UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Learning and Decision-Making in Active Cognition

Permalink

https://escholarship.org/uc/item/7jm9x97x

Author

Ahmad, Sheeraz

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Learning and Decision-Making in Active Cognition

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Computer Science

by

Sheeraz Ahmad

Committee in charge:

Professor Angela J. Yu, Chair
Professor Sanjoy Dasgupta
Professor Charles Elkan
Professor Tara Javidi
Professor Lawrence Saul

2015

Copyright

Sheeraz Ahmad, 2015

All rights reserved.

The Dissertation of Sheeraz Ahmad is approved and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

This dissertation is dedicated to my Mom for her unconditional love and constant encouragement to pursue my education.

EPIGRAPH

I am always doing that which I cannot do, in order that I may learn how to do it.

Pablo Picasso

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Figures viii

Acknowledgements xi

Vita xii

Abstract of the Dissertation xiii

Chapter 1 Introduction 1
1.1 Outline 2
1.2 Organization 4

Chapter 2 Active Sensing 6
2.1 Motivation 7
2.2 Methods 10
2.2.1 Inference 11
2.2.2 C-DAC Decision Policy 11
2.2.3 Greedy MAP Policy 14
2.2.4 Infomax Decision Policy 15
2.2.5 Approximate Decision Policy 16
2.2.6 Experimental Design 21
2.3 Results 24
2.3.1 Policy Comparison 24
2.3.2 Behavioral Comparison 27
2.3.3 Model Predictions 32
2.3.4 Visual Search with Peripheral Vision 33
2.4 Discussion 36

Chapter 3 Competitive Foraging 42
3.1 Motivation 43
3.2 Background 46
3.2.1 Human Experiments 46
3.2.2 Existing Models 47
3.3 Socially Aware Bayesian Agents (SABA) 49

3.3.1 Bayesian Inference 49
3.3.2 Decision Strategies 51
3.4 Naive Agents 52
3.5 Results 53
3.6 Discussion 55

Chapter 4 Multi-attribute Decision Making 59
4.1 Motivation 60
4.2 Methods 63
4.2.1 Option Model 63
4.2.2 Inference and Decision Making 64
4.2.3 Consumption Model 68
4.3 Results 71
4.4 Discussion 75

Chapter 5 Conclusion 79

Appendix A 82
A.1 Greedy MAP Policy: Analytical Solution 82

Bibliography 85

LIST OF FIGURES

Figure 2.1. C-DAC Q-factors for different actions. Blue: stop & declare. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3. 17

Figure 2.2. Simple visual search task, with gaze contingent display. 22

Figure 2.3. Experiment design and data for the visual search task. On each trial, subjects must find the target amongst the two distractors. First and second fixations are probabilistically matched. 23

Figure 2.4. Decision policies – Infomax resembles C-DAC. Blue: stop. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3. Environment $(c, c_s, \beta) = (0.1, 0, 0.9)$. Threshold for infomax and greedy MAP = 0.8 25

Figure 2.5. C-DAC policy for different environments (c, c_s, β) – high cost of time, c $(0.2, 0, 0.9)$, high noise $(0.1, 0, 0.7)$, and added switch cost, c_s $(0.1, 0.1, 0.9)$. 26

Figure 2.6. Decision policies as a function of switching cost. (A) No switch-cost environment: $(c, c_s, \beta) = (0.1, 0, 0.9)$. (B) Switch-cost environment: $(c, c_s, \beta) = (0.1, 0.1, 0.9)$. 27

Figure 2.7. Confirmation bias: the influence of prior spatial knowledge on sensory processing and decision-making, in humans and models, manifesting in accuracy as well as response time. 30

Figure 2.8. Behavioral predictions as a function of switching cost. (A) Accuracy, (B) search time, (C) number of switches, and (D) total behavioral cost, for C-DAC and Infomax for two environments $(c, c_s, \beta) = (0.005, 0, 0.68)$ and $(0.005, 0.1, 0.68)$. 32

Figure 2.9. Schematics of visual search task. The general task is to find the target amongst distractors. (A) Task 1: agent fixates one of the target patches at any given time. (B) Task 2: agent fixates one of the blue circle regions at any given time. 34

Figure 2.10. Decision policies. Azure: stop and choose location l_1. Blue: stop and choose l_2. Indigo: stop and choose l_3. Green: fixate l_1. Seagreen: fixate l_2. Olive: fixate l_3. Red: fixate l_{12}. Brown: fixate l_{23}. Yellow: fixate l_{13}. 35

Figure 2.11. Comparison between C-DAC and Infomax on Task 2 for two environments $(c, \beta 1, \beta_2, \beta_3, \beta_4) = (0.05, 0.62, 0.6, 0.55, 0.5)$, $c_s = 0$ and $c_s = 0.005$. C-DAC adjusts the time steps and number of switches depending on the environment. 36

Figure 3.1. Stationary Environment: [7 3]. Dotted lines indicate the patch richness level (hidden). Solid lines indicate the normalized agent density on each patch, blue: patch 1 and green: patch 2. 54

Figure 3.2. Inferred beliefs of agents at different patches. The boundaries indicate the standard deviation of beliefs. Dotted lines show the true richness level of the two patches. 54

Figure 3.3. Non-stationary Environment: [6 4] $\rightarrow$ [7 3]. Dotted lines indicate the patch richness level (hidden) which changes at time-step 150. Solid lines indicate the normalized agent density on each patch, blue: patch 1 and green: patch 2. 55

Figure 4.1. Contextual effects. Options X and Y are equi-preferable in the absence of a third option. Arrow denotes the direction of the preference shift when a decoy, Z, is introduced. (A) Attraction. (B) Compromise. (C) Similarity. (D) Phantom decoy effect. 61

Figure 4.2. Graphical model showing how options are generated in the market. The square box (plate) represents repetitions, so that each i corresponds to a different option. The unshaded nodes represent hidden variables and the shaded one represents observations. 64

Figure 4.3. Augmented model to incorporate the additional information provided by phantom decoys. $c = \mathbf{1}(v_1 > v_2)$ is the variable encoding which option was consumed, where $\mathbf{1}(.)$ is the indicator function. 68

Figure 4.4. Graphical model showing how options are generated in the market. (B) All options lying on the same curve have the same value/utility as denoted on the curve. (C) All options lying on the same curve have the same attribute trade-off (ζ), as denoted on the curve. . . . 69

Figure 4.5. Proportion of choices for different effects and their explanation. (A) & (E) X and Y are equi-preferable. (B) & (F) Attraction effect. (C) & (G) Compromise effect. (D) & (H) Similarity effect. 72

Figure 4.6. (A) Information about an unavailable, or sold out, option increases the preference for Y relative to X. (B) The phantom decoy makes Y appear better on average. 73

Figure 4.7. (A) Individual difference in contextual effects. Compromise effect becomes insignificant for individuals that value attribute 1 less. (B) & (C) Evolution of belief about market preference with experience for two individuals starting with different priors. 74

Figure A.1. (A) Analytical Greedy MAP policy plot for the region $\mathbf{p}_t^1 > \mathbf{p}_t^2 > \mathbf{p}_t^3$. (B) Analytical Greedy MAP policy plot for the complete simplex. (C) Simulated Greedy MAP policy. When more that one actions is optimal in a region, the action is chosen randomly. 84

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Angela J. Yu, for her guidance, tutelage and many stimulating discussions. Her support and mentorship has helped shape my life academically and beyond.

I would also like to thank my peers and my group members for broadening my research outlook, for providing valuable feedback through different projects, and for making graduate school an experience I will forever cherish.

Chapter 2 contains material from S. Ahmad and A.J. Yu, Active sensing as bayes-optimal sequential decision-making, Uncertainty in Artificial Intelligence (UAI), 2013; S. Ahmad, H Huang, and A.J. Yu, Context-sensitive active sensing in humans, Advances in Neural Information Processing Systems (NIPS), 2013; and S. Ahmad, H Huang, and A.J. Yu, Cost-sensitive bayesian control policy in human active sensing, Frontiers in Human Neuroscience, 2014. The dissertation author was the primary investigator and author of these papers.

Chapter 3, in full, is a reprint of the material from S. Ahmad and A.J. Yu, A socially aware bayesian model for competitive foraging, Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2014. The dissertation author was the primary investigator and author of this paper.

