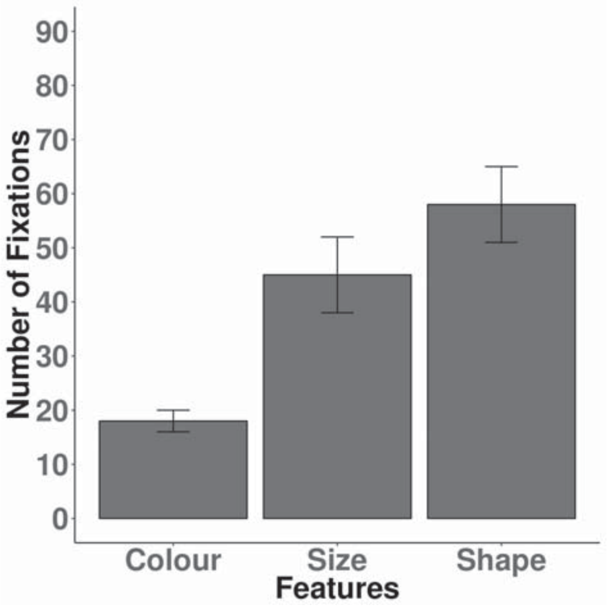


*Figure 5.3: figure re-plotted from (Kieras et al., 2015*a*) showing mean number of fixations required to successfully complete a trial for each target cue.*

5.3 Background

The search behaviour in Williams object search task was previously explained by using an EPIC cognitive architecture (Kieras et al., 2015*a*). In their model, the search strategy was adapted to the constraints in the human visual system. They utilised the *active vision* approach to constraint the amount of information being gathered on each fixation. The strategy adopted by the model to explain the search behaviour was defined using production rules (see section 2.3.4 for review of cognitive architectures like EPIC). In the model, the strategy to fixate on next item is described by a priority scheme. Here, the priority was defined on the basis the most available information over the least available information. For example, colour information had the highest availability in the parafoveal region in the model and hence had the highest priority, followed by size and finally shape. The priority is also dependent on the pre-cued feature shown to the participants. Alternatively, when no pre-cued feature information is available in the parafoveal region, a location is randomly selected. The strategy employed by the EPIC model as described by (Kieras et al., 2015*a*) can explain the guided gaze movement to the corresponding known target feature objects. However, it fails to explain the unguided gaze movements when no known target feature is available in the parafoveal region.

In contrast, to the EPIC architecture, we present a novel model of visual search that explains the search behaviour as an optimal adaptation to both human information processing constraints in the visual system and the task reward. Here in the model, we use two sources of constraints, i.e., the *feature noise* and *memory decay*. Both sources of noise have been previously shown to be essential to modelling the visual search behaviour in Willams' object search task (Kieras and Hornof, 2014, Kieras et al., 2015*a*, Williams, 1967). In this chapter, we contrast our model with the EPIC models that have been used previously to provide an interpretation to the search task.

What follows is a description of the underlying theory used in our model to interpret the search behaviour. Then the description and results of the model are presented in the chapter. Finally, the implication of the model is discussed.

5.4 Theory

In this chapter, we present a control model with the assumption that the search behaviour in William's object search task can be explained as an adaptation to the feature noise that constraints the human visual system and the decay in information retention time that constraints the human working memory. The feature noise can be described as the decline in the ability of the human visual system to discriminate and perceive information with an increase in eccentricity (the distance of the object from the centre of gaze) (Strasburger et al., 2011). This decline is also influenced by the low-level feature that defines an object, for example, colour, shape, size and orientation (Findlay and Gilchrist, 2003). Kieras and Hornof (2014) showed that the colour features have the highest degree of availability in the parafoveal region, followed by the size feature (larger objects are easily perceived compared to smaller objects), and finally shape features.

In addition to the feature noise, the model also includes a constraint on the working memory. Peterson et al. (2001) in their experiment showed that visual search retains information about previous fixations. They recorded the search behaviour of participants for a visual search task where 12 letters were randomly distributed in a display. In the display, the target letter 'T' was surrounded by letter 'L' and the goal was to find the letter T as quickly as possible. Peterson et al. (2001) noted that the re-fixation to the previously visited location was significantly lower than what would be in a memoryless system. They concluded that this behaviour is only possible if people maintained a memory of previous fixations. In addition to people maintaining a history of the previous fixation, there is a limit to the capacity of what they can retain. McCarley et al. (2003) in their study showed that the probability of re-fixating to a previously visited location increases with time.

Therefore, our theory here is that because of feature noise the availability of information is limited. For the colour feature, the availability of information in parafoveal region spans the highest and hence, we predict that proportion of fixations that will land on objects with the colour target will be the highest followed by size feature and finally shape. We predict this behaviour will also replicate for the number of fixation due to the availability of information, where the target with known colour cue will require fewer fixations in comparison to size and shape.

Due to decay in memory, the information about previously fixated location will be lost with time. Unless the location is revisited, or information is available in parafovea. We predict due to memory decay the number of fixations will increase and the saccadic distance will be the highest for the colour feature since information is available further. However, to optimise information gain versus memory decay, we predict a sizeable saccadic distance for weak cue like shape, to reduce the overall uncertainty in the display quickly.

In the following section, we formalise this theory in the form of a computational model in which the above-specified behaviour emerges.

5.5 CRVS Model with History based update

To construct the approximately optimal control model description of the (a) visual acuity, (b) what information is gathered and stored and (c) for how long the information persists in memory is required. In sections below, A walk-through of the 75 elements visual search task is presented with a description of the above three components.

5.5.1 External Display

In the model, we represent the display by randomly distributing the 75 objects in a grid, where each object can span multiple cells based on the size of the object. In the model, we used a matrix of size 5 x 15 to represent the display with 75 objects randomly distributed in the display. In the display, a minimum gap of 1 degree (visual angle) was maintained between objects and an additional gap was randomly sampled from a distribution. In the model, a uniform distribution over gap with [0.0, 1.0, 2.0, 3.0] degrees were used to sample additional gap size.

Each cell in the display is represented by the presence or absence of target feature at each object location where the presence or absence of a feature at each location in the model is represented numerically by the number 1 for presence and 0 for absence. The presence and absence of each object feature in the model is maintained in a vector. In the model, four vectors are kept for colour, shape, size and text features.

5.5.2 Actions

The action space consists of (1) fixate on a location, and (2) click the fixated object. In our study, there was a grid of 75 objects with unique colours, shapes and sizes. Therefore, a total of 75 possible fixation actions were possible for the model to take. A trial was terminated when the click fixated object action was chosen.

5.5.3 Reward

A reward was given after clicking the fixated object. The reward distribution was defined as a value 5, 12 or 21 for a correct response depending on the difficulty of the search condition, a value of -5 for an incorrect response and a value of -0.4 * $SearchTime$, -0.3 * $SearchTime$, or -0.15 * $SearchTime$ for each fixation depending on the difficulty of the search condition. Where the search time (in seconds) is defined in equation 5.1 We assume in the model that the mean fixation duration for high-speed visual search is around 275ms (Rayner and Castelhano, 2008). Here, in the experiment, the level of search difficulty in ascending order is described as colour, size and shape. The penalty on each fixation imposes a speed-accuracy trade-off. More fixations give greater accuracy but at a cost.

5.5.4 Observation Model

visual constraint

The observation model is the same as described in section 4.5.4

memory constraint

In addition to the visual constraint, the model also constraints the duration of information being persisted in the model. This is achieved by maintaining a decay threshold in the model. If last perceived information from a location exceeds the decay time threshold, the last seen information is removed. In the model, the time cost added to each fixation is described below,

$$SearchTime = SaccadicDuration + FixationDuration \tag{5.1}$$

$$SaccadicDuration = 37 + 2.7 * Eccentricity \tag{5.2}$$

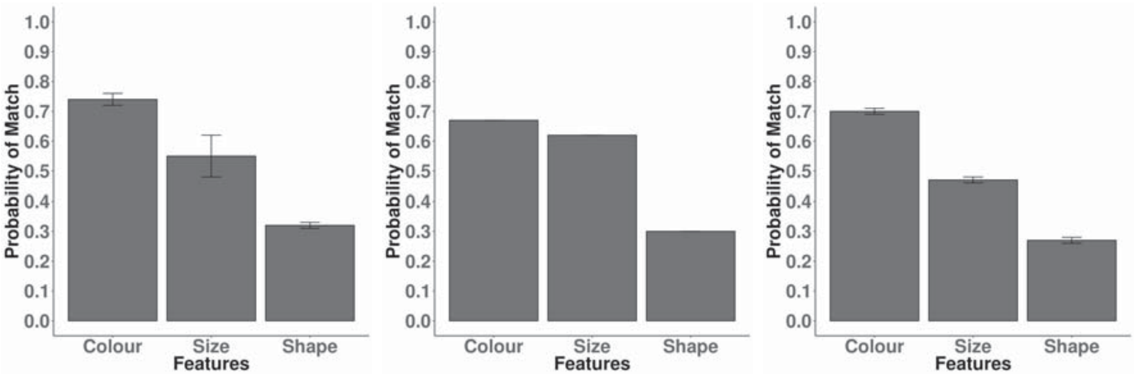
Where the Fixation Duration was kept constant at 275ms (Rayner and Castelhano, 2008) for all features, eccentricity is the distance between fixated location and item of interest in degrees visual angle.

5.5.5 State Space

In the model, we combine the available feature and text observation vectors to obtain a scalar score for each location. This scalar score is defined as a *relevance score* in the model. The *relevance score* for each cell is a Euclidean distance between the target features, i.e., $[1, 1]$ and the available feature and text observations for each location (6.3). The *relevance score* is capped between 0 and 1 where a value 1 represents high relevance to task and value 0 represents low relevance to the task.

$$Score_t[i, j] = min(max(0, \sqrt{(1 - \delta_{feature}[i, j])^2 + (1 - \delta_{text}[i, j])^2}), 1.0) \tag{5.3}$$

This combined score is then used as a state vector s_t. The state vector according to the model persists in the working memory. Every time new information is obtained at a specific location the model updates the information at that location in the state space. The model also maintains the persistence time for each information in the state space. The persistence time acts as a time counter for each location in the state space. It monitors when last the information was updated by adding a time cost (see section 5.5.4) to the counter on each fixation. On reaching a decay threshold, the information is removed from the state space and replace with a value 0.5 which represents uncertainty. On the availability of each information in the state space, the persistence time is reset to 0 for the location being updated.



(a) Human (b) EPIC (c) CRVS

*Figure 5.4: Mean search performance for proportion of fixations that match the target feature for Human, EPIC (plotted from (Kieras et al., 2015*a*)) and CRVS model.*

5.6 Model Learning

5.6.1 Model Architecture

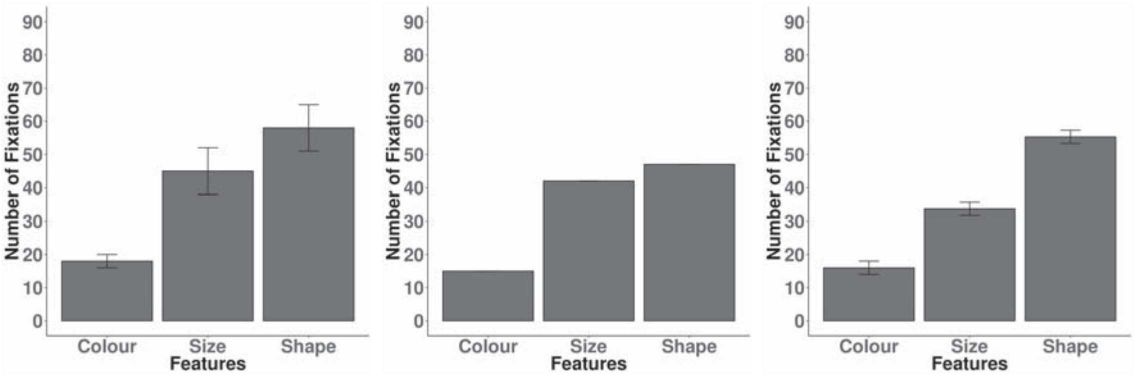
The relevance score estimates $Score_t$ (75 element vector) from the state estimator (above) was taken as the input. This input was connected to three fully connected hidden layer consisting of nodes equivalent to the number of elements in the display, i.e., 75, with rectifier activation function. Finally, the output layer was a fully connected linear layer of 76 nodes with a single output for each action in the task. To avoid over-fitting of the network $l2$ regularisation of the weights was applied with value 10^{-5} and Dropout with rate 0.2 was added to each hidden layer except the final layer.

5.6.2 Policy Learning

The algorithm used to learn policy is same as described section 4.5.6.

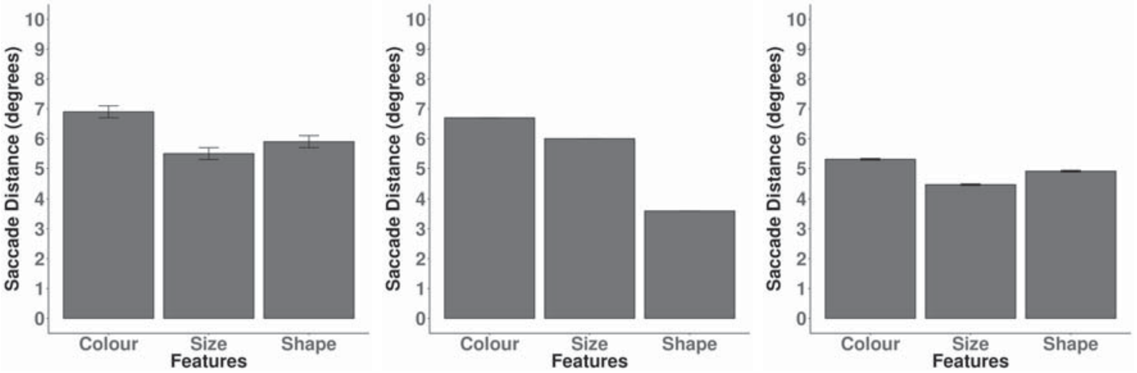
5.7 Results

In the section below, the search performance of the CRVS model is reported and compared to the human performance and the EPIC architecture performance reported by Kieras



(a) Human (b) EPIC (c) CRVS

*Figure 5.5: Mean search performance for number of fixations required to successfully complete a trial for each target cue for Human, EPIC (plotted from (Kieras et al., 2015*a*)) and CRVS model.*



(a) Human (b) EPIC (c) CRVS

*Figure 5.6: Mean search performance for saccade distance from the previous fixation to the fixated object for Human, EPIC (plotted from (Kieras et al., 2015*a*)) and CRVS model.*

et al. (2015*a*) for the Williams object search task. The results reported here are from a model (best fit model) that learns the approximately optimal policy through training and was used to generate the behaviour (last 50,000 trials of the simulation).

The plot for the proportion of fixations that landed on the objects that shared the same target features is presented in Figure 5.4. The result shows that the CRVS model predicted the highest proportion of fixations that landed on the objects that shared the same features as the pre-cued colour feature followed by the size and shape feature. This behaviour is consistent with human search performance with the goodness of fit for the model being $R^2 = 0.98$ and the mean squared error value being $RMSE = 0.04$. Here, the model replicated the human performance, where both people and the model found the colour as the most reliable cue with larger availability span across para-foveal region followed by size and finally shape feature.

The plot for the number of fixations per trial for successful completion is presented in Figure 5.5. The results show that the colour cue required a fewer number of fixations where guided fixations dominated. This was followed by the size feature where larger objects were further available in para-foveal, and hence fixations were guided towards those objects. For smaller sized target objects, more fixations were required due to the low availability of features in the parafoveal region. Finally, shape feature which was dominated by unguided fixations due to lowest availability of information in the para-foveal region and required the highest number of fixations. This behaviour is consistent with human search performance with the goodness of fit for the model being $R^2 = 0.97$ and the mean squared error value being $RMSE = 17.89$. However, the magnitude of the fixations was not replicated as seen in human performance. This could be due to the uniform distribution used to generate the ecology of the gaps between objects. As shown in chapter 4, a small increase in gap increases the number of fixations.

The plot for the saccadic distance from the previous fixated location to current fixated location is presented in Figure 5.6. The results show that the saccadic distance was highest when the cued feature was the colour feature due to the high availability of information making larger jumps. In comparison to colour feature, the size and shape features were comparatively lower due to lower availability of information. However, the saccadic distance for shape feature was higher in comparison to size feature although the availability of information is lowest for shape feature. This behaviour is contributed to the memory decay where the strategy is adapted to the information retention time where large retention time leads to larger saccadic jumps and lower retention time lead to smaller saccadic jumps. This behaviour is consistent with human search performance with the goodness of fit for the model being $R^2 = 0.92$ and the mean squared error value being $RMSE = 1.24$.