Chapter 4 contains material from S. Ahmad and A.J. Yu, A rational model for individual differences in preference choice, Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2015.The dissertation author was the primary investigator and author of this paper.

Chapter 4 also contains material being prepared for submission. The dissertation author is the primary investigator and author of this material.

VITA

2008	Bachelor of Technology, Indian Institute of Technology, Kanpur
2008-2010	Member, Research Team, Anveshan Telecom, Bangalore, India
2015	Doctor of Philosophy in Computer Science University of California, San Diego

PUBLICATIONS

- S. Ahmad and A.J. Yu. A rational model for individual differences in preference choice. In Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2015.
- S. Ahmad, H Huang, and A.J. Yu. Cost-sensitive bayesian control policy in human active sensing. Frontiers in Human Neuroscience, 2014.
- S. Ahmad and A.J. Yu. A socially aware bayesian model for competitive foraging. In Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2014.
- S. Ahmad, H Huang, and A.J. Yu. Context-sensitive active sensing in humans. Advances in Neural Information Processing Systems (NIPS), 2013.
- S. Ahmad and A.J. Yu. Active sensing as bayes-optimal sequential decision-making. Uncertainty in Artificial Intelligence (UAI), 2013.
- S. Ahmad and A.J. Yu. Active sensing as bayes-optimal sequential decision-making. In NIPS: Workshop on Information in Perception and Action, December 2012.
- S. Ahmad and A.J. Yu. Active sensing: An optimal inference model (poster). In NIPS Workshop: Bayesian Nonparametric Models (BNPM) For Reliable Planning And Decision-Making Under Uncertainty, December 2012.

ABSTRACT OF THE DISSERTATION

Learning and Decision-Making in Active Cognition

by

Sheeraz Ahmad

Doctor of Philosophy in Computer Science

University of California, San Diego, 2015

Professor Angela J. Yu, Chair

Learning and decision making is one of the universal cornerstones of human and animal life. There are several situations where these processes are passive or secondary, but there are several others where the individual can actively control where to gain information to learn more effectively. For example, selecting which restaurant to visit next with the aim of finding out a consistently good dining option. Even though there have been some efforts to model problems in active cognition, a large class of such models is algorithmic, and only concerns itself with how such behavior may be replicated, but not how and why the algorithm actually comes about. This leads to a dearth of

computational models that can explain the behavior in a principled manner starting from basic, justifiable assumptions.

Employing tools such as hierarchical Bayesian modeling and stochastic control, this dissertation proposes novel models for several problems in cognition with the goals of enriching our understanding of how the brain might tackle such problems, as well as to provide useful guidelines for building artificial agents that could rival human performance on such problems. More specifically, we investigate three related problems and propose Bayesian models for each of these. Active sensing problem, which deals with optimally allocating sensory resources in a given sensing task. Competitive foraging problem, which deals with understanding foraging in an uncertain environment and in the presence of competing agents who affect each other's reward. And lastly, multi-attribute decision-making problem, which deals with understanding how humans make choices between multi-attribute alternatives and how such a choice mechanism could lead to the well-studied contextual effects.

Chapter 1

Introduction

Our daily lives are rife with situations where we need to learn from varying sources of information, be it learning a new route to our favorite restaurant through some hit and trial or learning how busy a workday will be based on the day of the week or the state of some ongoing project. Often accompanying these learning scenarios are the the opportunities to act on the learned information to optimize for some goals. For example, based on the information about possible routes, expected traffic and other conditions, needing to decide which route to take to get to our favorite restaurant the fastest; or deciding whether it would be a good idea to get an extra hour of sleep based on how much loss might be accrued in terms of lost work. In fact, such problems are not unique to humans, as animals have to learn about when and where they can expect a predator or a good source of food, and subsequently use that knowledge to decide when and where to camp.

For the purpose of this dissertation, we are going to define a broad classification for such learning and decision making problems. The passive setting, where the individual has no control over the source of information, like watching an online lecture before deciding if it's alright to move to the next topic. This contrasts with the active setting, where the individual can assert overt control over the source of information, even switching from one source to another to optimize some goals, like a setting where

the student can switch between watching the lecture, asking real time questions and solving problems before deciding to move to the next topic. Even though there are situations when the learning maybe passive, like figuring out which traffic lights in your neighborhood are the longest as you drive to your intended destinations and back, there are several others where the source of information can be chosen to facilitate learning, like consciously choosing a somewhat longer route to get more information about how long the lights are on that route. Confirming with common intuition, active cognition is usually expected to lead to faster and often more robust learning.

This dissertation focuses on the problems of how to build powerful computational models to explain learning and decision making in the active cognition setting starting from justifiable assumptions about the representation of the environment, generation of observables and update mechanisms to incorporate information to the existing state of knowledge. As a high level summary, we represent the environment and generation of observables probabilistically to allow for the inclusion of uncertainty about those, and we use Bayesian inference to update the state of knowledge. For the decision-making phase, we either use greedy policies for their simplicity, or policies that optimize long term gains/costs for their exactness.

1.1 Outline

We tackle three such problems, each of which constitutes a chapter detailing our research contribution. A brief description follows:

- **Active Sensing:** The problem of deciding how to allocate sensory resources to the most pertinent aspects of the environment when performing a specific task. We present a Bayes-optimal framework, termed C-DAC (Context-Dependent Active Controller) from here on, that directly optimizes for task-dependent variables, such

as speed, effort and accuracy. We compare our model's behavior to human behavior on a visual search task and discover that it not only matches coarse performance measures such as speed and accuracy, but also a finer, peculiar phenomenon known as confirmation bias, wherein responses are biased towards the most likely outcome. To enable the application of our model to more complex tasks, we investigate several approximation schemes based either on learning simpler representations of the long-term reward structure or using a greedy approach, and note that these can still capture interesting behavior while providing substantial computational savings. Our model also generates novel prediction on tasks that can be used as guidelines for future experiment design. Lastly, we present an extension of our model that allows for the incorporation of peripheral vision as sub-quality data, potentially paving the way for sophisticated models of more natural tasks where peripheral vision needs to be considered.

- **Competitive Foraging:** The problems of deciding how to allocate resources to the most pertinent aspects of the environment when performing a specific task, *in the presence of competitors working towards the same goals*. Motivated by the problem of how animals search for food, we present a model that captures how the reward in a location depends on both the quality of the location and the competition there, and how to handle a new trade-off resulting from the presence of other agents, besides the usual exploration-exploitation trade-off. This new follow-avoid trade-off means that following the crowd can be beneficial since the competitors might know something about the environment that we don't and potentially lead us to a location with abundant rewards, but at the same time, being in a location with a lot of other competitors reduces the amount of reward one might get, so avoiding the competition and favoring less crowded locations might be more beneficial. We

present a model, termed Socially Aware Bayesian Agents (SABA) from here on, that takes into account both self-information gained from personal experience and public-information gained from observing others while making decisions. This is the first model for foraging that directly considers competition and shows that optimal strategies can emerge even when the agents' goals are "selfish" as opposed to social.

- **Multi-attribute Decision Making:** The problem of understanding how the mere presence of options in a choice preference task can affect the judgement about the relative values of other options. When faced with multi-attribute options, like cars with different ratings on safety and economy, humans exhibit some peculiar, seemingly irrational behavior, wherein the introduction of arguably irrelevant options can affect choices. Extensive empirical studies have revealed four well known effects – attraction, compromise, similarity and phantom decoy effect. We present the first Bayesian model for the problem that can explain all four effects within the same paradigm. We also present extensions to explain individual differences and discuss planned work on how the choices maybe guided by more than just value-based rewards, aimed at actively discovering preferences.

1.2 Organization

The rest of this dissertation is organized as follows. In Chapter 2, we describe in detail the problem of active sensing and present a Bayes-optimal model that outperforms existing models on both comparison with human behavior as well as other artificial tasks. Chapter 3 motivates the problem of competitive foraging from animal behavior, and presents a model that incorporates competition and still converges to the optimal strategy. In Chapter 4, we first introduce contextual effects as have been observed in

multi-attribute choice scenarios and present a model that can explain all major effects in the same paradigm. We also provide extensions that can explain individual differences and hint at a potential way to model preference discovery. We conclude with a summary of our research contributions in Chapter 5.

Chapter 2

Active Sensing

An important but poorly understood aspect of sensory processing is the role of *active sensing*, the use of self-motion such as eye or head movements to focus sensing resources on the most rewarding or informative aspects of the sensory environment. Here, we present behavioral data from a visual search experiment, as well as a Bayesian model of within-trial dynamics of sensory processing and eye movements. Within this Bayes-optimal inference and control framework, which we call C-DAC (Context-Dependent Active Controller), various types of *behavioral costs*, such as temporal delay, response error, and sensor repositioning cost, are explicitly minimized. This contrasts with previously proposed algorithms that optimize abstract statistical objectives such as anticipated information gain (Infomax) (Butko and Movellan, 2010) and one-step look-ahead accuracy (greedy MAP) (Najemnik and Geisler, 2005). We find that C-DAC captures human visual search dynamics better than previous models, in particular a certain form of "confirmation bias" apparent in the way human subjects utilize prior knowledge about the spatial distribution of the search target to improve search speed and accuracy. We also examine several computationally efficient approximations to C-DAC that may present biologically more plausible accounts of the neural computations underlying active sensing, as well as practical tools for solving active sensing problems in engineering applications. To summarize, this chapter provides an introduction to the problems of

active sensing and describes our key contributions: human visual search behavioral data, a context-sensitive Bayesian active sensing model, a comparative study between different models of human active sensing, and a family of efficient approximations to the optimal model.

2.1 Motivation

The brain excels at performing sensory inference under conditions of uncertainty. One important tool the brain has at its disposal is *active sensing*, the ability to utilize self-motion in order to allocate sensing resources toward the most rewarding or informative aspects of the environment. Most prominent models of sensory processing presume *passiveness* in data collection (Riesenhuber and Poggio, 1999; Simoncelli and Heeger, 1998; Lewicki, 2002; Wilson and Mainen, 2006), considering only how to represent or compute with given inputs, and not how to actively intervene in the input collection process itself, especially with respect to behavioral goals or environmental constraints. Having a formal understanding of active sensing is not only important for advancing neuroscientific progress but also for engineering applications, such as developing context-sensitive, interactive artificial sensing systems.

The most well-studied aspect of human active sensing is saccadic eye movements (Yarbus, 1967). While most existing models have focused on bottom-up saliency factors related to image statistics (Koch and Ullman, 1985; Itti and Koch, 2000; Rao et al., 2002; Lee and Yu, 2000; Itti and Baldi, 2006), empirical data have long indicated eye movements to be under the influence of a variety of cognitive factors (Yarbus, 1967; Henderson, 2007; Rayner, 1998; Land and Hayhoe, 2001; Land and Lee, 1994), such as prior knowledge about target location (He and Kowler, 1989), temporal onset (Oswal et al., 2007), and reward probabilities (Roesch and Olson, 2003). Models that assume saccadic eye movements to be passive or automatic responses to the visual scene sidestep

the computational consequences of saccadic decisions, and thus cannot explain how behavioral goals and environmental context ought to influence eye movements. In part to overcome this limitation, the notion of saliency has been reframed probabilistically and dynamically in terms of maximizing the future informational gain (Infomax) given current beliefs about the visual scene (Lee and Yu, 2000; Itti and Baldi, 2006; Butko and Movellan, 2010). Separately, it has also been proposed that saccades may be dynamically chosen to maximize the greedy, one-step look-ahead probability of finding the target (Greedy MAP), conditioned on an individual's visual acuity map (Najemnik and Geisler, 2005).

While both Infomax and Greedy MAP brought a new level of sophistication to the modeling of eye movements – in particular, representing sensory processing as iterative Bayesian inference, quantifying the knowledge gain of different saccade choices, and incorporating knowledge about sensory noise – they are still limited in some key respects: (1) they optimize abstract computational quantities that do not directly relate to behavioral goals (eg, speed and accuracy) or task constraints (eg, the cost of moving sensors from one location to another); (2) there is no explicit representation of time in these algorithms, and thus no principled means of trading off fixation duration or number of fixations with search accuracy.

In this chapter, we propose a behaviorally-grounded inference and control framework for active sensing, which we call C-DAC (Context-Dependent Active Controller). In contrast to Infomax and Greedy MAP, which optimize abstract statistical objectives, C-DAC explicitly optimizes behaviorally defined objectives. Specifically, we assume that the observer aims to optimize a context-sensitive objective function that takes into account behavioral costs such as temporal delay, response error, and the cost of switching from one sensing location to another. C-DAC uses this compound objective to choose when and where to collect sensory data, based on a continually updated, statistically

optimal (Bayesian) representation of the history of sequential observations. This framework allows us to derive behaviorally optimal procedures for making decisions about (1) where to acquire sensory inputs, (2) when to move from one observation location to another, and (3) how to negotiate the exploration-exploitation tradeoff between collecting additional data and terminating the observation process.

Complementing the modeling work, we also present data from a visual search experiment, in which subjects are required to find a designated target stimulus amongst distractor stimuli, in the presence of sensory noise. A critical experimental manipulation is that some locations are more likely to contain the target than others, during each block of trials. As reported elsewhere (Yu and Huang, 2014), we found that subjects' first fixation choice on each trial is consistent with having internalized spatial statistics of target location based on experienced target location in previous trials. In this work, we analyze how this prior information exerts control over the within-trial, fine-temporal dynamics of eye movements and perceptual processing. We find that humans exhibit a certain confirmation bias on this task – a tendency to favor the most likely location in processing speed and perceptual choice. We compare the various models with human behavior, in terms of fixation choice and duration, to investigate whether human behavior is better explained by taking into account task-specific considerations, as in C-DAC, or whether it is sufficient to optimize a generic statistical goal, like that of Infomax or Greedy-MAP.

Due to its contextual sensitivity, C-DAC may also improve on current algorithms in engineering applications, such as the digital eye (Butko and Movellan, 2010), where Infomax has been applied. However, the standard numerical solution for C-DAC is computationally intense – it scales exponentially with the number of potential target locations, making it impractical for most real-world applications. To address this issue, we propose two kind of approximations for C-DAC: (1) based on a low-dimensional

parametric and non-parametric approximation of the value function, (2) based on one step look-ahead myopic approximation of the value function. These approximations retain context-sensitivity while significantly reducing computational complexity. We discuss how these approximations compare to C-DAC as well as their limitations that demand further investigation.

The rest of the chapter is organized as follows. We first describe in detail the C-DAC model and different approximations, as well as the visual search experiment that we use to validate the models. We then present model comparison in terms of the decision policies they yield, followed by the behavioral data from the experiment along with comparison of models with human behavior. We also provide model predictions in novel scenarios that can be tested empirically. We conclude with a discussion of the implications of this work and its relationship to previous work.

2.2 Methods

In the following, we model active sensing as solving a sequential (Markov) decision process problem, where the observer can control both the location where the data is collected, and the amount of data collected in each location. Although we use language specific to visual search for concreteness, the framework discussed below easily generalizes to a broad range of active sensing problems. As a Markov decision process problem, active sensing consists of two separable computational components, an *inference* component that extracts a useful representation of the external environment based on the observed stream of noisy data, and a *decision* component that decides how to maintain or modify or terminate the sensing process as a function of this extracted representation. In the following, we begin with a discussion of the inference process, common to C-DAC, Greedy MAP and Infomax, and then discuss the respective decision policies. We then discuss various possible approximations to the C-DAC policy, and

conclude with a description of the visual search experiment.

2.2.1 Inference

We assume the observer starts with a prior belief over the true target location s (out of k possible locations), based on prior searching encounters (trials in the experimental setting), and then updates his/her belief about target location via Bayes' rule, upon receiving each new observation x_t. The observation noise is assumed to be i.i.d. conditioned on target location s and sensing (fixation) location λ_t. Thus, if we denote the sequence of fixation locations up to time t as $\lambda_t := \{\lambda_1, \ldots, \lambda_t\}$, the sequence of observations up to time t as $\mathbf{x}_t := \{x_1, \ldots, x_t\}$, and the belief state at time t as $\mathbf{p}_t := (P(s=1|\mathbf{x}_t;\lambda_t), \ldots, P(s=k|\mathbf{x}_t;\lambda_t))$, then the belief-update rule according to Bayes' theorem is:

$$\mathbf{p}_t^i = P(s=i|\mathbf{x}_t;\lambda_t) \propto p(x_t|s=i;\lambda_t)P(s=i|\mathbf{x}_{t-1};\lambda_{t-1}) = f_{s,\lambda_t}(x_t)\mathbf{p}_{t-1}^i \tag{2.1}$$

where $f_{s,\lambda}(x_t)$ is the likelihood function, and $\mathbf{p}_0$ the prior belief distribution over target location.