The repeated fixations seen for 5 seconds of information retention time threshold set in the model was 23% across all features. Individually, for the colour feature, the repeated fixation was 14.4%, for size feature the repeated fixation was 21.3% and for shape feature the repeated fixation was 35.1%. The overall reported fixation was around 25% for human performance. While we did explore the retention time threshold (see appendix B) the best fit model was for $5secs$. By decreasing the threshold time further, the magnitude of fixation did improve but at the cost of an increase in repeated fixations.

5.8 Model Comparison

The figures 5.4, 5.5 and 5.6 show a comparison between the search performance of the EPIC architecture (Kieras et al., 2015a) and the CRVS model for the Williams object search task. Our results show that both the models were able to predict the difference in the proportion of fixations and the number of fixations per trial for each pre-cued features. However, the magnitude of fixations was low for both the EPIC and CRVS model.

The two models differed in how the strategy was defined. In the case of EPIC architecture, a hand-coded rule was defined where the model chose a candidate object according to a priority rule. The priority rule made fixation choices according to the availability of the features, and so the objects with matching colour feature were given the highest priority followed by size and then shape. If there are no candidate objects, then a random object is chosen whose properties are unknown. In contrast, the CRVS model was provided with just the theory of constraints and the reward. The strategy emerged as an approximately optimal (optimisation algorithm like Deep Reinforcement learning used to find the solution) adaptation to the underlying theory.

Furthermore, a clear distinction of strategy choice was seen for unguided fixations in case of shape features (Figure 5.6), where the EPIC architecture made small saccadic jumps. However, the CRVS model predicted a large saccadic jump which was consistent with human behaviour. The strategy used by the CRVS model adapted to the memory decay constraint.

5.9 Discussion

In this chapter, we presented a computational model for explaining and predicting users' visual search strategies and the role of working memory for Williams object search task. The model was able to generate human-like behaviour, given a theory of constraints in the human visual system and the memory decay. The model predicted where to fixate next strategy as a function of the acuity of low-level features like colour, size and shape. The results show that the proportion of fixation that landed on target features was similar to the human performance where the colour feature had that highest proportion followed by size and shape features. Since the colour feature has the highest range of availability across the periphery (Gordon and Abramov, 1977), the model was able to perceive target colour at a broader range. As a result, the model learned to guide its attention towards those objects that matched the target colour features. The next highest availability was seen for size followed by the shape feature which saw the least proportion of fixations.

The number of fixations was also predicted by the model as an adaptation to both the constraints on the acuity of features in the periphery and the memory decay. In the case of colour feature the memory decay had a little effect. Since the colour feature could be perceived at a broader range, the model required fewer fixations and completed the trail before the set decay threshold and with fewer repeated fixations. In contrast, the size and shape feature saw a higher number of repeated fixations due to memory decay. For size feature, the larger sized objects were perceived at a wider range, but the smaller sized objects were difficult to perceive which contributed to a higher number of fixations and saw higher repeated fixations. Finally, the shape feature had the least availability of the three and required the highest number of fixations and repeated fixations.

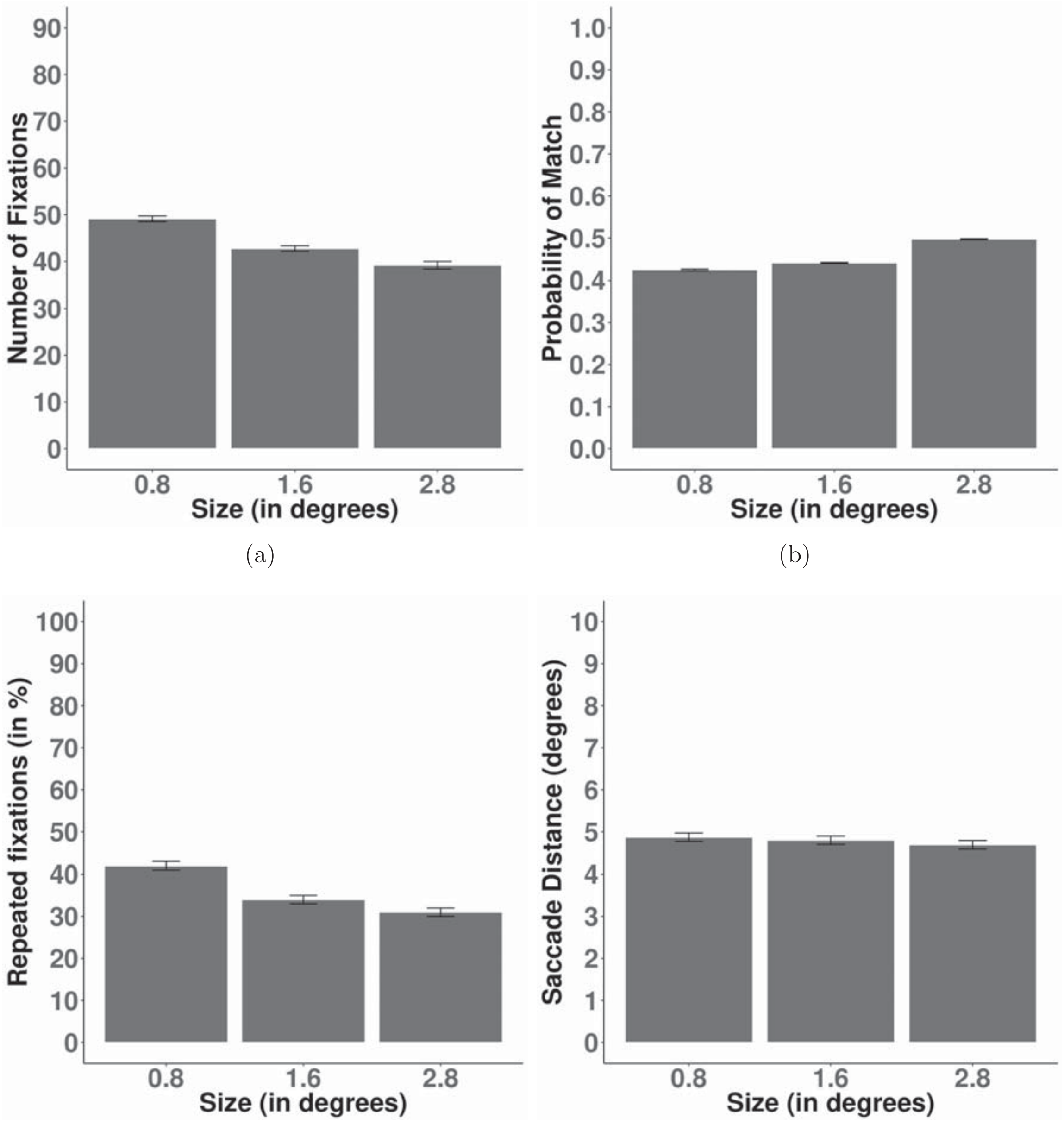
Furthermore, the saccadic distance predicted by the model was adapted to the acuity of features in the periphery, and the memory decay as a consequence of colour feature having the highest acuity had the largest saccadic distance followed by shape and size. The model made saccadic distance jumps as an adaptation to the information decay and the feature acuity. When no decay was present, the model chooses to cover a larger area because once the information enters the working memory, it remains throughout the trail. In contrast, when the information decays very quickly, the model chose to make smaller jumps to keep the previously fixated information available in the parafovea. Also, the saccadic distance adapted to the acuity of features, for example, in figure 5.7, the saccadic distance increased with the difficulty of perceiving the object as a function of its size. This explains the lower saccadic distance seen in human performance between size and shape cue. As the cue availability reduces the number of unguided fixations increases and hence the model makes larger jumps so as to reduce overall uncertainty in the display.

Also, the model shows that a working memory that represents the spatial position of objects in the display is essential in guiding attention when the target information changes on each trial. The model supports the view on visual working memory playing an active role in finding targets and reject distractors in a visual search task as suggested previously in literature Peterson et al. (2001), McCarley et al. (2003), Woodman and Chun (2006). This was further evident from our results (see appendix), as the models' ability to retain information decreased the model ability to reject previously seen distractors reduced and thereby reducing the proportion of fixations landing to target features.

Previous approaches like the EPIC architecture that required pre-defined hand-coded rules provided an interpretation to the Williams search task. They showed that the pattern of human fixations could be explained as a consequence of differences in colour, shape and size acuity. However, the strategy encoded was sub-optimal and couldn't explain the unguided fixation pattern for low acuity feature, e.g., when the target feature information was a shape cue. In contrast, the CRVS explained the search behaviour as an adaption to the acuity of the low-level features and the decay in memory. In the CRVS model, no handed-coded rules were provided. Instead, the strategy emerged as an adaptation to the constraints. The use of optimisation algorithm ensured the causality between the acuity constraint and the memory decay with the human behaviour.

Further work needs to be done to be confident about the model behaviour. We need to explore the gap parameter space of the CRVS model fully. For example, changing the gap distribution between the objects may improve the fit to the number of fixations. As shown in Chapter 4, a small change in object density changes the behaviour. This might also impact saccadic distance reported by the model. Also, the CRVS model needs to be extended with recurrent based state estimation approach and explore the behaviour of this model. In the current version, we did not use the recurrent architecture due to its property to retain long-term information. As part of future work, we will explore the Basic recurrent neural network architecture that is prone to information decay.

In conclusion, we have reported and evaluated a model of the human scan path. The model predicted where to fixate next strategies as a function of feature acuity. Here, the strategies are derived by formulating the problem as a POMDP and solving the problem to find the approximately optimal solution that is adapted to the human perceptual constraints and memory. +



(a)

(b)

(c)

(d)

Figure 5.7: Plot showing strategy learned by the model for different cue sizes. (a) Average number of fixations per trial for each size cue. (b) Proportion of fixation landing on target size cue. (c) Repeated fixations observed per trial for each size cue. (d) Saccadic distance observed per trial for each size cue.

CHAPTER 6

THE DISTRACTOR RATIO TASK

6.1 Introduction

In this chapter, we apply the *approximately optimal control model* to a visual search experimental task, i.e., the Distractor Ratio (DR) task. In this task, the participant has to find a target item (which may or may not be present) surrounded by some distractors which share one common feature with the target. Here the control problem involves learning both when to stop searching and where to look next.

Previous modelling approaches, like the bottom-up salience and maximum a posteriori (MAP), have utilised heuristic rules to explain the search behaviour in the DR task. In contrast, we show that an approximately optimal control model can explain the search behaviour without any description of the heuristic rule. Instead, it explains the search behaviour as a consequence of adaption to constraints in the human visual system and the reward. Here, we demonstrate that the two empirical phenomena, i.e., the distractor ratio curve (inverted U shape) and the saccadic selectivity can be explained as an adaptation to the uncertainty introduced by the crowding of objects which results in smearing of features in human vision, and the speed-accuracy trade-off.

The *CRVS* model presented in this chapter represents the environment as a display of 36 objects. Where each object in the model is represented as presence or absence of the target feature (e.g., colour and shape). The actions than the model can take consist of 36 fixations action and choice of present or absent action. In this chapter, we use a recurrent neural network with history based state update. We do not use any time decay in this model because the search task terminates within few seconds for memory decay to have any behavioural effect. Furthermore, we compare the recurrent model against two other behaviour model previously reported in literature. First, is the Kalman filter based state update model (adapted to the CRVS model) first reported by Sprague and Ballard (2004). Second, is a heuristic model (adapted to the CRVS model) reported by Myers et al. (2013). Here, we fit the noise parameters that constraint the amount of information perceived by the model that best fit the human data. We did not fit the model behaviour/strategy to the human data.

6.2 The Task

In the distractor ratio task, the participants were asked to find a target object surrounded by some distractor objects. The participants were informed about the features that define a target and the distractor's. At the start of the trial, a display is presented to the participant which may or may not consists of a target object (randomly positioned) and some distractor objects each of which shares at least one common feature with the target. The number of colour to shape distractors present in the display is a controlled parameter in the experiment and is unknown to the user. The goal is to respond whether the target is present or absent on each trial. An example display is shown in Figure 6.1 where the target is a red letter O. The distractors in this display share either a same-colour 'red' or same-shape letter 'O' feature with the target.

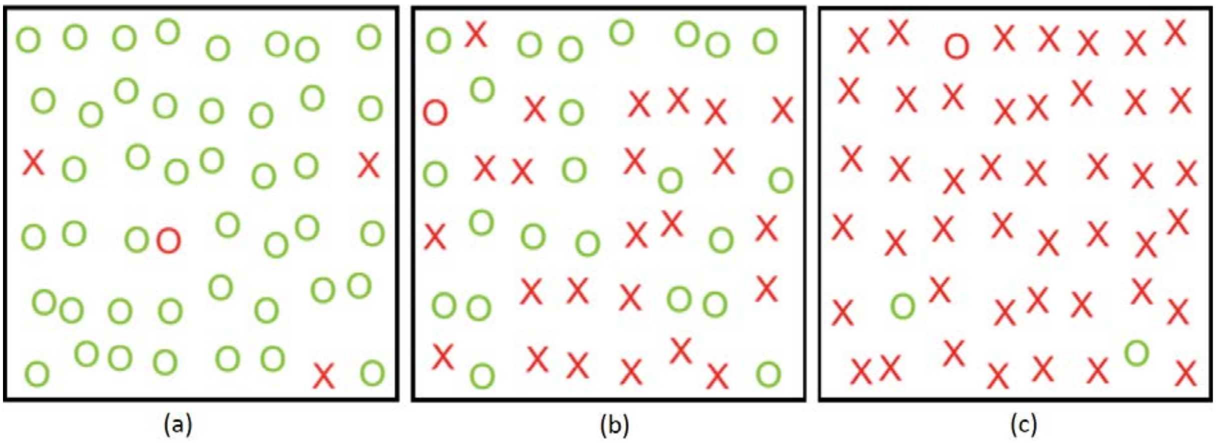
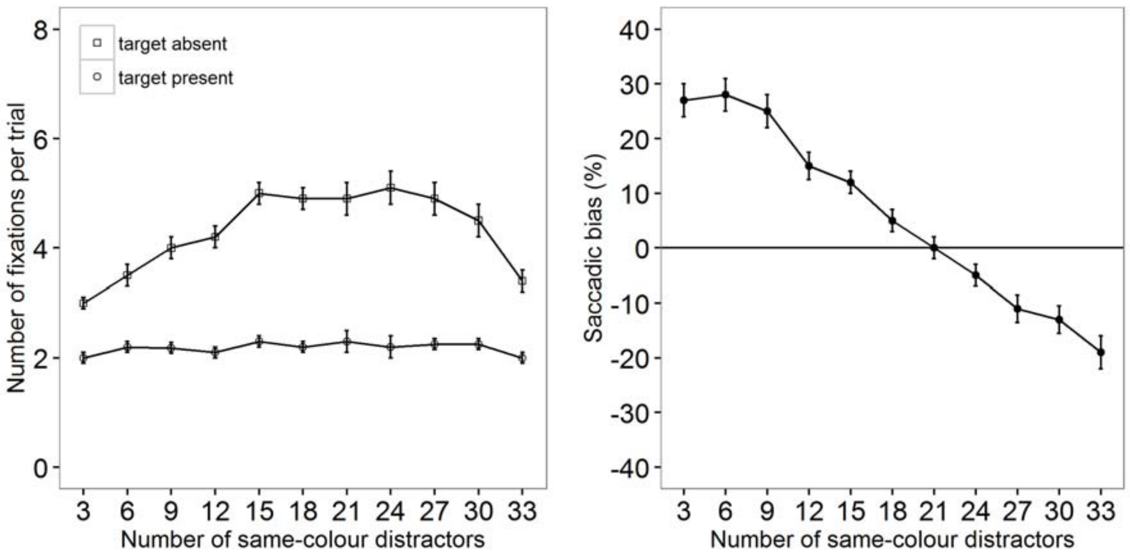


Figure 6.1: A sample distractor-ratio display with ratio distribution: (a) 3:45, (b) 24:24, (c) 46:2 and target stimuli a colour red letter O.

Shen et al. (2003) through his experiment observed and reported that the participants took fewer fixations and responded quickly whether the target was present or absent when the distractor ratio for same colour to shape and the same shape to colour distractors was in the minority (as shown in Figure-6.2 (a)). For example, in Figure-6.1 participants took fewer fixation to find the target 'red letter O' in the display (a) and (c) with ratios 3:45 and 46:2 respectively in contrast to display (b) which required more number of fixations to respond Shen et al. (2003). This effect was especially significant for target absent trials. The inverted-U like shape (see Figure-6.2 (a)) is termed as distractor-ratio effect.