2.2.2 C-DAC Decision Policy

The action selection component of active vision is a stochastic control problem where the agent chooses the sensing location and the number of data points collected, and we assume the agent can optimize this process dynamically based on ongoing data collection and size of sensory data, but the exact consequence of each action is not perfectly known ahead of time. The goal is to find a good decision policy π, which maps the augmented belief state $(\mathbf{x}_t, \lambda_t)$ into an action $a \in A$, where A consists of a set of termination actions, stopping and choosing a response, and a set of continuation actions, obtaining data point from a certain observation location. The policy π produces for each

observation sequence $(x_1, \ldots, x_t, \ldots)$, a stopping time τ (number of data points observed), a sequence of fixation choices $\lambda_\tau := (\lambda_1, \ldots, \lambda_\tau)$, and an eventual target choice δ.

In the Bayes risk minimization framework, the optimization problem is formulated in terms of minimizing an *expected* cost function, $L_\pi := \mathbb{E}[l(\tau, \lambda_\tau, \delta)]_{\mathbf{x},s}$, averaged over stochasticity in the true target location s and the data samples $\mathbf{x}$. We assume that the cost incurred on each trial takes into account *temporal delay*, *switch cost (cost associated with each switch in sensing location)*, and *response error*, respectively. In accordance with the typical Bayes risk formulation of the sequential decision problem, we assume the cost function to be a linear combination of the relevant factors:

$$l(\tau, \delta; \lambda_\tau, s) = c\tau + c_s n_\tau + \mathbf{1}_{\{\delta \neq s\}} \tag{2.2}$$

where n_τ is the total number of switches ($n_\tau := \sum_{t=1}^{\tau-1} \mathbf{1}_{\{\lambda_{t+1} \neq \lambda_t\}}$), c parameterizes the cost of temporal delay, c_s the cost of a switch, and unit cost for response errors is assumed (as we can always divide c and c_s by the appropriate constant to make it 1). The expected cost is $L_\pi := c\mathbb{E}[\tau] + c_s\mathbb{E}[n_s] + P(\delta \neq s)$, where the expectation is taken over τ, λ, δ, and $\mathbf{x}_\tau$.

Bellman's dynamic programming equation (Bellman, 1952) tells us that the problem is optimized if at each time point, the agent chooses the action associated with the lowest expected cost (the *Q-factor* for that action), given his current knowledge or *belief state*, $\mathbf{p}_t$. The Q-factors for the stopping actions are straight forward: $\bar{Q}_t^i(\mathbf{p}_t, \lambda_t) := \mathbb{E}[l(t,i)|\mathbf{p}_t, \lambda_t] = ct + c_s n_t + (1 - \mathbf{p}_t^i)$. Obviously, the best stopping action δ is to minimize the probability of error. Thus, the stopping cost associated with the optimal stopping

action ($i^* := \text{argmax}_i \, \mathbf{p}_t^i$) is:

$$\begin{aligned} \bar{Q}_t^*(\mathbf{p}_t, \lambda_t) &:= \mathbb{E}[l(t, i^*) | \mathbf{p}_t, \lambda_t] \\ &= ct + c_s n_t + (1 - \mathbf{p}_t^{i^*}) \end{aligned} \tag{2.3}$$

The Q-factor associated with each continuation action j (continue sensing in location j) is:

$$Q_t^j(\mathbf{p}_t = \mathbf{p}, \lambda_t) := c(t+1) + c_s(n_t + \mathbf{1}_{\{j \neq \lambda_t\}}) + \min_{\tau', \delta, \boldsymbol{\lambda}_{\tau'}} \mathbb{E}[l(\tau', \delta) | \mathbf{p}_0 = \mathbf{p}, \lambda_1 = j] \tag{2.4}$$

with the optimal continuation action being $Q_t^* := \min_j Q_t^j = Q_t^{j^*}$. The expected cost of continuing observing in location j is equivalent to solving the original optimization problem with the prior belief set to the posterior after the previous t time-steps, and the first observation location being j. Suppose we define the *value function* $V(\mathbf{p}, i)$ as the expected cost associated with the optimal policy, given prior belief $\mathbf{p}_0 = \mathbf{p}$ and initial observation location $\lambda_1 = i$:

$$V(\mathbf{p}, i) := \min_{\tau, \delta, \boldsymbol{\lambda}_\tau} \mathbb{E}[l(\tau, \delta) | \mathbf{p}_0 = \mathbf{p}, \lambda_1 = i] \, . \tag{2.5}$$

Then the value function satisfies the following recursive relation:

$$\begin{aligned} V(\mathbf{p}, k) &= \min(\bar{Q}_1^*(\mathbf{p}, k), Q_1^*(\mathbf{p}, k)) \\ &= \min \Bigg(\left(\min_i \bar{Q}_1^i(\mathbf{p}, k) \right), \\ &\qquad \min_j \left(c + c_s \mathbf{1}_{\{j \neq k\}} + \mathbb{E}[V(\mathbf{p}', j)] \right) \Bigg) \end{aligned} \tag{2.6}$$

where $\mathbf{p}'$ is the belief state at next time-step, and the expectation is taken over the

stochasticity in the next observation x. The optimal policy effectively divides the belief state space into a *stopping region* ($\bar{Q}^* \leq Q^*$) and a *continuation region* ($\bar{Q}^* > Q^*$), each of which further divided into subregions corresponding to alternative continuation and stopping actions. Note that the optimal decision policy is a *stationary* policy: the value function depends only on the belief state and observation location at the time the decision is to be taken, and not on time *t per se*.

Bellman's dynamic programming principle implies a numerical algorithm for computing the optimal policy: guess an initial setting $V'(\mathbf{p},k)$ of the value function (e.g., minimal stopping cost associated with each belief state $\mathbf{p}$ and observation location k), then iterate Eq. 2.6 until convergence, which yields the value function $V(\mathbf{p},k) = V^{\infty}(\mathbf{p},k)$.

2.2.3 Greedy MAP Policy

The *greedy MAP* algorithm (Najemnik and Geisler, 2005) suggests that agents should try to maximize the expected one-step look-ahead probability of finding the target. Thus, the reward function is:

$$\begin{aligned} R^g(\mathbf{p}_t, j) &= \mathbb{E}_{x_{t+1}}[\max_i P(s=i|\mathbf{x}_t, x_{t+1}, \lambda_t, \lambda_{t+1}=j)] \\ &= \mathbb{E}_{x_{t+1}}[\max_i (\mathbf{p}_{t+1}^i)|x_{t+1}, \lambda_{t+1}=j] \end{aligned}$$

To keep the notations consistent, we define the associated *Q-factor*, cost and policy as:

$$\begin{aligned} Q^g(\mathbf{p}_t, j) &= -R^g(\mathbf{p}_t, j) \\ V^g(\mathbf{p}_t, j) &= \min_j Q^g(\mathbf{p}_t, j) \\ \lambda_{t+1}^g &= \operatorname*{argmin}_j Q^g(\mathbf{p}_t, j) \end{aligned}$$

2.2.4 Infomax Decision Policy

The objective of Infomax is to maximize long-term future information gain (or minimize cumulative future entropy of the posterior belief state) (Butko and Movellan, 2010). Thus, the action-values, value function, and the resultant policy can be formulated as follows in the Markov decision process framework:

$$\begin{aligned} Q^I(\mathbf{p}_t, j) &= \sum_{t'=t+1}^{T} \mathbb{E}[H(\mathbf{p}_{t'})|\mathbf{p}_t; \lambda_{t+1} = j]; \\ V^I(\mathbf{p}_t, j) &= \min_j Q_I(\mathbf{p}_t, j); \\ \lambda^I_{t+1} &= \operatorname{argmin}_j Q_I(\mathbf{p}_t, j) \end{aligned}$$

where $H(\mathbf{p}) = -\sum_i \mathbf{p}^i \mathbf{log} \mathbf{p}^i$ is Shannon's entropy.