(a)

(b)

Figure 6.2: (a) Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for high discriminability condition. (b) Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target- absent trials for high discriminability condition (Shen et al., 2003)

In addition to the distractor-ratio effect Shen et al. (2003) also reported and observed a systematic shift in saccadic towards the distractor features which were in the minority. In Figure 6.2 (b), the frequency of saccades to same-colour distractors is plotted against the number of same-colour distractors. In the plot, the saccadic frequencies are higher for rare features (colour or shape) than should be expected by chance (represented by the horizontal line). When the same-colour distractors are rare in the display, the participants were more likely to make eye movements towards them than when they were common. Conversely, when the number of same-colour distractors was high, the participants were more likely to make eye movement towards same-shape distractors.

6.3 Background

Previously, two approaches were used to explain the two empirical phenomena in distractor ratio task. First, is the map-based approach as described by Kowler (2011). In this approach, different information sources are combined together to form a hypothetical map-based representation, such as *activation map* (Pomplun et al., 2003, Wolfe, 2007) or *saliency map* (Findlay and Gilchrist, 1998, Shen et al., 2000). These map-based representations consist of topological peaks, where the peak may either represent high neural activity due to the presence of target item (Wolfe, 1994) or how different they are from their surroundings items/location (Findlay and Gilchrist, 1998). The foveated vision is then directed towards these peaks sequentially by using a heuristic control, such as winner take all (Itti et al., 1998). These are the class of models that describe the distractor ratio effect using the bottom-up or stimulus-driven approach.

Alternatively, Myers et al. (2013) presented a Bayesian model to explain the distractor ratio task. According to these models, the distractor ratio effect can be explained as an adaptation to human information processing constraints in the visual system. In these models, the agent starts with a prior probability of where the target could be present. These probabilities are then updated by making a series of eye movements. The model assumes that people make eye movements to gather evidence in order to identify the underlying true display. Once enough evidence is gathered, the decision maker then responds that the target is either present or not. These decisions are made based on a threshold.

In contrast, we present a novel model of visual search that explains the distractor ratio effect as an optimal adaptation to both human information processing constraints in the visual system and the task reward. As shown in chapter 4 and 5 both the visual constraint and the reward is essential to explaining when to stop and where to look next strategies. Here in the model, we use two sources of constraints, i.e., the *feature noise* and *spatial noise* on the human visual system. Both sources of noise have been previously shown to be essential to modelling the DR Effect (Myers et al., 2013, Chen, 2015). In this chapter, we contrast our model with two other control models that have been used previously to provide an interpretation to the distractor ratio effect.

What follows is a description of the underlying theory used in our model to interpret the DR-effect. Then the description and results of each of the model present in the chapter are provided. Finally, the performances of the models are compared against each other.

6.4 Theory

In this chapter, we present a control model with the assumption that the distractor ratio effect can be explained as an adaptation to the spatial smearing that constraints the human visual system and task reward. Spatial smearing can be explained as a localisation error (Levi, 2008), where information located in the parafovea may erroneously combine features from one location with its adjacent locations. It is a well-known constraint that has previously been shown as one of the potential causes for crowding effect (Levi, 2008, Yu et al., 2009). In Eriksen's flanker task (Eriksen and Eriksen, 1974), this effect was shown for letters, where the target letter was flanked by distractor letters on either side. The participants were explicitly instructed to identify the target stimuli, i.e., the letter in the middle position and ignore the two flankers on either side. Despite the instruction, the participants were unable to ignore the distractors completely. Yu et al. (2009) postulated that this behaviour could be the consequence of smearing of information from neighbouring flanker letters on the target letter when the objects are in parafovea. The neuronal explanation was that, when the participant observes the display, an area of the display will fall within the same neural unit or within a single receptive field, which would include a number of stimuli. Yu et al. (2009) thereby assumed 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. They showed that performance of the flanker task having target letter 'S' with compatible flanker (letter 'S') and incompatible flanker (letter 'H') could be accounted for by a spatial smearing model.

We assume that the accuracy of response to identify an object with a colour and a shape will deteriorate with eccentricity due to the spatial smearing effect. For example, If the colour red and shape X is surrounded by colour green and shape O around it in the parafoveal region, as a consequence of spatial smearing, the participants would be uncertain whether the red coloured X is truly red-X or a green coloured O. Alternatively, the uncertainty reduces if the red coloured X is surrounded by red-X's since the mixing of information has little effect.

Therefore, our theory here is that, because of smearing of both shape and colour information, the observer will function under uncertainty for any distractor ratio, when the object is in the parafoveal region and surrounded by distractors. As a consequence of spatial smearing the observer will be highly uncertain for the objects in minority set, and therefore, may saccade towards those items in order to reduce the uncertainty. In addition, the observer would also be highly uncertain for the entire display when the distractor ratio is closer to 1 since there is an equal number of red-X and green-O distractors in the display. Based on our theory described above, we predict that the observer should take a longer time to respond (target present/absent) when the ratio is closer to 1 due to higher overall uncertainty in the display.

In the following section, we formalise this theory in the form of a computational model in which the above-specified behaviour emerges. We hypothesise that the control model encoded with these simple assumptions about the reward structure and spatial smearing constraint will learn policies that maximise reward. The policies are, to saccade towards minority set objects to resolve the colour and shape uncertainty caused by smearing. Also, to make more saccadic movements when the distractor ratios are closer to 1 to resolve overall uncertainty in the display.

6.5 CRVS model with Recurrent Update

In the following section, we provide a detailed overview of each of the individual component used in the optimal control model with recurrent architecture.

6.5.1 External Display

In the model, we represent the display by randomly distributing the target (if present) and the distractors in a grid-like space, where each cell consists of either a target object or a distractor object that either share a common colour or shape feature with the target. In the display, the number of common colour or shape distractors are determined by randomly sampling a ratio from a set r (where, $r = \{$ 3:33, 6:30, 9:27, 12:24, 15:21, 18:18, 21:15, 24:12, 27:9, 30:6, 33:3 $\}$) per trail. The display is represented by two feature vectors, one for colour and one for shape. Each cell can contain a value either 1 or 0. Where for colour vector value 1 represent the feature red and value 0 represents feature green. Similarly, for shape vector a value 1 represents the letter 'O', and value 0 represents letter 'X'.

6.5.2 Action Space

In the model, the action space consists of (1) fixate on a cell, (2) respond present and (3) respond absent. In our study, there was a grid of 6x6 coloured shapes, and there were, therefore, a total of 36 possible fixation actions. A trial was terminated by choice of the present or absent action.

6.5.3 Reward

The reward function uses the time cost to reward or penalise the model during the learning process. A reward of 10 for a correct response and a value of -10 for an incorrect response was given after choosing a present or absent action. For every fixation action, a time penalty is given to the model defined as $-1 * (FixationDuration + SaccadicDuration)$ where the $FixationDuration$ in target present trial is 0.230 sec and in the target-absent trial is 0.200 sec. These mean values across distractor ratio set for $FixationDuration$ is taken as reported in (Shen et al., 2003). The $SaccadicDuration$ is defined in equation-6.1 (Baloh et al., 1975),

$$SaccadicDuration = 37 + 2.7 * Amplitude \tag{6.1}$$

Where the $Amplitude$ is the distance between the current and previous fixated location, by providing a penalty on each fixation, a speed-accuracy trade-off is imposed in the model. More fixations give greater accuracy but at a cost.

6.5.4 Observation Model

Every time the model fixates, it also makes an observation. The observation obtained by the model is constrained by the noise in the human visual system. Two types of noise are added to the signal: spatial smearing noise and feature noise. Both sources of noise have been shown to be essential to modelling the DR Effect (Myers et al., 2013, Chen, 2015).

1. **Feature Noise:** The human eye's ability to discriminate and perceive object features degrades with eccentricity according to a hyperbolic function (Strasburger et al., 2011). To model this function we added Gaussian white noise with mean 0 and standard deviation as eccentricity, i.e., a function of visual angle 'θ' between the fovea and the given location, and a scalar weight '$w_{featural}$' to scale the effect of distance to the fovea for feature noise. Therefore, the equation for the observation after adding feature noise at location j given that the eye is focused on location k is as follows,

$$\delta_{featural}(S_t, j) = v[s_t] + N(\theta, \sigma_f(\theta_{jk}, w_{featural}))$$

$$\sigma_{featural}(\theta_{jk}, w_{featural}) = \frac{\theta_{jk}}{(w_{featural})} + c$$

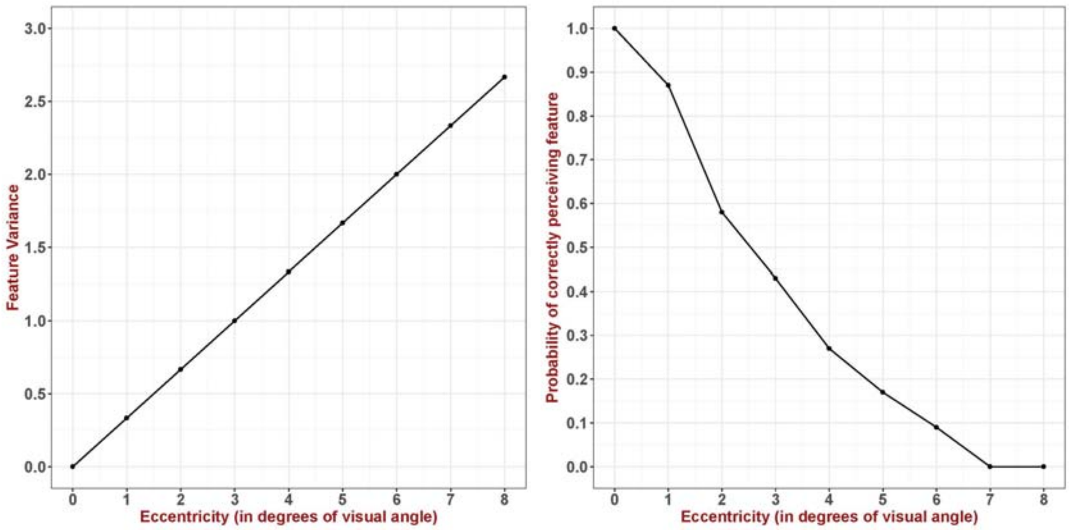
where, $v[s_t]$ is the feature vector as defined in section 6.5.1, c is a constant with value 10^{-4} to avoid 0 variance in the model, $\sigma_f(\theta, w_f)$ is the variance to simulate the degrading eccentricity and 'θ' is the distance between the fixated cell and location j.

2. **Spatial Smearing:** Another source of uncertainty in the human visual system is the localisation error (Levi, 2008), where information in the parafovea may erroneously combine features from one location with adjacent locations. Therefore, for each location in the colour and shape vector, a weighted sum is calculated for the location and its adjacent eight locations. For example, If a red X is surrounded by green Os in the parafovea then, as a consequence of spatial smearing, the participant would be uncertain whether they are actually looking at a red X or a green O. In the model, spatial smearing is represented by a weighting function (Gaussian kernel) with standard deviation as a function of visual angle 'θ' between the fovea and the given location, and a scalar weight '$w_{spatial}$' to scale the effect of distance to the fovea for spatial noise. The weighting function here is a normalised function. As 'θ' (distance) increases the acuity decreases and the standard deviation of the Gaussian kernel increases, this means that the precept of the item at a given location suffers greater interference from surrounding items. This encoding is done for each location in the display. Thus, the equation for the observation after adding spatial noise at location j given that the target features are at location $S_t \in (1, 2, ..., n)$ and the eye is focused on location k is as follows,

$$\delta_{percept}(S_t, j) = K(s, \sigma_s(\theta_{jk}, w_{spatial})) \times \delta_{featural}(S_t, j)$$

$$\sigma_{spatial}(\theta_{jk}, w_{spatial}) = \frac{\theta_{jk}}{(w_{spatial})} + c$$

where, K is the Gaussian kernel with kernel size 3x3, $\sigma_s(\theta_{jk}, w_s)$ is the variance. $\delta_{percept}(S_t, j)$ is calculated separately for both shape and colour feature vectors. c is a constant with value 10^{-4} to avoid 0 variance in the model.



(a)

(b)

Figure 6.3: (a) acuity function represented using a linear variance model with $w_{featural}$ = 3. (b) plot for acuity drop using the linear variance model.

An important thing to note here is, even though the variances $\sigma_f(\theta, w_f)$ is generated using a linear function (see Figure 6.3 (a)). The decline in acuity as a consequence of using the linear variance function is still hyperbolic (see Figure 6.3 (b))(Strasburger et al., 2011).

Now each percept ($\delta_{percept}$) (one for colour and one for shape) is represented as a vector of noisy observations for each location. A consequence of introducing the noise is uncertainty in the content of the location. The extreme values, <= 0.0 or and >= 1.0, represent strong evidence that the feature is either red or green or O or X, while a value of 0.5 represents the absence of evidence in favour of either feature value.

6.5.5 State Estimation

At each time step t on which fixation is made the model receives a noisy observation for each location. The observation space consists of a noisy colour and shape feature vector. Now, we combine the colour and shape observation vectors to obtain a scalar score for each location as shown in equation 6.3. This scalar score is defined as a *relevance score* in the model. The *relevance score* for each cell is a Euclidean distance between the target features, i.e., $[1, 1]$ and the observed colour and shape observations for each location (6.3). The *relevance score* is capped between 0 and 1 where a value 1 represents high relevance to task and value 0 represents low relevance to the task.

$$Score_t[i,j] = min(max(0, \sqrt{(1-\delta_{colour}[i,j])^2 + (1-\delta_{shape}[i,j])^2}), 1.0) \tag{6.2}$$

This vector of *relevance score* is input to a recurrent neural network which integrates information across fixation. In this model, by using a recurrent neural network, we update the underlying hidden state by maintaining a summarised history of previous observations (see section 3.6.3).

6.5.6 Model Learning

Model Architecture

The model architecture consists of the 36-element *relevance score* vector. This is then forwarded as an input to the recurrent neural network with a network size of 36 and tanh activation. The output from the network is a 36-element vector which we denote as the underlying hidden state. The hidden state is then mapped to a feed-forward neural network (layer2) of size 36 nodes and a sigmoid activation function. The output of the layer2 is mapped to a feed-forward layer of 38 nodes and soft-max activation function. This network outputs the parameterised policy distribution. Also, the layer2 is mapped to another feed-forward network with a single node and linear activation that approximates the value of the given state. The model uses Adam Optimisation with a learning rate of 0.0001 to train the network with 4 worker nodes. The hidden layer weights were initialised using Xavier initialisation (Glorot and Bengio, 2010).

Policy Learning

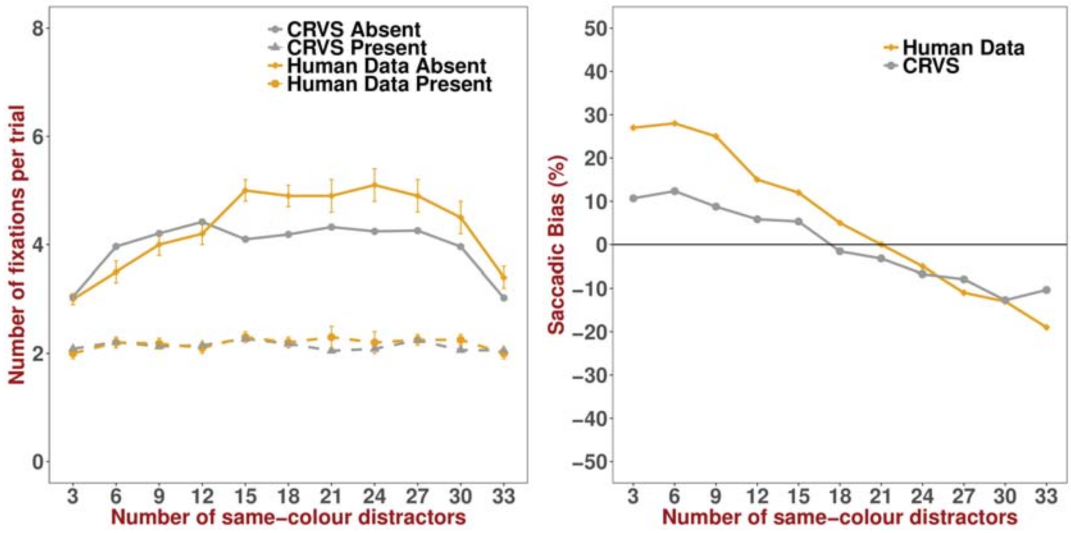
As described above, at each point in time, the model observes the external display through a noisy percept with a high-resolution fovea and low-resolution parafovea and receives an observation (section 6.5.4). The model then extracts the high-resolution local information from the environment by taking actions (section 6.5.2) to move the fovea (e.g., choose where to move the fovea). Since the environment is only partially observed the model needs to integrate information over time in order to determine how to act and how to make eye movements most effectively. It does this using the Recurrent network as state estimator described above. At each step, the model receives a scalar reward (section 6.5.3) (which depends on the action taken by the agent), and the goal of the agent is to learn policies/strategies that maximise the total sum of such rewards $R = E[\sum \gamma^{t-1} r^t]$ where $\gamma \in (0, 1)$ is the discount factor.