Note that neither the original Greedy MAP nor the Infomax directly prescribes when to stop, since there are only continuation actions and no *stopping action*. A general heuristic used for such strategies is to stop when the confidence in one of the locations being the target (the belief about that location) exceeds a certain threshold, which is a free parameter challenging to set for any specific problem. We can augment these policies using the bound developed in Theorem. 1, which yields a stopping region that is a provable subset of the optimal stopping region for C-DAC, and which is sensitive to observation noise parameterized by β and the cost of time parameterized by c.

Simulations suggest that with the addition of the parametric threshold to Infomax and myopic C-DAC, these policies perform at par with C-DAC as long as there is no switch cost. Addition of switch code reveals the disparity, since Infomax has no systematic way to modulate the switching boundaries, whereas C-DAC, low-dimensional approximations and myopic C-DAC do by direct incorporation of the cost.

2.2.5 Approximate Decision Policy

Our model is formally a variant of POMDP (Partially Observable Markov Decision Process), or, more specifically, a Mixed Observability Markov Decision Process (MOMDP) (Ong et al., 2010; Araya-López et al., 2010), which differs from ordinary POMDP in that part of the state space is partly hidden (target location in our case) and partly observable (current fixation location in our case). In general, POMDPs are hard to solve since the decision made at each time step depends on all the past actions and observations, thus imposing enormous memory requirements. This is known as the curse of history, and is the first major hurdle towards any practical solution. An elegant way to alleviate this is to use belief states which serve as a sufficient statistic for the process history, thus requiring to maintain just a single distribution instead of the entire history. Converting a POMDP to a belief-state MDP is in fact a prevalent technique and the one we employ. However, this leads to another computational hurdle, known as the curse of dimensionality, since now we have a MDP with a continuous state-space, making tabular representation of value function infeasible. One way to work around the problem is to discretize the belief state space into a grid, where instead of finding the value function at all the points in the belief state simplex, we only do so for a finite number of grid points. The grid approximation, that we also use, has appealing performance guarantees which improve as the density of the grid is increased (Lovejoy, 1991). To evaluate the value function at the points not in this set, we use some sort of interpolation technique (value at the nearest grid point, weighted average value at k-nearest grid point, etc.). However, although grid approximation may work for small state spaces, it does not scale well to larger, practical problems. For example, when used for the active sensing problem with k sensing locations, a uniform grid of size n has $O(kn^{k-1})$ complexity.

Although there is a rich body of literature on approximate solutions of POMDP

(e.g. (Powell, 2007; Lagoudakis and Parr, 2003; Kaplow, 2010)) tackling both general as well as application-specific approximations, most are inappropriate for dealing with the MOMDP problem such as the one encountered here. Furthermore, most of the POMDP approximation algorithms focus on discounted rewards and/or finite-horizon problems. Our formulation does not fall into these categories and thus require novel approximation schemes. We present two approximate decision policies that limit the complexity of dynamic programming (Eq. 2.6) by either approximating the value function using a low-dimensional basis function representation, or by foregoing value iteration altogether by using a myopic horizon to assess action values.

Value Function Approximation

As shown in Fig. 2.1, the Q-factors (and the resulting value function) are *smooth and concave*, making them amenable to low dimensional approximations. At each step, we find a low dimensional representation of the value function, and use that for the update step of the value iteration algorithm. Specifically, instead of re-computing the value function at each grid point, here we generate a large number of samples uniformly on the belief state space, compute a new estimate of the value function at those locations, and then extrapolate the value function to everywhere by improving its parametric fit.

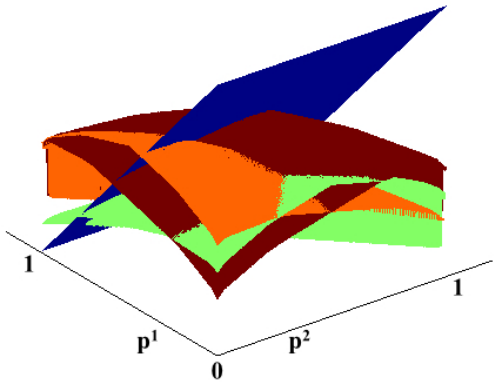


Figure 2.1. C-DAC Q-factors for different actions. Blue: stop & declare. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3.

The first low-dimensional approximation we consider is Radial Basis Function (RBF) approximation (Buhmann, 2003), which has been proposed to characterize neural encoding and computation in the visual pathways (Poggio, 1990), and also having been applied to many engineering applications:

1. Generate M RBFs, centered at $\{\mu_i\}_{i=1}^{M}$, with fixed σ: $\phi(\mathbf{p}) = \frac{1}{\sigma(2\pi)^{k/2}} e^{\frac{||\mathbf{p}-\mu_i||^2}{2\sigma^2}}$
2. Generate m random points from belief space, $\mathbf{p}$.
3. Initialize $\{V(\mathbf{p}_i)\}_{i=1}^{m}$ with the stopping costs.
4. Find minimum-norm $\mathbf{w}$ from: $V(\mathbf{p}) = \Phi(\mathbf{p})\mathbf{w}$.
5. Generate new m random belief state points $(\mathbf{p}')$.
6. Evaluate required V values using current $\mathbf{w}$.
7. Update $V(\mathbf{p}')$ using value iteration.
8. Find a new $\mathbf{w}$ from $V(\mathbf{p}') = \Phi(\mathbf{p}')\mathbf{w}$.
9. Repeat steps 5 through 8, until w converges.

While we adopt a Gaussian kernel function, other functional forms are also possible, e.g. multiquadratic ($\phi(\mathbf{p}) = \sqrt{1+\varepsilon||\mathbf{p}-\mu_i||^2}$), inverse-quadratic($\phi(\mathbf{p}) = (1+\varepsilon||\mathbf{p}-\mu_i||^2)^{-1}$), and thin plate spline ($\phi(\mathbf{p}) = ||\mathbf{p}-\mu_i||^2 \ln||\mathbf{p}-\mu_i||$). We have implemented some of these for our problem without observing significant deviations from the RBF results (data not shown).

The RBF approximation requires setting several parameters (number, mean, and variance of bases), which can be impractical for large problems, when there is little or no information available about the properties of the true value function. We thus also implement a nonparametric variation of the algorithm, whereby we use Gaussian Process Regression (GPR) (Williams and Rasmussen, 1996) to estimate the value function (step 4, 6 and 8). In addition, we also implement GPR with hyperparameter learning (Automatic Relevance Determination, ARD), thus obviating the need to pre-set model parameters.

The approximations lead to considerable computational savings. The complexity of the RBF approximation is $O(k(mM+M^3))$, for k sensing locations, m random points chosen at each step, and M bases. For the GPR approximation, the complexity is $O(kN^3)$, where N is the number of points used for regression. In practice, all the approximation algorithms we consider converge rapidly (under 10 iterations), though we do not have a proof that this holds for a general case.

Myopic Approximation

This approximation attempts to optimize the contextual cost of C-DAC, but only for one step in the future. In other words, the planning is based on the inherent assumption that the next action is the last non-terminal action permissible (or considered), and so the goal is to minimize the cost incurred in this single step. The actions available are, thus, stop and declare the current location as the target, or choose another sensing location before stopping. Similar to eq. 2.6, we can write the value function as:

$$V(\mathbf{p},k) = \min\left(\left(1-\mathbf{p}^k\right), \min_j\left(c + c_s\mathbf{1}_{\{j\neq k\}} + \min_{l_j}\left(1-\mathbb{E}[\mathbf{p}^{l_j}]\right)\right)\right) \tag{2.7}$$

where j indexes the possible sensing locations, and l_j indexes the possible stopping actions for the sensing location j.

Note that the value function computation does not involve any recursion, just a comparison between simple-to-compute action values for different actions. Furthermore, this online policy only needs to be calculated for the current belief state, making the computational complexity to be constant time. For the visual search problem considered below, because the stopping action is restricted to only the current sensing location, $l_j = j$,

the right-hand side simplifies to

$$V(\mathbf{p},k) = \min\left(\left(1-\mathbf{p}^k\right), \min_j \left(c + c_s \mathbf{1}_{\{j\neq k\}} + \left(1-\mathbb{E}[\mathbf{p}^j]\right)\right)\right)$$
$$= \min\left(\left(1-\mathbf{p}^k\right), \min_j \left(c + c_s \mathbf{1}_{\{j\neq k\}} + \left(1-\mathbf{p}^j\right)\right)\right) \tag{2.8}$$

the last equality due to $\mathbf{p}$ being a martingale. It can be seen that this myopic policy overestimates the size of the stopping region: if there is only step left, it is never optimal to continue looking at the same location, since such an action would not lead to any improvement in expected accuracy, but incur a unit cost of time c. We present a novel theoretical result that gives an inner bound of the stopping region, which is sensitive to the sampling cost c and the signal-to-noise ratio of the sensory input.