Here, the model learns the optimal control policy by using one type of reinforcement learning algorithm, i.e., the policy gradient algorithm (for detail refer to section 3.8.1).

6.5.7 Results

The search performance of the CRVS model with recurrent update is reported and compared to the human performance for the Distractor-Ratio task. The results reported here are from a model that learns the approximately optimal policy through training and was used to generate the behaviour (last 50,000 trials of the simulation).

Plots of fixation frequency versus same colour distractor-ratio for the best fit model is shown in Figure 6.4 (a). The results show that the model generates similar distractor ratio curves to humans (Figure 6.2) for target absent, where more fixations are required for ratios close to 1. The RMSEs for the model $RMSE = 0.40$ and the goodness of fit against Human performance for the model was $R^2 = 0.92$. The model here predicts the invert-U curve or the distractor ratio effect only in the absent condition and a flatter curve in target present condition. This predicted behaviour is consistent with human behaviour.



(a) (b)

Figure 6.4: (a) Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for the CRVS model with recurrent update. (b) Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target-absent trials for the CRVS model with recurrent update. Noise parameters used: Feature noise weight = 18, Spatial noise weight = 4.

The saccadic bias effect is shown in Figure 6.4 (b). The results show that the model chose to make eye movements towards colour or shape distractors when they were in minority set. The RMSEs for the Naive Bayes model $RMSE = 9.41$ and the goodness of fit against Human performance for the model was $R^2 = 0.95$. A weakness in the model was it could not predict the magnitude of saccadic selectivity shown by participants.

6.6 CRVS model with Naive Bayes update

6.6.1 Model Environment

The environment details used in this model is same as described in section 6.5.1, section 6.5.2, section 6.5.3, section 6.5.4

6.6.2 State Estimation

In the model, we combine the colour and shape observation vectors to obtain a scalar score for each location. This scalar score is defined as a *relevance score* in the model. The *relevance score* for each cell is a Euclidean distance between the target features, i.e., $[1, 1]$ and the observed colour and shape observations for each location (6.3). The *relevance score* is capped between 0 and 1 where a value 1 represents high relevance to task and value 0 represents low relevance to the task.

$$Score_t[i, j] = min(max(0, \sqrt{(1 - \delta_{colour}[i, j])^2 + (1 - \delta_{shape}[i, j])^2}), 1.0) \tag{6.3}$$

This combined score is then used as a state vector s_t. The state s_t is updated on each fixation using the Naive Bayes method described in section 3.6.2.

Here, in this model each cell in the state vector s_t is parameterised by representing it with a mean score value $Score_t[i, j]$ and variance $\sigma_t[i, j]$ (which represents the uncertainty). On each time-step the previous score and variance is maintained in the model ($Score_{t-1}[i, j]$ and $\sigma_{t-1}[i, j]$). On receiving a new observation, i.e., $Score_t[i, j]$ and variance $\sigma_t[i, j]$ (here variance is the eccentricity) the model uses the Naive Bayes method (section 3.6.2) to update the current estimates of the score and variance.

6.6.3 Model Learning

Model Architecture

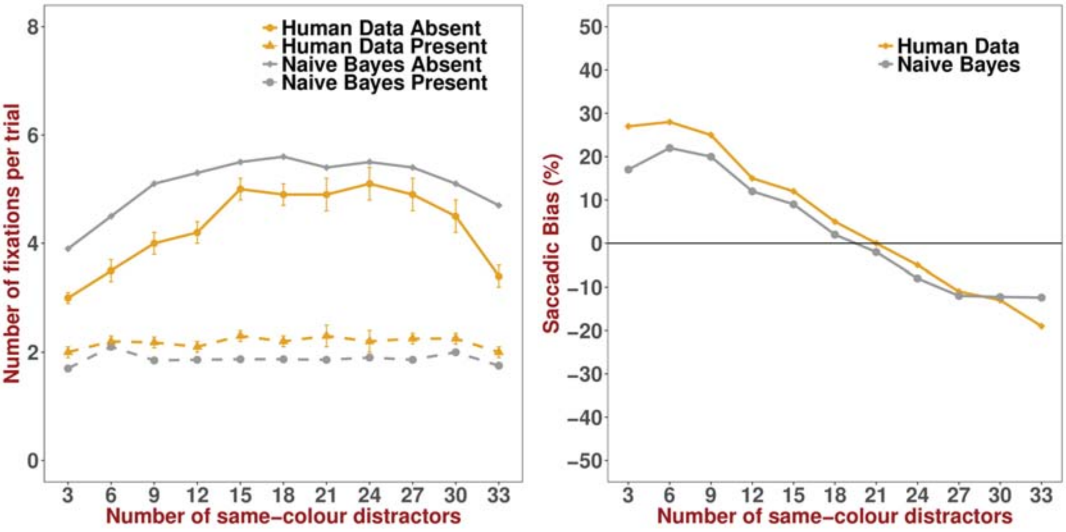
The relevance score estimates $Score_t$ (36 element vector) from the state estimator (above) was taken as the input. This input was connected to a fully connected hidden layer consisting of nodes equivalent to the number of elements in the display, i.e., 36, with rectifier activation function. This is followed by a second fully connected hidden layer consisting of again nodes equivalent to the number of elements in the display, i.e., 36, with sigmoid activation function. Finally, the output layer was a fully connected linear layer of 38 nodes with a single output for each action in the task. To avoid over-fitting of the network $l2$ regularisation of the weights was applied with value 10^{-5}.

Policy Learning

Here, the model learns the optimal control policy by using one type of reinforcement learning algorithm explained in section-3.8.1 (Mnih et al., 2016).

6.6.4 Results

The search performance of the CRVS model with Naive Bayes update is reported and compared to the human performance for the Distractor Ratio task. The results reported here are from a model that learns the optimal policy (reward value per trial converges) through training and was used to generate the behaviour (last 50,000 trials of the simulation).



(a) (b)

Figure 6.5: (a) Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for the CRVS model with Naive Bayes update. (b) Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target-absent trials for the CRVS model with Naive Bayes update. Noise parameters used: Feature noise colour weight = 14, Feature noise shape weight = 13 and Spatial noise weight = 4.

Plots of fixation frequency versus same colour distractor-ratio for the best fit model is shown in Figure 6.5 (a). The results show that the Naive Bayes model generates similar distractor ratio curves to humans (Figure 6.2) for target absent, where more fixations are required for ratios close to 1. The RMSEs for the Naive Bayes model $RMSE = 0.98$ and the goodness of fit against Human performance for the model was $R^2 = 0.92$. The model here predicts the invert-U curve or the distractor ratio effect only in the absent condition and a flatter curve in target present condition. This predicted behaviour is consistent with human behaviour. A weakness in the model is the magnitude of fixations it predicts for the target absent condition.

The saccadic bias effect is shown in Figure 6.5 (b). The results show that the Naive Bayes model chose to make eye movements towards colour or shape distractors when they were in minority set. The RMSEs for the Naive Bayes model $RMSE = 6.64$ and the goodness of fit against Human performance for the model was $R^2 = 0.96$. The saccadic selectivity predicted by the model is consistent with the strategy used by people in the distractor ratio task.

6.7 Heuristic Model for Distractor Ratio Task

6.7.1 Model Environment

The environment details used in this model is same as described in section 6.5.1, section 6.5.2, section 6.5.3, section 6.5.4

6.7.2 State Estimation

The state estimation function used in this model is same as described in section 6.6.2

6.7.3 Model Learning

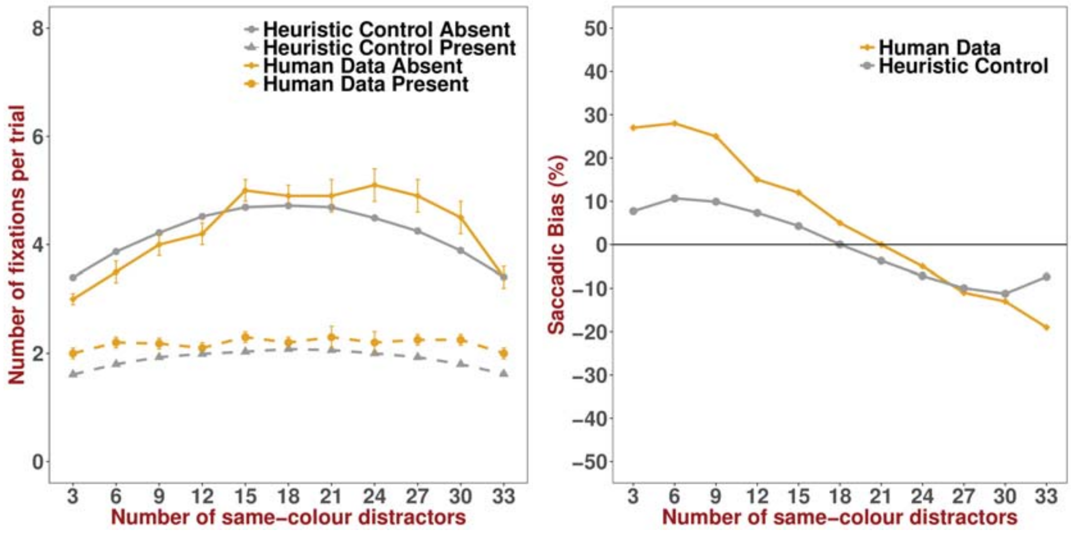
The Heuristic model begins each trial with a prior score for each location as 0.5. Once a random display is presented to the model, it begins by fixating on a location (the first fixation is random) and obtains a noisy observation. It then integrates the noisy observation with previously acquired information from the trial (as described in section 6.6.2). It then uses the integrated state vector to choose the next action to take. These choices in the model are made using decision variables (target present or a target absent). If the decision variables are greater or less than the threshold, then the model responds appropriately. If neither decision variables reach the required threshold, then the model selects a new location to fixate. The next location to fixate is made using a look for best strategy (max rule),i.e., the model selects the location with highest *relevance score* to fixate.

6.7.4 Results

The search performance of the Heuristic model is reported and compared to the human performance for the Distractor Ratio task. The Heuristic control model was run for 50,000 trials, and 10 regression runs to check for consistency.

Plots of fixation frequency versus same colour distractor-ratio for the best fit model is shown in Figure 6.6 (a). The results show that the Heuristic Control model generate similar distractor ratio curves to humans (Figure 6.2) for target absent, where more fixations are required for ratios close to 1. The RMSEs for the Heuristic Control model $RMSE = 0.35$, and the goodness of fit against Human performance for the model was R^2 = 0.93. The Heuristic Control model produced the shape and magnitude of the distractor ratio curve for target absent similar to humans. However, a weakness of the Heuristic control model was that it produced DR effects for both target present and target absent.

The saccadic bias effect is shown in Figure 6.6 (b). The results show that the Heuristic Control model chose to make eye movements towards colour or shape distractors when they were in minority set. The RMSEs for the Heuristic Control model $RMSE = 10.45$,



(a)

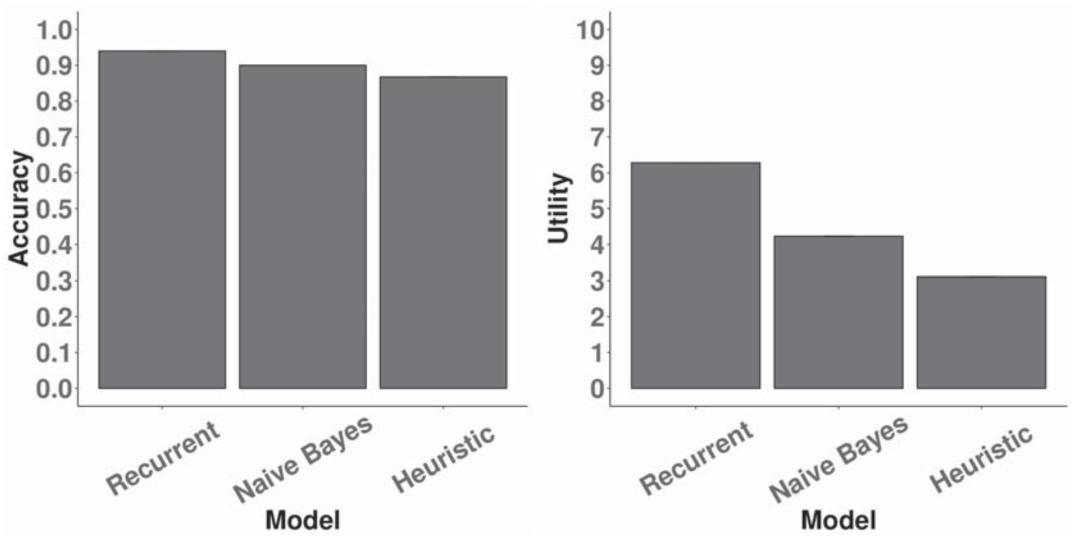
(b)

Figure 6.6: (a) Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for Heuristic Control Model. (b) Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target-absent trials for Heuristic Control Model. Noise parameters used: Feature noise weight $= 6$*, Spatial noise weight* $= 8$*, Absent Threshold = 0.35 and Present Threshold = 0.90.*

and the goodness of fit against Human performance for the model was $R^2 = 0.92$. The saccadic selectivity predicted by the Heuristic Control model is not consistent with the strategy used by people in the distractor ratio task. The Heuristic Control model produces a small saccadic bias which is not a characteristic of human performance.

6.7.5 Model Comparison

In this section we compare the performance of the CRVS model with recurrent update, the CRVS model with Naive Bayes update and the heuristic control model on the Distractor-Ratio task.



(a)

(b)

Figure 6.7: (a) Mean accuracy achieved by the three best fit models (CRVS Recurrent update model, CRVS Naive Bayes update model and Heuristic model). (b) Mean utility gained by the three best fit models (CRVS Recurrent update model, CRVS Naive Bayes update model and Heuristic model).

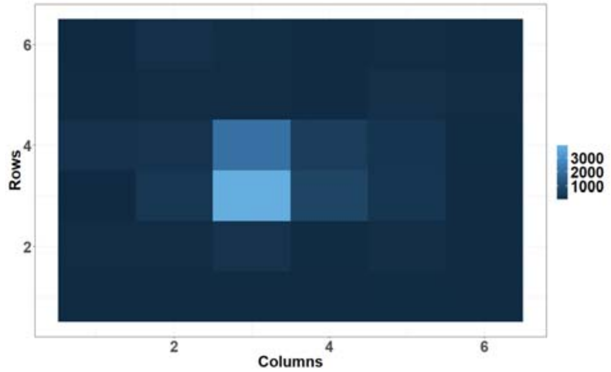


Figure 6.8: plot shows frequency distribution of the first fixation over the grid like display made by the CRVS model. The distribution shows that majority of the first fixation landed in the centre of the display.

CRVS vs Heuristic model

The results from the simulation shows that the both the CRVS and Heuristic control models were able to generate the distractor ratio curves in the target absent condition. However, the Heuristic model produced DR effects for both target present and target absent. Also, the magnitude of the saccadic selectivity was better captured by the CRVS model in comparison to the Heuristic model.

Our results show that the look for best heuristic strategy does not explain the saccadic selectivity as shown by people. These findings are consistent with the analysis reported by Najemnik and Geisler (2008), that the people use more sophisticated strategies than a simple look for best.

Another effect generated by the CRVS model is the bais towards the centre of screen on the first fixation (see Figure-6.8). A characteristic which is also shown by people Tatler (2007). The CRVS tends to produce this effect as an adaptation to the decay in acuity, so as to maximise information gain on the first fixation. In contrast, the heuristic model requires such strategy to be externally encoded.

CRVS Naive Bayes vs Recurrent Update

Both the Naive Bayes and the Recurrent CRVS model were able to generate the distractor ratio effect and the saccadic selectivity. However, the magnitude of saccadic selectivity was better explained by the naive Bayes model. Our analysis shows that this lower saccadic selectivity in the recurrent CRVS model can be attributed to the state estimation used by the model. In the recurrent model, the states are defined as the noisy relevance score estimate. The relevance score uses an euclidean distance metric as a similarity score between the target features and the observed features. Due to this metric, information is lost regarding which feature contributed more, whether colour feature lead to high relevance score or shape feature. For example, the relevance score for observed features [colour = 0.6, shape = 0.4] and [colour = 0.4, shape = 0.6] is the same when target feature are defined as [colour = 1.0, shape = 1.0]. In contrast, the Naive Bayes model updates the feature estimate separately for colour and shape and then combines them to give a relevance score. The information loss impacts both the models, however, due to using sub-optimal information to estimate current state the recurrent model suffers the most.