Assuming that the observations are binary and Bernoulli distributed which is a sensible distribution for active sensing task accumulating evidence (positive and negative):

$$f_{s,\lambda}(x) = p(x|s=i;\lambda=j) = \mathbf{1}_{\{i=j\}}\beta^x(1-\beta)^{1-x} + \mathbf{1}_{\{i\neq j\}}(1-\beta)^x\beta^{1-x} \tag{2.9}$$

We can state the following result:

Theorem 1. *If p^* is the solution of the equation:*

$$p\frac{(2\beta-1)(1-p)}{\beta p + (1-\beta)(1-p)} = c$$

where c is the cost per unit time as defined in sec. 2.2.2, then for all $\mathbf{p}^i > p^*$, the optimal action is to stop and declare location i under the cost formulation of C-DAC.

Proof. The cost incurred for collecting each new sample is c. Therefore stopping is optimal when the improvement in belief from collecting another sample is less than the cost incurred to collect that sample. Formally, stopping and choosing i is optimal for the

corresponding belief $\mathbf{p}^i = p$ when:

$$\max_{p' \in \mathscr{P}} (p') - p \leq c$$

where $\mathscr{P}$ is the set of achievable beliefs starting from p. Furthermore, if we solve the above equation for equality, to find p^*, then by problem construction, it is always optimal to stop for $p > p^*$ (stopping cost $(1-p) < (1-p^*)$). Given the likelihood function $f_{s,\lambda}(x)$ (eq. 2.9), we can use eq. 4.7 to simplify the above relation to:

$$p \frac{(2\beta - 1)(1-p)}{\beta p + (1-\beta)(1-p)} = c$$

□

Fig. 2.6 shows the myopic C-DAC policy augmented with the stopping threshold bound for the environment $(c, c_s, \beta) = (0.1, 0, 0.9)$ and $(0.1, 0.1, 0.9)$.

2.2.6 Experimental Design

The data are from eleven subjects, recruited from the UCSD undergraduate students (five females). Subjects first performed a random-dot coherent motion direction discrimination task (Britten et al., 1992) training session and achieved an accuracy exceeding 75% for 12%-coherence stimuli, before continuing onto the main experiment. In the main visual search experiment, subjects must identify one of the three random-dot stimulus patches as the target (left-moving for five subjects, right-moving for six subjects), the other two being distractor stimuli moving in the opposite direction. Subjects began each trial by fixating a central cross, then sequentially fixated one or more stimulus patches until pressing a space bar, which indicated that the last viewed stimulus was the chosen target. The three stimulus patches were circular and equidistant from the central

cross, rotationally symmetrically positioned at non-cardinal angles.

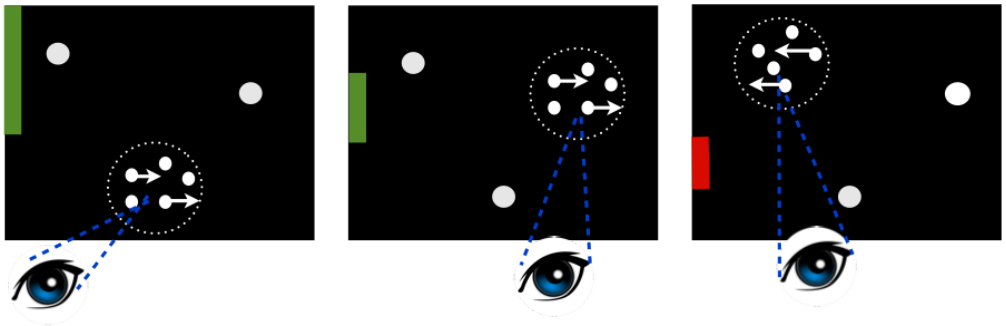


Figure 2.2. Simple visual search task, with gaze contingent display.

In the 1:1:1 condition (2 blocks), the target appeared in the three locations with equal likelihood on each trial; in the 1:3:9 locations (6 blocks, one of each possible configuration), the target appeared in the three locations with correspondingly biased probabilities. The order of the eight blocks (six biased blocks and two uniform ones, 90 trials per block) were randomized for each subject. Before the main experiment, subjects experienced 3 practice blocks: respectively, they consisted of 30, 40, and 40 trials, each with target location distribution drawn randomly from the configuration in the main experiment (2/8 probability of a 1:1:1 block, 1/8 probability of each of the 6 1:3:9 blocks). The random-dot motion coherence of the three blocks were 30%, 20%, and 12%, respectively. A target identification accuracy of 80% had to be reached in the first two practice two blocks, or else the same block has to be repeated; similarly, in the third practice block, an accuracy of 68% had to be reached before the subject can proceed to the main experiment. Other than experiencing practice blocks with similar statistics as in the main experiment, subjects did not receive explicit instructions on the spatial distribution of target location. Fig. 2.2 shows schematic illustration of the task at three time points in a trial.

The gaze-contingent display only revealed a motion stimulus in the fixated location, with the remainder replaced by a central dot; boundaries for fixation determining

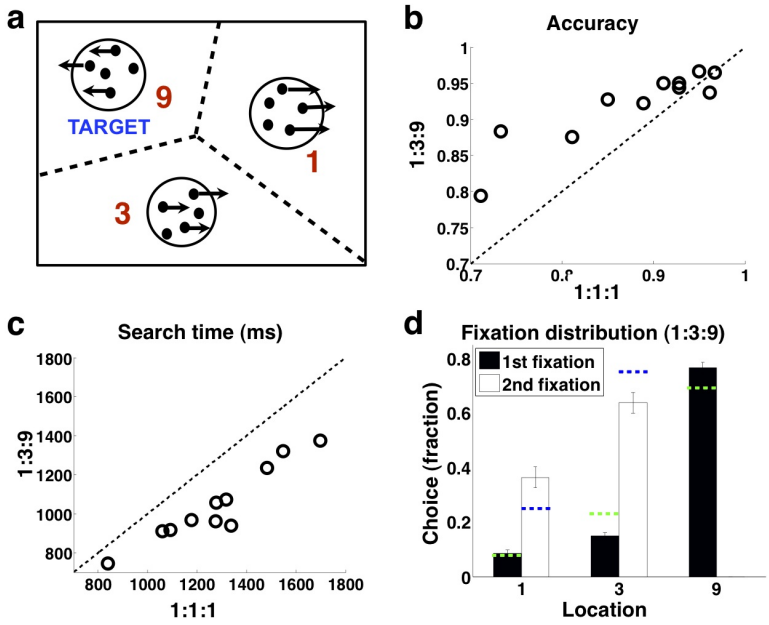


Figure 2.3. Experimental design and data. (a) On each trial, two of the random-dot stimuli are distractors, one is the target; subjects must find the target (see Methods). (b) Subjects are more accurate in finding the target in the 1:3:9 condition than the 1:1:1 condition, and (c) faster. (d) 1:3:9 condition, allocation of fixation location on first fixation (black), and second fixation when subjects first fixated the 9 location and found that it was not the target (white), averaged over all subjects. Green dashed lines indicate the matching probabilities on the first fixation, $(1/13, 3/13, 9/13)$; blue dashed lines indicate matching probabilities on the second fixation, $(1/4, 3/4)$. $n = 11$. Errorbars: s.e.m. across subjects

which stimulus was shown at any given time are as shown in Fig. 2.3a. Subjects' eye movements were monitored using a SR Research Eyelink 1000 eye tracker. A timing bar on the left side of the screen indicated time elapsed since onset of first stimulus fixation, first decreasing in length (green) until 8 seconds elapsed, and then growing in length in the opposite direction (red) at the same rate, though subjects were told that points were deducted indefinitely in proportion to their total response time (12.5/sec). At the end of each trial, subjects were shown the true target location and the total points gained/lost for that trial: 100 - 12.5 $\times$ (seconds taken to respond) + 25$\times$(number of fixation switches, 0 if only one patch fixated) $\pm$ 50 (+ if final response correct, - if incorrect). Subjects

were paid at the end of the experiment proportional to the total points earned, which were calibrated so as to award the average subject about 10 an hour.