The performance of the model is shown in fig 6.7. The results show that the recurrent update model out-performs the naive bayes update model. The recurrent update model achieves higher utility and accuracy than the naive bayes update model.

6.8 Discussion

In this chapter, we presented a control model for explaining and predicting peoples search behaviour for the distractor-ratio task. Our results show that the saccadic selectivity and the distractor-ratio effect emerges as an approximately optimal adaptation to the constraints imposed by the human visual system specifically, the noise introduced by crowding of neighbouring features in the periphery (feature smearing). Unlike previous work, including Myers et al. (2013), our results are based on a model that makes approximately optimal control decisions to choose fixation locations rather than a model that uses MAP-like heuristics. Furthermore, our results show that the MAP-like heuristic fails to capture the magnitude of saccadic selectivity shown by people in DR-Task. Furthermore, our results show that the distance metric for feature binding used in the model has some information loss. As a consequence of this way of feature binding, the saccadic selectivity is lower than what is shown in human performance.

The model presented was tested against the human performance reported in Shen et al. (2003) for a 36-element distractor-ratio task. Our results explain that people choose where to fixate next so as to reduce the overall uncertainty in the display about the target and distractor item caused by the feature smearing in the periphery. Also, the when to stop decision, that is, to decide if the target is present or absent emerges so as to maximise the trade-off between the accuracy and dwell time.

Achieving these results required two contributions to cognitive modelling. The first is the novel application of POMDPs to the framing of the distractor-ratio problem, further extending the work of Butko and Movellan (2008). The POMDP framing is important because it provides a rigorous basis for exploring the *computationally rational* adaptation of human strategies to known information processing constraints Lewis et al. (2014), Howes et al. (2009). It thereby helps make the crucial link between cognitive mechanism and rationality that supports in-depth explanations of behaviour.

The second contribution is the novel application of Deep Reinforcement-Learning Mnih et al. (2016) to determine the optimal policy given a theory of human visual information processing capacities. The role of reinforcement learning based algorithm has previously been proposed as means of explaining human learning processes Dayan and Daw (2008) and also, as means of deriving rational analyses of what a person should do in particular task Chater (2009). Our work is more aligned with the goals of Chater (2009). The purpose of our reinforcement learner was not to model the step-by-step learning process, but rather to model the rational outcome of the learning process – an approximately optimal adaptation to information processing limits.

A future extension to the work will be to extend the recurrent architecture with a state estimate that uses colour and shape estimate separately rather than a combined estimate. Rather than using a sub-optimal state estimate the alternative may help in further improving the fit to the saccadic selectivity.

In conclusion, we have demonstrated that framing the visual search problem as a POMDP and solving this problem with deep reinforcement learning is a viable approach to explaining effects such as distractor-ratio and saccadic selectivity.

CHAPTER 7

GENERAL DISCUSSION

7.0.1 Main Results

To summarise, the thesis presents a computational theory of visual search and operationalises it using the CRVS model that explains the human search behaviour. The model explains the eye movement strategies as an *emergent* consequence of *ecology*, *reward* and critically the *architecture* defined for the task. The theory was tested against three visual search tasks.

In the Image search task (Chapter 4), the, when to stop strategy, emerged as an adaptation to the structure of the feedback reward (power and linear) and the ecology (skewed distribution of target features in an image in the display). The model assumes that the availability of information declines with eccentricity (also termed as feature noise in this thesis). The model uses this constraint to define the model architecture. The results showed that the model strategy adapted to the change in reward and the density of images where the model took fewer fixation and terminated the search when the reward for images increased linearly as compared to the power-law-like increase. The model also adapted to the gap size between images, where, the model took fewer fixation when the gap between images was small as compared to the large gap between images.

In the Williams' object search task (Chapter 5), the where to look next strategy emerged as an adaptation to the constraints in the human visual system. The model reuses the architecture defined in chapter 4, where the availability of information declines with eccentricity and the rate of decline varies for each low-level feature like colour, shape, size. Also, the model assumed that the information previously perceived decays with time. The results showed that when the colour feature was provided as a pre-cued feature the proportion of fixations landed on target colour was the highest due to high acuity of the colour feature, followed by size and shape feature. Furthermore, a fewer number of fixations were required for the colour feature due to the high acuity. Although the model failed to capture the magnitude of saccadic distance, it still did capture the shape which was consistent which human search performance. This cause of difference was due to the difference in the ecology of the gap between the simulated display and the actual display experience by participants. The model also revisited previously fixated locations which were a function of the memory decay rate.

In the Distractor-Ratio task (Chapter 6), the where to look next and when to stop strategy emerged as an adaptation to the theory of constraints on the architecture and the ecological distribution of the distractor set. The model made the same architectural assumption as in chapter 4. In addition, the model also assumed a smearing perceptual noise due to crowding and similarity of distractors with the target on a single feature dimension. The results showed that the saccadic selectivity and the distractor ratio effect could be explained by the model as a function of the distractor set and the acuity. Furthermore, the model showed that the centre of gravity effect emerged as an adaptation to the acuity function. Where, by fixating at the centre of the display, the para-foveal region is able to gather more information and effectively guide fixations. Also, our results showed that MAP-like strategy is not a characteristic of human behaviour and failed to produce the magnitude of saccadic selectivity shown by people in DR-Task. This finding is consistent with Geisler (2011) study. Where, they also reported that MAP-like heuristic does not capture the human strategy, rather people utilise a more complex strategy.

In all three studies presented in the thesis, the *strategy* (saccadic selectivity, and saccadic distance due to feature sensitivity) emerged as a consequence of the *architecture* defined in the model (the perceptual feature noise, spatial noise, and memory decay), the *reward* (speed accuracy trade-off) defined for the model for each task, and the *ecology* of the task (distribution of target features, distribution of distractor set, and gap between stimulus). Furthermore, in all the three tasks no prior assumptions were made about the strategy space nor were any hand-coded rule or heuristics were defined. Instead a set of theories were defined, and the strategy emerged as an adaptation to those theories.

7.0.2 Machine Learning for Computational Models

In addition to the scientific contribution, a new methodological contribution is also provided where strategies are not hand-coded, rather derived as an adaptation to the theory defined.

The predictive performance of the existing models is impressive. However, their further development is limited by how the strategies are defined to build the models. Cognitive architectures like EPIC (Kieras and Meyer, 1997) and ACT-R (Anderson, 1996) models have required hand-coded rules and the coding of rules not only require domain expertise it is also a significant development work on the application of the models to new tasks. Bayesian models are data intensive or make strong assumptions about the likelihood distribution (e.g., Gaussian distribution) (Butko and Movellan, 2008) which makes them difficult to scale up to real-world tasks efficiently.

Alternatively, with a recent breakthrough in machine learning, cognitive models have utilised machine learning techniques to derive strategies. However, the models have relied upon relatively simple reinforcement learning methods (Sprague and Ballard, 2004, Chen et al., 2015) or classifiers (Li et al., 2018). The efficiency concerns limit the range of tasks to which the models can be applied.

In this thesis, we investigated a more scalable and efficient approach to modelling human behaviour. The model presented utilised a model-free Deep Reinforcement Learning (DRL) algorithm as an optimisation technique to solve the visual search problem given a set of theory about human cognition. The role of reinforcement learning in decision making have previously been proposed as a means of explaining human behaviour (Dayan and Daw, 2008). Also, as a means of deriving optimal strategy of what a person should do in particular task (Chater, 2009). The studies presented in this thesis is more aligned to Chater (2009). The purpose of our reinforcement learner was not to model the step-by-step learning process, but rather to model the optimal outcome of the learning process – an approximately optimal adaptation to the theory of constraints.

The model-free reinforcement learning works by ignoring the underlying environment dynamics (model) and utilise only the past experience to take future actions. The focus is on answering what is the best action to take given only the current observation. This assumption is especially useful for finding optimal policies when the underlying system is complex to model. In this thesis, the environment used is static, and the saccadic movements are assumed to be without any errors. In other words, the transition probability of moving from one state to next $(P(s_(t+1)|s_t, a))$ is 1. Due to this nature of the environment, the model-free reinforcement learning is an appropriate algorithm to find the optimal policy.

An optimisation algorithm was used for the assumption that the behaviour emerges as an approximately optimal adaptation to the theory of the architecture, reward and ecology. The optimality here is assumed to derive a causal relationship between the theory and the emergent behaviour.

In this thesis, we formulated the visual search problem as a POMDP, and used a deep reinforcement learning algorithm to solve the POMDP problem, we have shown that the approach can scale to everyday tasks that require a visual search. The POMDP framework breaks down the problem into two sub-tasks of *state estimation* and *controller*. The *state estimation* is a one-to-one mapping of current environment state and the *controller* then uses the state to learn how to act in the environment. The role of *state estimation* has previously been attributed to Orbitofrontal cortex (OFC) that maintains the current environment state (Wilson et al., 2014) for a given task. Also, integrating information from cortical and subcortical areas, with information from memory to update current state in partially observable setting (Wilson et al., 2014). In this thesis, through the three visual search tasks, we investigated three different state estimation mechanisms. Our intention here was not to vary the theory of how people maintain or update information across fixation. Rather, to test the efficiency of the approaches.

In the Image search task (Chapter 4), the Bayes optimal and recurrent network approaches were investigated for updating information across fixations. Our results showed that both the state estimation techniques were able to explain the change in stopping behaviour with the change in item density and the reward. However, the Bayes state estimation approach was difficult to scale up to the actual task size in comparison to the recurrent approach.

In the Distractor-Ratio task (Chapter 6), the Naive Bayes and the recurrent network approaches were investigated for updating information across fixations. Our results showed that both the techniques were able to explain the distractor ratio effect and saccadic selectivity. However, the Naive Bayes approach imposed a strong constraint on the noise model to be used, i.e., the noise model needs to be linear and Gaussian. This constraint might not be feasible to model real-world problems that are more often non-linear in nature. Furthermore, the CRVS model with recurrent update achieved higher accuracy and utility as compared to the CRVS model with Naive Bayes update.

In the studies presented above, the recurrent architecture with deep reinforcement learning was shown to be a more scalable and efficient solution to modelling human behaviour. However, decoding what representation a recurrent network has learnt is difficult.

The fact that the CRVS model generates human-like strategies using machine learning distinguishes it from approaches which use hand-coded rule (Kieras and Hornof, 2014) and approaches that learn from human examples (Li et al., 2018). It, therefore, represents a progression to models that require fewer inputs from analysts in order to model human cognition. Also, the model represents a new application of deep reinforcement learning in modelling human behaviour. Previous research has focused on the application of learning human-level control policies on a variety of different Atari games (Mnih et al., 2015). By defining a deep reinforcement learning algorithm to solve a POMDP, as we have done, we have shown that the approach can scale to everyday tasks that require a visual search. Also, by utilising a recurrent neural network to approximate state estimation, we further extended (Rao, 2010) idea to solve a larger POMDP problem, and addressed some of the high computational cost associated with POMDPs to provide a more efficient approach to the modelling problem.

7.0.3 Further Contributions to Cognitive Modelling

Through the studies reported, the thesis makes a contribution to the field of cognitive science by presenting a new computational model that further explains the role of constraints, ecology and reward in understanding human behaviour and decision making. This work extends the current practice of computational modelling by (a) by showing how to operationalise computational rationality to model human behaviour, (b) extend deep reinforcement learning with recurrent neural network to explain visual search, and (c) explaining behaviour as bounded optimal and focus on the role of constraints on the behaviour.

Research on human behaviour has often focused on explaining behaviour as either being optimal (Anderson, 1991, Chater and Oaksford, 1999, Geisler, 2011) or sub-optimal (Rahnev and Denison, 2018). This thesis claims that behaviour is bounded optimal. The assumption of bounded optimality (computational rationality) in the computational model is crucial because it enables models to be predictive and explanatory (Howes et al., 2009, Lewis et al., 2014, Howes et al., 2016, Acharya et al., 2017). By selecting a strategy through an optimisation algorithm, it allows a causal relationship between the theoretical assumption made in the model and the resulting behaviour that emerged. Furthermore, the strategy derived is also predictive. Here, the success of the model in predicting behaviour does not lead to the conclusion of behaviour being optimal or sub-optimal. Instead, it provides evidence in favour of the theory of the constraints, ecology and rewards used.

The visual constraints assumed in this thesis worked well to predict users search behaviour to locate target items in the display. The model was able to address what information can be perceived on each fixation by showing that the difference in the effective field of view for the low-level feature (colour, shape and size) helps to predict how quickly participants were able to find the target. An important prediction of the model was to explain the fixation proportions (where to fixate next). The model was able to address this question by showing that the difficulty in discriminating features in the periphery and the ecology of the task plays a critical role in explaining and predicting the emergent behaviour. For example, in chapter 6 the spatial noise made it difficult for the model to discriminate features and thereby adopted a strategy to guide fixations towards the minority set.

The model-free reinforcement learning used to explain search behaviour suggests that the people may utilise a trial-and-error approach to perform goal-directed tasks. These models operate by caching rewards/values accumulated through repeated interaction with the environment. In other words, the decision is made by estimating future reward based on the rewards that have been encountered in the past. For example, in chapter 4 the model learned when to terminate searching such that it maximised its rewards. In chapter 5 and 6 the model learned where to move next so that it minimised the search time.

Furthermore, the model prediction explains how the strategic choices that people make are affected by the task ecology, information processing constraints and user priorities described in the form of rewards. For example, in chapter 4 small changes in the visual angle between items and changes in the reward structure that reflects user preference causes qualitative changes in strategy. In chapter 6, smearing of information due to crowding effect led to saccadic selectivity towards target features. By describing cognitive models as an adaptation to the underlying theory, enables the model to explain behaviour, rather than merely describing them. Furthermore, the model provides a framework to test the theory of constraints that led to observed behaviour.

To summarise, the model presented in the thesis shows that the human eye movement behaviour can be explained as computationally rational strategy that adapts to both the visual and cognitive influence. Both factors need to be taken into account for a theory to explain how eye movements operate in the real world. For example, we saw that the eye movement are dependent on the information processing constraints, user expectations, memory and the constraints imposed by the structure of the visual display.

7.1 Future Work

Through this thesis, the CRVS model we presented was able to generate human-like search behaviour in 3 different visual search tasks. Though the progress made in this thesis is a step forward, it does so for visual search tasks. This section discusses some of the possible extensions to the work.

7.1.1 Integrated Model for Human behaviour

The model reported in this thesis interacts with its environment by converting rich visual information into a symbolic representation. With the complex visual layouts of todays displays, converting to a symbolic representation may prove to be a challenge. A possible extension would be to use Convolutional Neural Networks (CNN) to process raw images as an input to the model rather than symbolic values. Recent, breakthrough in CNN's capability of extracting meaningful information has led to high performing object detection classifiers (Girshick, 2015, Dai et al., 2016, Liu et al., 2016). The model will extend its input layer to incorporate a CNN layer to extract task-relevant information given a foveated image input and then use the information to learn the strategy. Xu et al. (2015) have previously used attention-based learning to generate image caption. With the newer displays being more visually rich and complex a CNN extension will help the model to be applied to more real work tasks. The first test task would be to redo the modelling of image search task (Chapter 4) with a CNN layer.

Another extension to the model will be to incorporate a motor control module. Much of the real-world interaction involves both the eye movement for localising or feedback and a corresponding motor movement. For example, driving, typing text, cooking or grabbing a cup on the table. A unified model of perceptual-motor control will be rich enough to answer questions like how attention and motor control is coordinated. For example, in the chapter 4 the model was missing the motor control model of mouse movement. Future work will be to see if a model of mouse movement recovers the additional search time. It will further validate the novel hypothesis explored in this thesis with a complex architecture where the model has to adapt to both the perceptual noise and the motor control noise.

7.1.2 Optimal Control versus Approximately Optimal Control

One potential direction of research is to look at the consequence of approximation to visual search behaviour. While many researchers have advocated that human search behaviour can be explained as Bayes optimal (Najemnik and Geisler, 2005, 2008, Butko and Movellan, 2008, Myers et al., 2013, Nunez-Varela and Wyatt, 2013). Where the research suggests people maintain an optimal state estimate of the world and use those estimates to act in the world optimally. However, the work presented in the thesis assumes that finding an optimal policy to the real-world problem is biologically intractable, and some form of approximations needs to be applied to solve these problems. For example, in chapter 4 the Bayesian model was not tractable to the full 6x6 grid display.