All subjects have normal or corrected vision. They were excluded from the analyses if they achieved less than 50% accuracy in the main experiment (lower than in the practice blocks), or showed unusually large first fixation spatial bias (>2 standard deviations away in Kullback-Leibler divergence from the population mean distribution of first fixation, in the 1:3:9 condition).

2.3 Results

In this section, we will first present the different policies, followed by behavioral modeling, model predictions and potential extensions.

2.3.1 Policy Comparison

Before we discuss the performance of different models in terms of the behavioral output, we first visually illustrate the decision policies (Fig. 2.4). The belief state $\mathbf{p}$ is represented by discretizing the two-dimensional belief state space $(\mathbf{p}^1, \mathbf{p}^2)$ with $m = 201$ bins in each dimension ($\mathbf{p}^3 = 1 - \mathbf{p}^1 - \mathbf{p}^2$). Although for C-DAC the policy also depends on the current fixation location, we only show it for fixating the first location; the other representations being rotationally symmetric. In Fig. 2.4, the parameters used for the C-DAC policy are $(c, c_s, \beta) = (0.1, 0, 0.9)$, and for the statistical policies, $(\beta, thresh) = (0.9, 0.8)$. The color encodes the action prescribed by the policy at any given belief state, such that blue stands for stop and declare the current location as target, green for fixate location 1, orange for fixate location 2 and brown for fixate location 3. Note that for this simple scenario with no switch cost, the infomax policy looks almost like the C-DAC policy – fixate the most likely location unless there is very strong evidence that the fixated location contains the target, in which case the observer should

stop. The greedy MAP policy, on the other hand, looks completely different, and is in fact *ambiguous* in the sense that for a large set of belief states the policy does not give a unique next fixation location. We show one instance of this seemingly random policy, and note that there are regions where the policy suggests to look at either location 1 or 2 or 3 (corner regions speckled with green, orange and brown). Similarly, there are regions where the policy suggests to look at 1 or 2 (green+orange region). In fact, we exclude Greedy-MAP from our subsequent comparisons altogether since it resembles a random policy for the simple visual search problem we consider (see appendix for proof).



Figure 2.4. Decision policies – Infomax resembles C-DAC. Blue: stop. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3. Environment $(c, c_s, \beta) = (0.1, 0, 0.9)$. Threshold for infomax and greedy MAP = 0.8

Fig. 2.6 shows the effects of how the C-DAC policy changes when different parameters of the task are changed. As seen in the figure, the stopping region expands if the cost of time increases (high c), intuitively this makes sense – if each time step is costlier then the observer should stop at a lower level of confidence, at the expense of higher error rate. Similarly, for the case when β is smaller (high noise), stopping with a lower level of confidence makes sense – the value of each additional observation depends on how noisy the data is, the noisier the less worthwhile to continue observing, thus leading to a lower stopping criterion. Lastly, and arguably the most interesting case, is when there is an additional switch cost (added c_s); this deters the algorithm from

switching even when the belief in a given location has reduced below $1/3$. In fact, this is the scenario where optimizing for behavioral objectives turns out to be truly beneficial, and although infomax can approximate the C-DAC policy when the switch cost is 0, it cannot do so when switch cost comes in to play.



Figure 2.5. C-DAC policy for different environments (c, c_s, β) – high cost of time, c $(0.2, 0, 0.9)$, high noise $(0.1, 0, 0.7)$, and added switch cost, c_s $(0.1, 0.1, 0.9)$.

Next we look at the policy plots for different approximations and how they compare with the exact policies. Fig. 2.6 shows different policies for a behavioral task on two environments: $(c, c_s, \beta) = (0.1, 0.1, 0.9)$ and $(0.1, 0.1, 0.9)$. For example, as expected from any sensible policy, if the belief state is $(1, 0)$ while looking at location 1, the suggested action is to stop and declare it as the target. Although the belief state is continuous, we discretize over a grid of size 201 for implementation.

We note that the parameter that accentuates the major difference between C-DAC and Infomax is the switch cost, c_s, since its introduction leads to a shift in action boundaries. Unlike Infomax, the approximations we introduce can be seen to be sensitive to this cost and can modulate their switching boundaries. The boundary shift is less pronounced in Myopic C-DAC because of its greedy nature which only allows for incorporating one switch at the most, whereas in C-DAC and other approximations, a trial can potentially have multiple switches, leading to accumulating cost. In addition, Infomax does not adjust the size of its stopping boundary, unlike C-DAC and its parametric

approximations. We used 49 Gaussian kernel bases for RBF (Radial Basis Functions) approximation, centered so as to uniformly grid the belief space ($\sigma^2 = 0.05$), and 200 points to approximate the value function over. GPR (Gaussian Process Regression) approximation also uses 200 points to approximate the value function, and the parameters used are length scale 1, signal variance 1 and noise variance 0.1. GPR with ARD (Automatic Relevance Determination) uses 200 points as well and learns the parameters simultaneously with the value function.

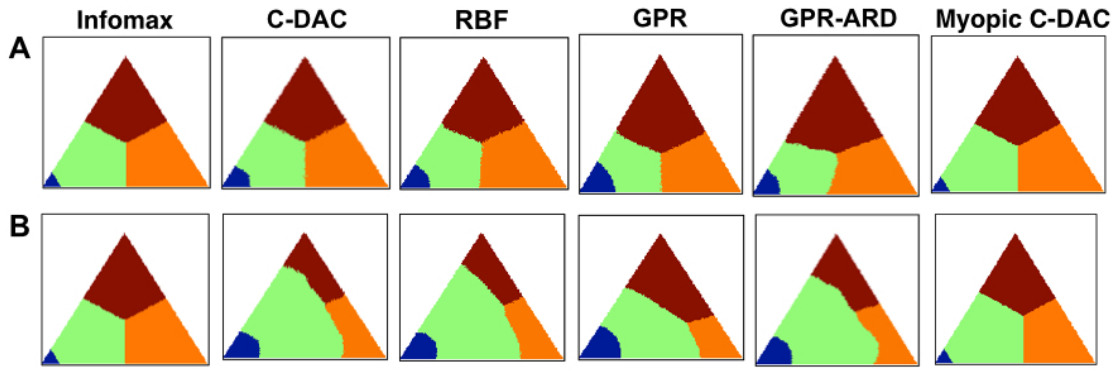


Figure 2.6. Decision policies as a function of switching cost. (A) No switch-cost environment: $(c, c_s, \beta) = (0.1, 0, 0.9)$. (B) Switch-cost environment: $(c, c_s, \beta) = (0.1, 0.1, 0.9)$. Triangle represents the belief state space, $(\mathbf{p}^1, \mathbf{p}^2)$, under affine transform for visual symmetry. Color-coding denotes the action prescribed by each policy for different values of the belief state vector. Blue: stop & declare location 1 as target. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3. All policies are deterministic (see Methods).

2.3.2 Behavioral Comparison

We first briefly describe the visual search task (see Methods, also (Yu and Huang, 2014)), before delving into the experimental findings and comparison to the various models. In the task (Fig. 2.3a), subjects must find a target stimulus (random-dot motion stimulus moving in a certain direction) in one of three possible locations, with the other two locations containing distractors (random-dot motion stimulus moving in the opposite direction). In the 1:3:9 condition, the target location is biased among the three options

with 1:3:9 odds. In the 1:1:1 control condition, the target appears in the three locations with equal probability on each trial. To eliminate the complications associated with the spatiotemporal dynamics of covert attention, which we cannot measure directly, the display is gaze-contingent: only the fixated stimulus is visible at any given time, with the other two stimuli being replaced by two small dots located at the center of the stimulus patches. Subjects receive feedback about true target location on each trial after making their choice, as well as their choice accuracy, search duration, and number of switches; they are encouraged through a point-based reward function to be fast, accurate, and efficient with the number of fixation switches (see Methods).