Approximations are especially necessary because in real-world information received is high-dimensional, continuous and partial. This implies that the data received is sparse and situations may or may not be encountered multiple times to learn best decisions to make. A consequence of approximation is the soft/no guarantees of convergence to global optimal solution. This may lead to a model learning local optimal solution for some problems. Further work needs to be done to understand the consequence of approximation to the underlying theory.

7.1.3 Model Constraints

Fixation Duration

The question of when do people move the gaze from one location to next is frequently asked in visual search (Rayner, 1998). In this thesis, the three studies presented doesn't answer this question directly. In chapter 4 for the image search task the fixation duration was set from the reported human data. In chapter 5 and 6 the fixation duration was set to 275ms. Visual search literature suggests that in high-speed visual search tasks fixation duration range from 250-300ms (Rayner and Castelhano, 2008). Previous research has given some insight into what factors effect fixation duration (Halverson and Hornof, 2004, Wolfe, 2007, Rayner, 1998). For example, Mackworth (1976) in his study showed that the fixation duration increased as the density of objects increased in the search display. He attributed this to the increase in cognitive demand imposed due to processing more objects in the display. As a future work, the model will be used to explain the increase in fixation duration as a function of object density.

Saccadic Movements

In the model, we assumed that when the gaze is moved from one location to next, they land on the object of attention without any error. Previous research has shown that the saccadic movements are noisy and imperfect (Kowler, 1990). Becker and Fuchs (1969) showed that saccades have a tendency of undershooting rather than overshooting the target location. Also, there is a distinction between where people are attending to and where the eyes move Salvucci (2001). For example, Schilling et al. (1998) showed that participants did not fixate on every word as they read a sentence, yet they could almost perfectly comprehend the sentence. Also, McConkie and Rayner (1975) showed that people are able to process information prior to them fixating on the word by processing it in the parafovea. The CRVS model presented in the thesis does not make the distinction between the deployment of attention using the noisy observation and then making eye movements. Since, it does not capture the imperfect eye movements. Previous work by Salvucci (2001) presented the EMMA model that captures the noisy eye movements. As part of the future work, we will use the constraints encoded in EMMA in our model. We will then use this model to the changes in model behaviour for the Williams object search task presented in chapter 5.

Discount Factor

Human behaviour studies have shown that people have a cognitive bias called 'Hyperbolic discounting' (Vincent, 2016), where they choose small immediate reward rather than a large reward later. This phenomenon is especially seen when the delay is closer to the present than the future. For example, when people are given 2 choices, get a 50 today or 70 in a week. People choose 50 today. In this thesis we did not explore the discounting of reward and its influence on human behaviour. Rather, we assumed based on the experimental instruction that people would maximise the future reward rather than intermediate reward. As part of future work, we need to further explore the discount factor parameter and its influence on model behaviour.

Parameter Exploration

The model presented in the thesis uses a grid search method to find model constraint parameters that best fit human performance. Two issues arise because of manual search: (1) an exhaustive search for finding values that leads to best fit is computationally expensive, (2) using values from literature can introduce additional bias on behaviour that is unwanted. An extension to the current modelling effort would be to use an inverse model to reduce ambiguity and bias due to model parameters that are derived from observed data. For example, Kangasrääsiö et al. (2017) used an Approximate Bayesian Computation (ABC) inference model to estimate model parameters from experiment data. Specifically, they were able to derive fixation duration and recall probability parameters for the model to improve prediction performance. As an extension, we will use the ABC model to derive fixation duration for the model in chapter 5. Furthermore, we will explore whether noise parameters can be derived from observed data.

7.2 Conclusion

The thesis demonstrates that a computationally rational strategy can explain and predict peoples' visual search behaviour. The CRVS model presented in this thesis derive strategy as an approximately optimal adaptation to the constraints imposed by the human visual system, the ecology of the task and the task rewards. In doing so, the model avoided any heuristic assumptions about the strategy. Three visual search task presented in this thesis supported the hypothesis.

APPENDIX A

GENERALISED ADVANTAGE ESTIMATE

The chapter explains the Generalised Advantage Estimate used in the thesis in section 3.8.1.

In the Generalised Advantage Estimate (GAE), the cumulative reward is described as Monte-Carlo return eq-A.1.

$$R_t = \sum_{k=0}^{T-t} \gamma^k r_{t+k} \tag{A.1}$$

Here, R_t is an unbaised expected return of future rewards in a given state. But, in a stochastic environment, the r_{t+k} is a random variable, and sum over rewards can have high variance. However, by using a n-step return that is represented by a value function V(s) that approximates the sum of return from remaining steps can be used as a low variance estimator.

$$R_t^{(n)} = \sum_{k=0}^{n-1} \gamma^k r_{t+k} + \gamma^n V(s_{t+n}) \tag{A.2}$$

Where, $R_t^{(1)}$ represents a 1-step look ahead commonly refered to Q-learning (section-3.7.1) and $R_t^{(\infty)}$ represents the Monte-Carlo return over the entire trajectory. Here, n-step acts as a trade-off between bais and variance, where, $R_t^{(\infty)}$ gives an unbiased but high variance estimator and $R_t^{(1)}$ gives a biased but low variance estimator.

Alternatively, a λ-return (Sutton and Barto, 1998) method is used for bias-variance trade-off, using a exponentially-weighted average function to estimate the value of state with λ as a decay parameter.

$$R_t(\lambda) = (1 - \lambda) \sum_{n=1}^{T} \lambda^{(n-1)} R_t^{(n)} \tag{A.3}$$

Where, $\lambda = 0$ refers to a 1-step look ahead similar to $R_t^{(1)}$, and $\lambda = 1$ refers to Monte-Carlo return $R_t^{(\infty)}$. Intermediate values are used balance the bias and variance trade-off for value estimation. Empirically, values between $[0.9, 0.99]$ works best (Schulman et al., 2015). In this thesis we set the value at 0.98.

Using the λ-return to update the value function results in a temporal difference algorithm (Sutton and Barto, 1998). Also, estimating the advantage function (eq-A.4) using the λ-return derives the generalized advantage estimator (Schulman et al., 2015).

$$A^t = R_t(\lambda) - V(s_t) \tag{A.4}$$

$$\nabla J(\theta) = \underset{\theta}{\overset{\pi}{\mathbb{E}}} \left[\nabla \sum_{t=1}^{T} logP_\theta(s, a) A^t \right] \tag{A.5}$$

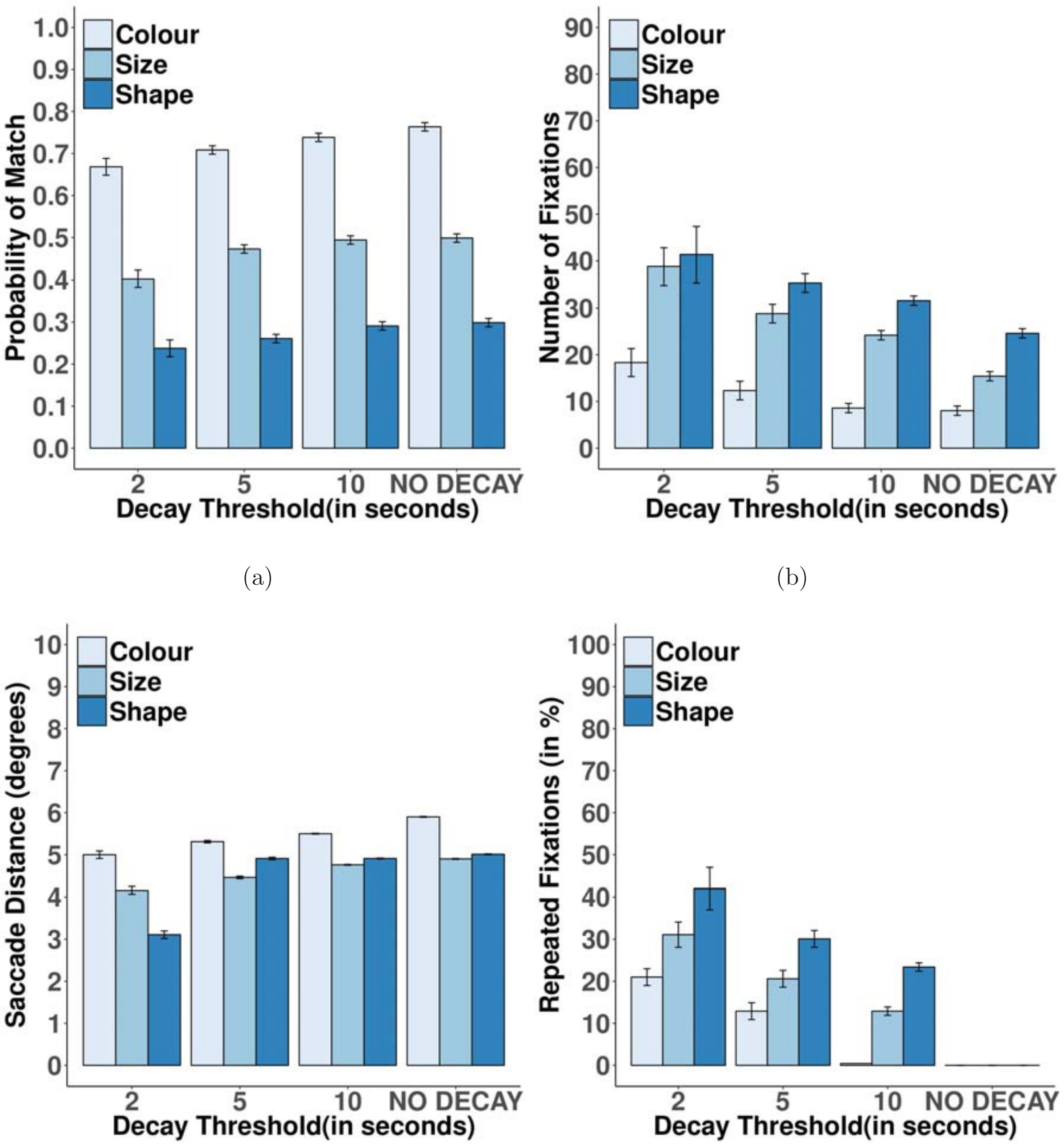
Empirically, the λ-return better performance than the n-step return (Schulman et al., 2015).

APPENDIX B

WILLIAM OBJECT SEARCH

B.1 Introduction

The chapter shows the model performance for the explored parameter space. Here, we highlight the model search performance for different decay threshold. Figure B.1 shows that the proportion of fixations landing on target items decreases as the models' ability to retain information in the working memory decreases. The number of fixations increases as the as the models' ability to retain information in working memory decreases. This is due to the model re-fixating on the previously visited object as seen in Figure B.1 (d). The model also adapts its saccadic distance as a function of decay where smaller jumps were made for para-fovea to be useful when the duration of information retention was small.



(a)

(b)

(c)

(d)

Figure B.1: Plots showing the search performance as a function of memory decay when the gap between objects was sampled from a uniform distribution over [0.0, 1.0, 2.0, 3.0]. (a) Proportion of fixation landing on target cue. (b) Average number of fixations per trial for each target cue. (c) Saccadic distance for each target cue. (d) Repeated fixations observed per trial for each target cue.

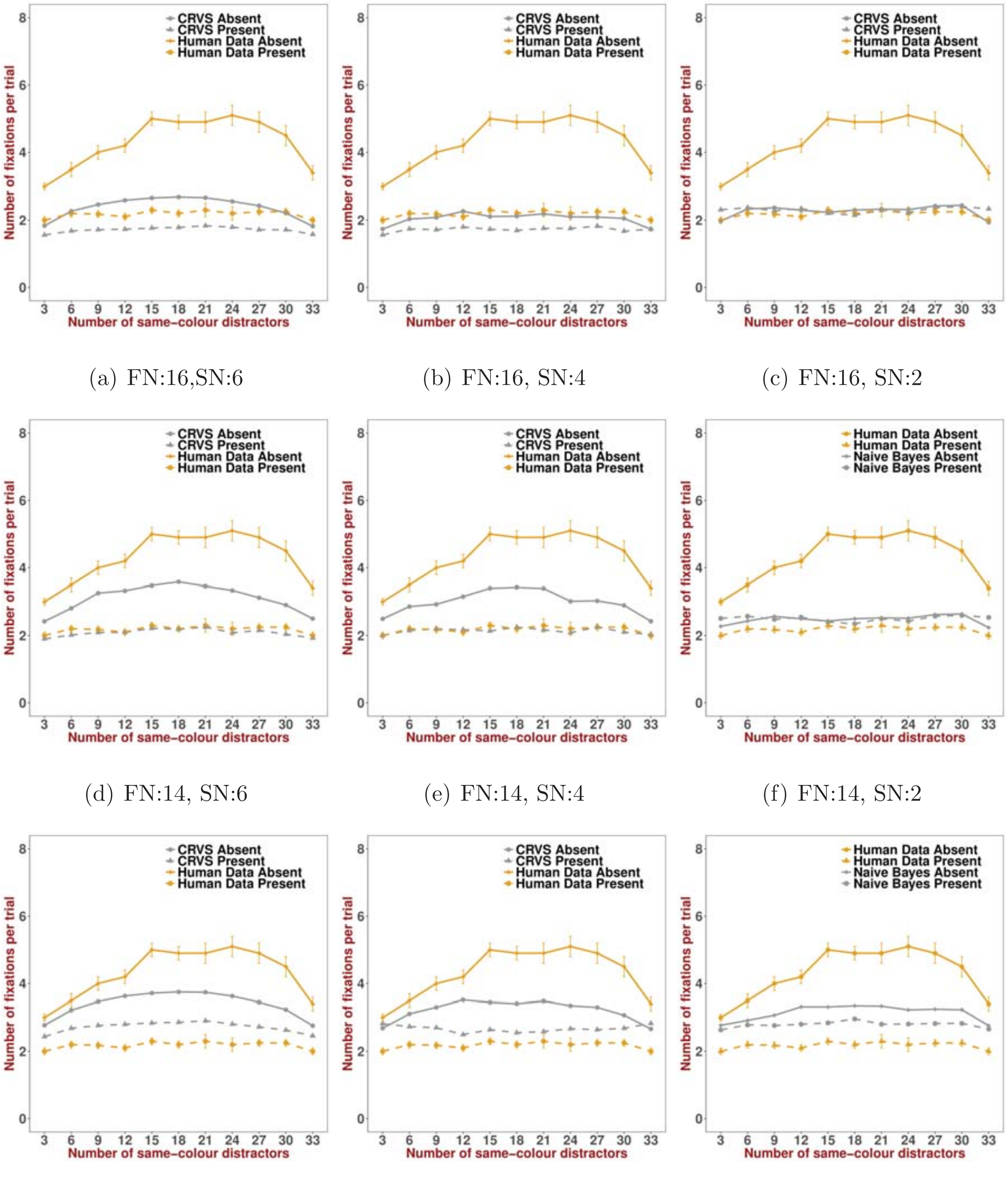
APPENDIX C

DISTRACTOR-RATIO TASK

C.1 Introduction

The chapter shows the CRVS model and the Heuristic model performance for the explored parameter space. Here the results show the effect of feature noise and spatial noise on the search behaviour. In both models, as feature noise increases the number of fixations required to respond target present or absent increases. Furthermore, the spatial noise effects the shape of the distractor ratio curve and the saccadic selectivity. As the spatial noise is increased the saccadic selectivity increases, but, as more noise is introduced the saccadic selectivity reduces.

C.1.1 Parameters explored for CRVS Model



(a) FN:16,SN:6

(b) FN:16, SN:4

(c) FN:16, SN:2

(d) FN:14, SN:6

(e) FN:14, SN:4

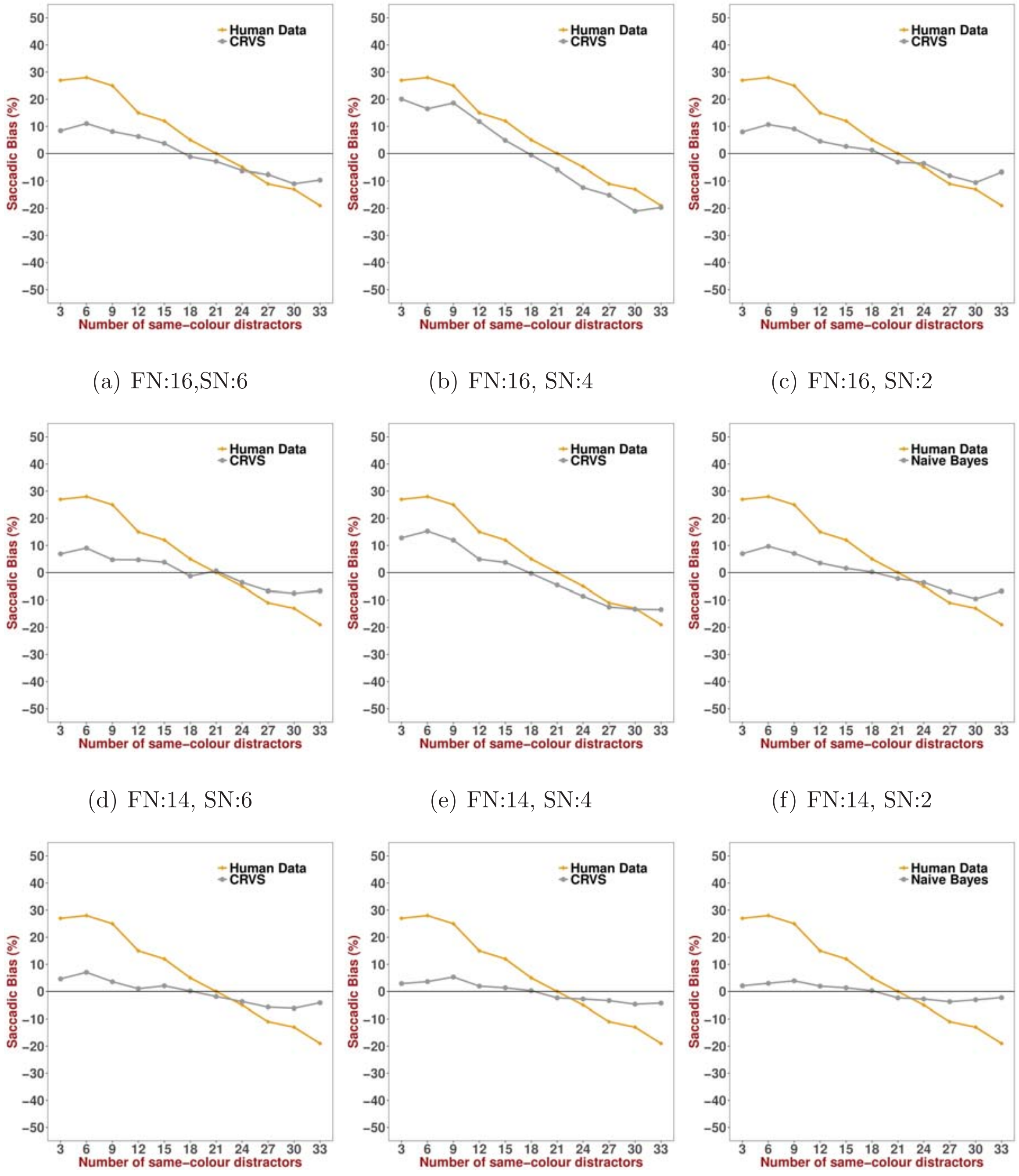
(f) FN:14, SN:2

(g) FN:12, SN:6

(h) FN:12, SN:4

(i) FN:12, SN:2

Figure C.1: Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for CRVS Model. FN is the Feature noise parameter and SN is the spatial noise parameter.



(a) FN:16,SN:6

(b) FN:16, SN:4

(c) FN:16, SN:2

(d) FN:14, SN:6

(e) FN:14, SN:4

(f) FN:14, SN:2

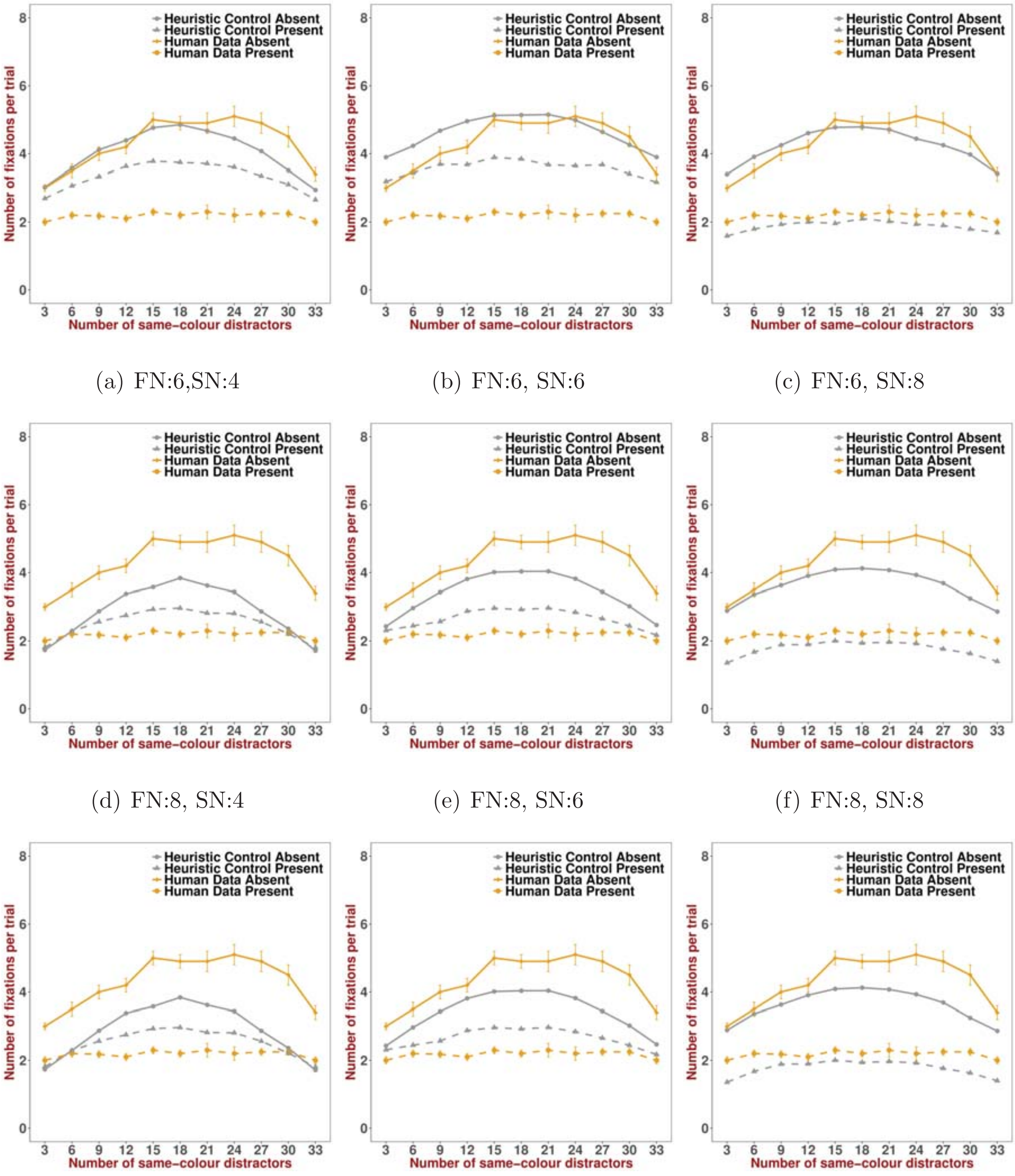
(g) FN:12, SN:6

(h) FN:12, SN:4

(i) FN:12, SN:2

Figure C.2: Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target-absent trials for CRVS Model. FN is the Feature noise parameter and SN is the spatial noise parameter.

C.1.2 Parameters explored for Heuristic Control Model

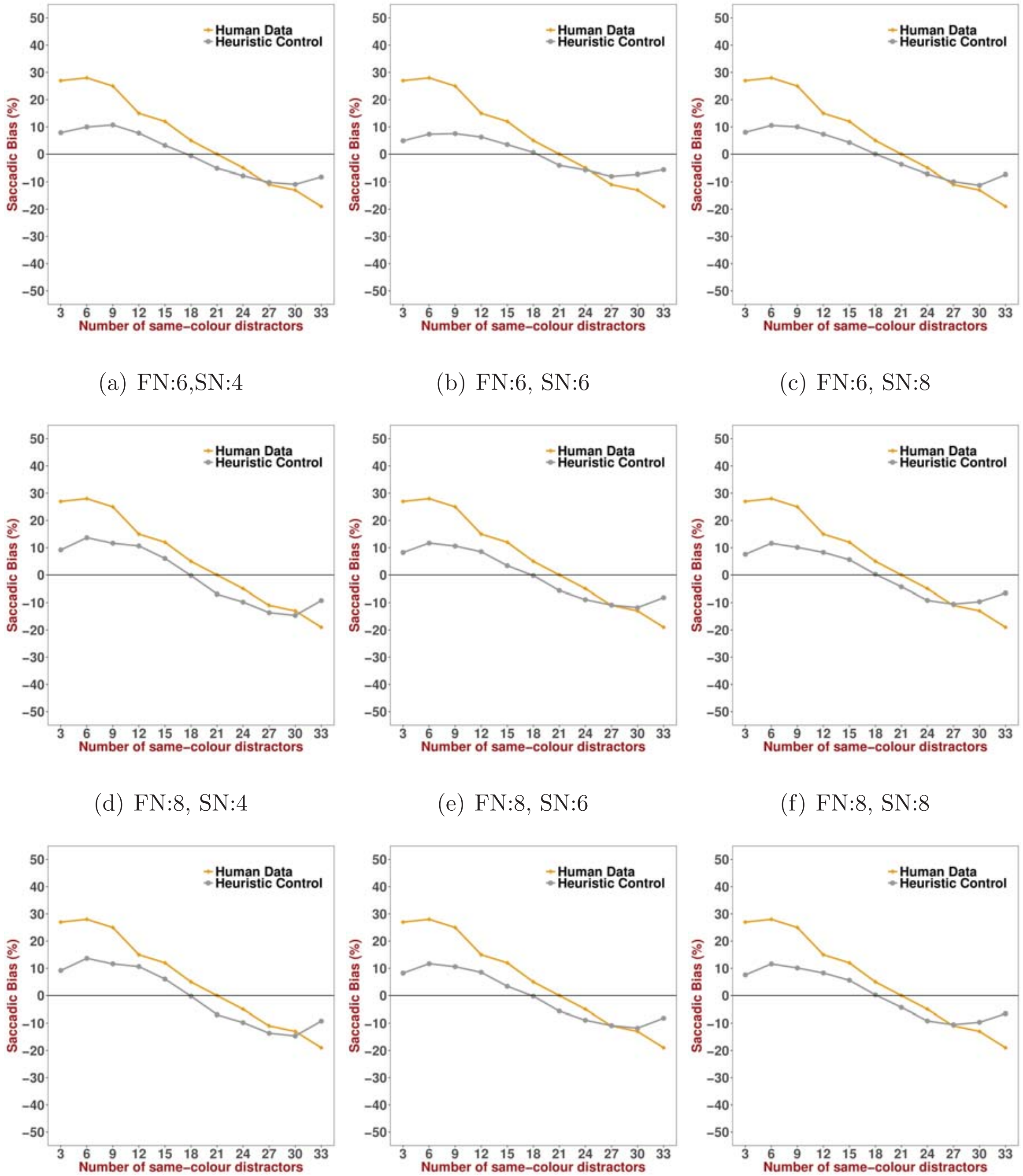


(a) FN:6,SN:4 (b) FN:6, SN:6 (c) FN:6, SN:8

(d) FN:8, SN:4 (e) FN:8, SN:6 (f) FN:8, SN:8

(g) FN:10, SN:4 (h) FN:10, SN:6 (i) FN:10, SN:8

Figure C.3: Average number of fixations per trial as a function of the number of distractors sharing colour with the search target in target-absent trials and target-present trials for Heuristic Control Model. FN is the Feature noise parameter and SN is the spatial noise parameter.



(a) FN:6,SN:4

(b) FN:6, SN:6

(c) FN:6, SN:8

(d) FN:8, SN:4

(e) FN:8, SN:6

(f) FN:8, SN:8

(g) FN:10, SN:4

(h) FN:10, SN:6

(i) FN:10, SN:8

Figure C.4: Saccadic bias (the difference between the observed frequency and chance performance) as a function of the number of same-colour distractors in target-absent trials for Heuristic Control Model. FN is the Feature noise parameter and SN is the spatial noise parameter.

APPENDIX

LIST OF REFERENCES

Acharya, A., Chen, X., Myers, C. W., Lewis, R. L. and Howes, A. (2017), Human visual search as a deep reinforcement learning solution to a pomdp., *in* 'Proceedings of the Annual Meeting of the Cognitive Science Society', pp. 51–56.

Anderson, J. R. (1991), 'Is human cognition adaptive?', *Behavioral and Brain Sciences* **14**(3), 471–485.

Anderson, J. R. (1996), 'Act: A simple theory of complex cognition.', *American Psychologist* **51**(4), 355.

Anderson, J. R., Lebiere, C., Lovett, M. and Reder, L. (1998), 'Act-r: A higher-level account of processing capacity', *Behavioral and Brain Sciences* **21**(6), 831–832.

Awh, E., Belopolsky, A. V. and Theeuwes, J. (2012), 'Top-down versus bottom-up attentional control: A failed theoretical dichotomy', *Trends in cognitive sciences* **16**(8), 437–443.

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

Baloh, R. W., Sills, A. W., Kumley, W. E. and Honrubia, V. (1975), 'Quantitative measurement of saccade amplitude, duration, and velocity', *Neurology* **25**(11), 1065–1065.

Baxter, J. and Bartlett, P. L. (2001), 'Infinite-horizon policy-gradient estimation', *Journal of Artificial Intelligence Research* **15**, 319–350.

Bechtel, W. (2008), 'Mechanisms in cognitive psychology: What are the operations?', *Philosophy of Science* **75**(5), 983–994.

Becker, W. and Fuchs, A. (1969), 'Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points', *Vision research* **9**(10), 1247–1258.

Bellman, R. (1952), 'On the theory of dynamic programming', *Proceedings of the National Academy of Sciences of the United States of America* **38**(8), 716.

Bertera, J. H. and Rayner, K. (2000), 'Eye movements and the span of the effective stimulus in visual search', *Perception & Psychophysics* **62**(3), 576–585.

Borji, A., Sihite, D. N. and Itti, L. (2011), Computational modeling of top-down visual attention in interactive environments., *in* 'BMVC', Vol. 85, pp. 1–12.

Brumby, D. P., Salvucci, D. D. and Howes, A. (2009), Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving, *in* 'Proceedings of the SIGCHI conference on human factors in computing systems', ACM, pp. 1629–1638.

Butko, N. J. and Movellan, J. R. (2008), I-pomdp: An infomax model of eye movement, *in* 'Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on', IEEE, pp. 139–144.

Carrasco, M. (2011), 'Visual attention: The past 25 years', *Vision research* **51**(13), 1484–1525.

Chater, N. (2009), 'Rational and mechanistic perspectives on reinforcement learning', *Cognition* **113**(3), 350–364.

Chater, N. and Oaksford, M. (1999), 'Ten years of the rational analysis of cognition', *Trends in cognitive sciences* **3**(2), 57–65.

Chen, B. and Perona, P. (2014), 'Towards an optimal decision strategy of visual search', *arXiv preprint arXiv:1411.1190* .

Chen, X. (2015), An optimal control approach to testing theories of human information processing constraints.
URL: *http://etheses.bham.ac.uk/5907/*

Chen, X., Bailly, G., Brumby, D. P., Oulasvirta, A. and Howes, A. (2015), The emergence of interactive behavior: A model of rational menu search, *in* 'Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems', ACM, pp. 4217–4226.

Chen, X., Starke, S. D., Baber, C. and Howes, A. (2017), A cognitive model of how people make decisions through interaction with visual displays, *in* 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', ACM, pp. 1205–1216.

Clarke, A. D., Green, P., Chantler, M. J. and Hunt, A. R. (2016), 'Human search for a target on a textured background is consistent with a stochastic model', *Journal of vision* **16**(7), 4–4.

Corbetta, M. and Shulman, G. L. (2002), 'Control of goal-directed and stimulus-driven attention in the brain', *Nature reviews neuroscience* **3**(3), 201.

Dai, J., Li, Y., He, K. and Sun, J. (2016), R-fcn: Object detection via region-based fully convolutional networks, *in* 'Advances in neural information processing systems', pp. 379–387.

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

Della Libera, C. and Chelazzi, L. (2009), 'Learning to attend and to ignore is a matter of gains and losses', *Psychological Science* **20**(6), 778–784.

Desimone, R. and Duncan, J. (1995), 'Neural mechanisms of selective visual attention', *Annual review of neuroscience* **18**(1), 193–222.

Eckstein, M. P. (2011), 'Visual search: A retrospective', *Journal of vision* **11**(5), 14–14.

Eckstein, M. P., Thomas, J. P., Palmer, J. and Shimozaki, S. S. (2000), 'A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays', *Perception & psychophysics* **62**(3), 425–451.

Eckstein, M., Schoonveld, W. and Zhang, S. (2010), 'Optimizing eye movements in search for rewards', *Journal of Vision* **10**(7), 33–33.

Ehinger, K. A. and Wolfe, J. M. (2016), 'When is it time to move to the next map? optimal foraging in guided visual search', *Attention, Perception, & Psychophysics* **78**(7), 2135–2151.

Eriksen, B. A. and Eriksen, C. W. (1974), 'Effects of noise letters upon the identification of a target letter in a nonsearch task', *Perception & psychophysics* **16**(1), 143–149.

Everett, S. P. and Byrne, M. D. (2004), Unintended effects: varying icon spacing changes users' visual search strategy, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', ACM, pp. 695–702.

Faragher, R. (2012), 'Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]', *IEEE Signal processing magazine* **29**(5), 128–132.

Findlay, J. M. and Gilchrist, I. D. (1998), Eye guidance and visual search, *in* 'Eye guidance in reading and scene perception', Elsevier, pp. 295–312.

Findlay, J. M. and Gilchrist, I. D. (2003), *Active vision: The psychology of looking and seeing*, number 37, Oxford University Press.

Friston, K., Adams, R., Perrinet, L. and Breakspear, M. (2012), 'Perceptions as hypotheses: saccades as experiments', *Frontiers in psychology* **3**, 151.

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

Girshick, R. (2015), Fast r-cnn, *in* 'Proceedings of the IEEE international conference on computer vision', pp. 1440–1448.

Glimcher, P. W. (2003), 'The neurobiology of visual-saccadic decision making', *Annual review of neuroscience* **26**(1), 133–179.

Glorot, X. and Bengio, Y. (2010), Understanding the difficulty of training deep feedforward neural networks, *in* 'Proceedings of the thirteenth international conference on artificial intelligence and statistics', pp. 249–256.

Glynn, P. W. (1990), 'Likelihood ratio gradient estimation for stochastic systems', *Communications of the ACM* **33**(10), 75–84.

Gordon, J. and Abramov, I. (1977), 'Color vision in the peripheral retina. ii. hue and saturation', *JOSA* **67**(2), 202–207.

Halverson, T. and Hornof, A. J. (2004), Local density guides visual search: Sparse groups are first and faster, *in* 'Proceedings of the human factors and ergonomics society annual meeting', Vol. 48, SAGE Publications Sage CA: Los Angeles, CA, pp. 1860–1864.

Hausknecht, M. and Stone, P. (2015), 'Deep recurrent q-learning for partially observable mdps', *arXiv preprint arXiv:1507.06527* .

Hayhoe, M. and Ballard, D. (2014), 'Modeling task control of eye movements', *Current Biology* **24**(13), R622–R628.

Hayhoe, M. M., Shrivastava, A., Mruczek, R. and Pelz, J. B. (2003), 'Visual memory and motor planning in a natural task', *Journal of vision* **3**(1), 6–6.

Heess, N., Hunt, J. J., Lillicrap, T. P. and Silver, D. (2015), 'Memory-based control with recurrent neural networks', *arXiv preprint arXiv:1512.04455* .

Hikosaka, O., Takikawa, Y. and Kawagoe, R. (2000), 'Role of the basal ganglia in the control of purposive saccadic eye movements', *Physiological reviews* **80**(3), 953–978.

Howes, A., Lewis, R. L. and Vera, A. (2009), 'Rational adaptation under task and processing constraints: implications for testing theories of cognition and action.', *Psychological review* **116**(4), 717.

Howes, A., Warren, P. A., Farmer, G., El-Deredy, W. and Lewis, R. L. (2016), 'Why contextual preference reversals maximize expected value.', *Psychological review* **123**(4), 368.

Itti, L., Gold, C. and Koch, C. (2001), 'Visual attention and target detection in cluttered natural scenes', *Optical Engineering* **40**(9), 1784–1794.

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

Itti, L. and Koch, C. (2001), 'Computational modelling of visual attention', *Nature reviews neuroscience* **2**(3), 194.

Itti, L., Koch, C. and Niebur, E. (1998), 'A model of saliency-based visual attention for rapid scene analysis', *IEEE Transactions on pattern analysis and machine intelligence* **20**(11), 1254–1259.

Jagacinski, R. J. and Flach, J. M. (2003), *Control theory for humans: Quantitative approaches to modeling performance*, CRC Press.

Kaelbling, L. P., Littman, M. L. and Cassandra, A. R. (1998), 'Planning and acting in partially observable stochastic domains', *Artificial intelligence* **101**(1), 99–134.

Kahneman, D. and Tversky, A. (2013), Prospect theory: An analysis of decision under risk, *in* 'Handbook of the fundamentals of financial decision making: Part I', World Scientific, pp. 99–127.

Kangasrääsiö, A., Athukorala, K., Howes, A., Corander, J., Kaski, S. and Oulasvirta, A. (2017), Inferring cognitive models from data using approximate bayesian computation, *in* 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', ACM, pp. 1295–1306.

Kieras, D. E. and Hornof, A. J. (2014), Towards accurate and practical predictive models of active-vision-based visual search, *in* 'Proceedings of the SIGCHI conference on human factors in computing systems', ACM, pp. 3875–3884.

Kieras, D. E., Hornof, A. and Zhang, Y. (2015*a*), Visual search of displays of many objects: Modeling detailed eye movement effects with improved epic, In Proceedings of the 13th international conference on cognitive modeling, pp. 55–60.

Kieras, D. E., Hornof, A. and Zhang, Y. (2015*b*), Visual search of displays of many objects: Modeling detailed eye movement effects with improved epic, *in* 'Proceedings of the 13th international conference on cognitive modeling', p. 55.

Kieras, D. E. and 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.

Koch, C. and Ullman, S. (1987), Shifts in selective visual attention: towards the underlying neural circuitry, *in* 'Matters of intelligence', Springer, pp. 115–141.

Kowler, E. (1990), 'The role of visual and cognitive processes in the control of eye movement.', *Reviews of oculomotor research* **4**, 1–70.

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

Land, M. F. and Hayhoe, M. (2001), 'In what ways do eye movements contribute to everyday activities?', *Vision research* **41**(25-26), 3559–3565.

Leibo, J. Z., d'Autume, C. d. M., Zoran, D., Amos, D., Beattie, C., Anderson, K., Castañeda, A. G., Sanchez, M., Green, S., Gruslys, A. et al. (2018), 'Psychlab: a psychology laboratory for deep reinforcement learning agents', *arXiv preprint arXiv:1801.08116* .

Levi, D. M. (2008), 'Crowdingan essential bottleneck for object recognition: A minireview', *Vision research* **48**(5), 635–654.

Lewis, R. L., Howes, A. and Singh, S. (2014), 'Computational rationality: Linking mechanism and behavior through bounded utility maximization', *Topics in cognitive science* **6**(2), 279–311.

Li, Y., Bengio, S. and Bailly, G. (2018), Predicting human performance in vertical menu selection using deep learning, *in* 'Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems', ACM, p. 29.

Littman, M. L. (1996), Algorithms for sequential decision making, PhD thesis, Brown University, Rhode Island.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C. (2016), Ssd: Single shot multibox detector, *in* 'European conference on computer vision', Springer, pp. 21–37.

Liu, W., Bailly, G. and Howes, A. (2017), Effects of frequency distribution on linear menu performance, *in* 'Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems', ACM, pp. 1307–1312.

Loftus, G. R. and Mackworth, N. H. (1978), 'Cognitive determinants of fixation location during picture viewing.', *Journal of Experimental Psychology: Human perception and performance* **4**(4), 565.

Ma, Y.-F., Hua, X.-S., Lu, L. and Zhang, H.-J. (2005), 'A generic framework of user attention model and its application in video summarization', *IEEE transactions on multimedia* **7**(5), 907–919.

Mackworth, N. H. (1976), 'Stimulus density limits the useful field of view', *Eye movements and psychological processes* pp. 307–321.

McCallum, A. K. (1996), Reinforcement learning with selective perception and hidden state, PhD thesis, University of Rochester. Dept. of Computer Science.

McCarley, J. S., Wang, R. F., Kramer, A. F., Irwin, D. E. and Peterson, M. S. (2003), 'How much memory does oculomotor search have?', *Psychological Science* **14**(5), 422–426.

McClelland, J. L. and Cleeremans, A. (2009), 'Connectionist models', *Oxford Companion to Consciousness* .

McConkie, G. W. and Rayner, K. (1975), 'The span of the effective stimulus during a fixation in reading', *Perception & Psychophysics* **17**(6), 578–586.

Mennie, N., Hayhoe, M. and Sullivan, B. (2007), 'Look-ahead fixations: anticipatory eye movements in natural tasks', *Experimental Brain Research* **179**(3), 427–442.

Miau, F., Papageorgiou, C. S. and Itti, L. (2001), Neuromorphic algorithms for computer vision and attention, *in* 'Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation IV', Vol. 4479, International Society for Optics and Photonics, pp. 12–24.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D. and Kavukcuoglu, K. (2016), Asynchronous methods for deep reinforcement learning, *in* 'International conference on machine learning', pp. 1928–1937.

Mnih, V., Heess, N., Graves, A. et al. (2014), Recurrent models of visual attention, *in* ‘Advances in Neural Information Processing Systems’, pp. 2204–2212.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G. et al. (2015), ‘Human-level control through deep reinforcement learning’, *Nature* **518**(7540), 529–533.

Müller, H. J. and Krummenacher, J. (2006), ‘Visual search and selective attention’, *Visual Cognition* **14**(4-8), 389–410.

Myers, C. W., Lewis, R. L. and Howes, A. (2013), Bounded optimal state estimation and control in visual search: Explaining distractor ratio effects, *in* ‘Proc. CogSci’.

Najemnik, J. and Geisler, W. S. (2005), ‘Optimal eye movement strategies in visual search’, *Nature* **434**(7031), 387–391.

Najemnik, J. and Geisler, W. S. (2008), ‘Eye movement statistics in humans are consistent with an optimal search strategy’, *Journal of Vision* **8**(3), 4–4.

Nakayama, K., Silverman, G. H. et al. (1986), ‘Serial and parallel processing of visual feature conjunctions’, *Nature* **320**(6059), 264–265.

Navalpakkam, V. and Itti, L. (2005), ‘Modeling the influence of task on attention’, *Vision research* **45**(2), 205–231.

Navalpakkam, V., Koch, C., Rangel, A. and Perona, P. (2010), ‘Optimal reward harvesting in complex perceptual environments’, *Proceedings of the National Academy of Sciences* **107**(11), 5232–5237.

Nelson, W. W. and Loftus, G. R. (1980), ‘The functional visual field during picture viewing.’, *Journal of Experimental Psychology: Human Learning and Memory* **6**(4), 391.

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

Pascanu, R., Mikolov, T. and Bengio, Y. (2013), On the difficulty of training recurrent neural networks, *in* ‘International Conference on Machine Learning’, pp. 1310–1318.

Payne, S. J. and 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.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E. and McCarley, J. S. (2001), ‘Visual search has memory’, *Psychological Science* **12**(4), 287–292.

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

Rahnev, D. and Denison, R. N. (2018), ‘Suboptimality in perceptual decision making’, *Behavioral and Brain Sciences* pp. 1–107.

Rao, R. P. (2010), 'Decision making under uncertainty: a neural model based on partially observable markov decision processes', *Frontiers in computational neuroscience* **4**, 146.

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

Rayner, K. and Castelhano, M. S. (2008), 'Eye movements during reading, scene perception, visual search, and while looking at print advertisements.', *Visual marketing: From attention to action* pp. 9–42.

Renninger, L. W., Verghese, P. and Coughlan, J. (2007), 'Where to look next? eye movements reduce local uncertainty', *Journal of Vision* **7**(3), 6–6.

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

Salvucci, D. D. (2001), 'An integrated model of eye movements and visual encoding', *Cognitive Systems Research* **1**(4), 201–220.

Schilling, H. E., Rayner, K. and Chumbley, J. I. (1998), 'Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences', *Memory & Cognition* **26**(6), 1270–1281.

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

Schulman, J., Moritz, P., Levine, S., Jordan, M. and Abbeel, P. (2015), 'High-dimensional continuous control using generalized advantage estimation', *arXiv preprint arXiv:1506.02438* .

Shen, J., Reingold, E. M. and Pomplun, M. (2000), 'Distractor ratio influences patterns of eye movements during visual search', *Perception* **29**(2), 241–250.

Shen, J., Reingold, E. M. and Pomplun, M. (2003), 'Guidance of eye movements during conjunctive visual search: the distractor-ratio effect.', *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* **57**(2), 76.

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

Sprague, N. and Ballard, D. (2004), Eye movements for reward maximization, *in* 'Advances in neural information processing systems', pp. 1467–1474.

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

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

Stritzke, M., Trommershäuser, J. and Gegenfurtner, K. R. (2009), 'Effects of salience and reward information during saccadic decisions under risk', *JOSA A* **26**(11), B1–B13.

Sutton, R. S. and Barto, A. G. (1998), *Reinforcement learning: An introduction*, Vol. 1, MIT press Cambridge.

Sutton, R. S., McAllester, D. A., Singh, S. P. and Mansour, Y. (2000), Policy gradient methods for reinforcement learning with function approximation, *in* 'Advances in neural information processing systems', pp. 1057–1063.

Tatler, B. W. (2007), 'The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions', *Journal of vision* **7**(14), 4–4.

Treisman, A. M. and Gelade, G. (1980), 'A feature-integration theory of attention', *Cognitive psychology* **12**(1), 97–136.

Tseng, Y.-C. and Howes, A. (2008), The adaptation of visual search strategy to expected information gain, *in* 'Proceedings of the SIGCHI conference on Human factors in computing systems', ACM, pp. 1075–1084.

Tseng, Y.-C. and Howes, A. (2015), 'The adaptation of visual search to utility, ecology and design', *International Journal of Human-Computer Studies* **80**, 45–55.

Verghese, P. (2001), 'Visual search and attention: A signal detection theory approach', *Neuron* **31**(4), 523–535.

Vincent, B. T. (2015), 'Bayesian accounts of covert selective attention: A tutorial review', *Attention, Perception, & Psychophysics* **77**(4), 1013–1032.

Vincent, B. T. (2016), 'Hierarchical bayesian estimation and hypothesis testing for delay discounting tasks', *Behavior research methods* **48**(4), 1608–1620.

Vlaskamp, B. N., Over, E. A. and Hooge, I. T. C. (2005), 'Saccadic search performance: the effect of element spacing', *Experimental brain research* **167**(2), 246–259.

Von Neumann, J. and Morgenstern, O. (2007), *Theory of games and economic behavior (commemorative edition)*, Princeton university press.

Wang, W., Chen, C., Wang, Y., Jiang, T., Fang, F. and Yao, Y. (2011), Simulating human saccadic scanpaths on natural images, *in* 'Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on', IEEE, pp. 441–448.

Wierstra, D., Förster, A., Peters, J. and Schmidhuber, J. (2010), 'Recurrent policy gradients', *Logic Journal of the IGPL* **18**(5), 620–634.

Williams, L. (1967), 'The effects of target specification on objects fixated during visual search', *Acta Psychologica* **27**(Supplement C), 355 – 360.
URL: *http://www.sciencedirect.com/science/article/pii/0001691867900807*

Williams, R. J. (1992), 'Simple statistical gradient-following algorithms for connectionist reinforcement learning', *Machine learning* **8**(3-4), 229–256.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. and Niv, Y. (2014), 'Orbitofrontal cortex as a cognitive map of task space', *Neuron* **81**(2), 267–279.

Wolfe, J. M. (1994), 'Guided search 2.0 a revised model of visual search', *Psychonomic bulletin & review* **1**(2), 202–238.

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

Wolfe, J. M. (2012), When do i quit? the search termination problem in visual search, *in* 'The influence of attention, learning, and motivation on visual search', Springer, pp. 183–208.

Wolfe, J. M. and Horowitz, T. S. (2004), 'What attributes guide the deployment of visual attention and how do they do it?', *Nature reviews neuroscience* **5**(6), 495.

Wolfe, J. M., Horowitz, T. S. and Kenner, N. M. (2005), 'Cognitive psychology: rare items often missed in visual searches', *Nature* **435**(7041), 439.

Woodman, G. F. and Chun, M. M. (2006), 'The role of working memory and long-term memory in visual search', *Visual Cognition* **14**(4-8), 808–830.

Wright, R. D. and Ward, L. M. (2008), *Orienting of attention*, Oxford University Press.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R. and Bengio, Y. (2015), Show, attend and tell: Neural image caption generation with visual attention, *in* 'International conference on machine learning', pp. 2048–2057.

Yarbus, A. (1967), 'Eye movements and vision. 1967', *New York* .

Yu, A. J., Dayan, P. and 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.

Zohary, E. and Hochstein, S. (1989), 'How serial is serial processing in vision?', *Perception* **18**(2), 191–200.