As shown in Fig. 2.3 and reported in (Yu and Huang, 2014), human subjects appeared to internalize and exploit the spatial statistics to locate the target stimulus more accurately (Fig. 2.3b) and more rapidly (Fig. 2.3c). Subjects were more accurate in finding the target in the 1:3:9 condition than the 1:1:1 condition (one-sided t-test, $p < 0.01$), and faster at finding the target than in the 1:3:9 condition than the 1:1:1 condition ($p < 0.001$). Underlying this performance improvement was a prioritized search strategy that favors the more probable locations as a fixation choice (Fig. 2.3d). For first fixation, subjects preferentially fixated the 9 location over the 3 location ($p < 0.0001$), which in turn was favored over the 1 location ($p < 0.01$). For second fixation, on trials in which the first fixation was at 9 and that was *not* the target, subjects then favored the 3 location over the 1 location ($p = 0.005$).

As we reported elsewhere (Yu and Huang, 2014), subjects first-fixation choices reflect a dynamically updated Bayesian prior belief, related to the true underlying spatial distribution of target location as well as incidental stimulus history experienced in the recent past. In this work, we focus on how that prior belief on each trial modulation sensory processing and perceptual decision-making within each trial, as reflected in the fine temporal dynamics of fixation patterns and durations. As shown in Fig. 2.7A,

subjects exhibit a certain confirmation bias in their fixation patterns. When viewing the same sensory input, subjects were more likely to identify the stimulus as a target if it is in the 9 location rather than in the 1 or 3 locations, thus exhibiting both a higher hit and false alarm rate in the 9 location compared to the other locations. Moreover (column 3 and 4), on correct trials, subjects appear to need *less* sensory evidence to confirm a stimulus as the target but *more* evidence to reject it as a distractor in the more probable location: subjects spent less time viewing a *target* stimulus in the 9 patch than the 1 and 3 patches; conversely, they spent longer viewing a *distractor* stimulus in the 9 patch than the 1 and 3 patches. To isolate the effect of prior spatial information on perceptual processing, and thus any indirect influence of prior information on motor response (evident in Fig. 2.3d), which in turn may affect sensory processing, all the data shown here (Fig. 2.7A) only include trials where subjects first fixated the labeled patch, i.e. the 9 patch in the "9" column and the 1 or 3 patch in the "1&3" column. Data for 1 and 3 patches were combined together due to the small number of trials in which subjects first fixated one of these locations.

We consider two different models for human eye movements: C-DAC and Infomax. We do not present results for Greedy MAP which is highly sub-optimal and behaves like a random policy for this task (see Appendix). C-DAC uses iterative Bayesian inference to update the belief about the target location based on sequential sensory data, and makes decision about how long to view a location before switching to another one or choosing a terminal response, so that the overall *behavioral cost* is minimized (a linear combination of the cost of time, the cost of making a switch among locations, and the cost of making an incorrect response). The Infomax formulation also employs Bayesian belief updates about target location, but makes fixation decisions so as to maximize the anticipated cumulative future entropy reduction about the target location.

We find that the confirmation bias phenomena exhibited by humans are captured

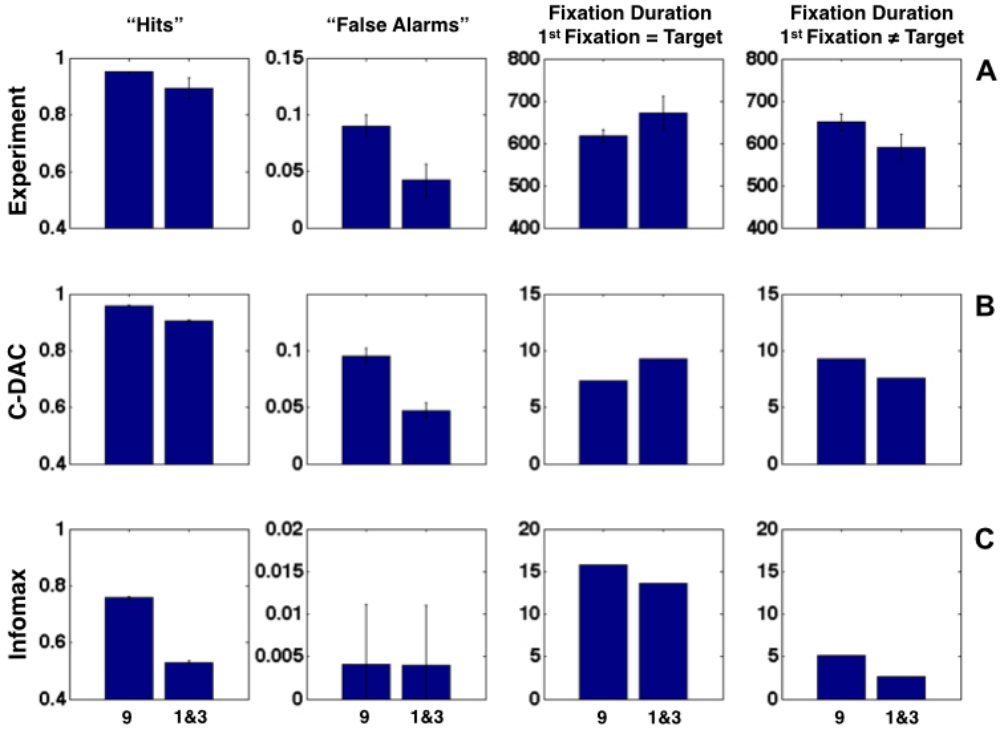


Figure 2.7. The influence of prior spatial knowledge on sensory processing and decision-making, in humans and models. (A) In terms of perceptual choice, subjects are more likely to choose the most probable target location ("9") than the less probable locations ("1&3") as containing the target, whether it actually contains the target ("hits", first column) or not ("false alarms", second column). In terms of fixation duration (milliseconds), subjects are faster to accept the most probable location, compared to the less probable locations, as containing the target, when the fixated location contains the target (third column); they are slower to accept the most probable location, as compared to the less probable locations, as *not* containing the target, when the fixated location does *not* contain the target (fourth column). All data included in the plots are from trials where the subject first fixated the labeled location. The data for the "1" and "3" locations are combined due to the relatively low incidence of subjects' first fixating these locations. (B) C-DAC reproduces the same pattern of behavior as human subjects. Fixation duration in number of simulation steps. (C) Infomax fails to reproduce the same pattern of behavior as human subjects. Simulation settings: $(c, c_s, \beta) = (0.005, 0.1, 0.68)$.

by C-DAC (Fig. 2.7B), but not by Infomax (Fig. 2.7C). This is due to Infomax's insensitivity to the cost of switching sensing location, parameterized by c_s in C-DAC. As seen in Fig. 2.6, the inability to incorporate switching cost makes Infomax decision policy quite different from the optimal policy (C-DAC). In C-DAC, larger c_s induces increasingly more asymmetric decision regions, whereby the optimal policy becomes more reluctant to switch away from the currently fixated location, despite sensory evidence pointing to the contrary. This would come into play when the observer is fixated in a stimulus location that does *not* contain the target (Fig. 3, column 4), as the switching cost accentuates a reluctance to switch away from this location, especially when it is *a priori* perceived to be more likely to contain the target. Separately, larger c_s also increases the size of the stopping region (Fig. 2.6), as the incentive of switching to other location just to "make sure" the currently fixated location contains the target diminishes, and the optimal policy instead chooses to terminate observations and select the current location as the target even for intermediate levels of confidence. This is why C-DAC has shorter fixation duration on target trials for the "9" location, as the higher prior probability, combined with the preliminary sensory evidence, is sufficient to push the optimal policy into the stopping region, whereas the lower prior confidence accorded to "1" and "3" locations does not succeed in doing so. In contrast, Infomax in its original formulation (Butko and Movellan, 2008) has no principled stopping policy at all, and in its augmented form here (Theorem 1) is still insensitive to c_s and moreover under-estimates the size of the stopping region (see Fig. 2.6). Thus, not only does Infomax fixate longer than C-DAC, in general, before stopping on the target trials (Fig. 3, column 3), but the higher prior for the "9" location is insufficient, combined with sensory data, to push it earlier into the stopping region than for the "1" and "3" locations. Altogether, the results suggest that humans, like C-DAC, modify the fine temporal dynamics of their visual search behavior in a way that is apparently sensitive to the switch cost, which may arise from intrinsic temporal or

energetic cost associated with saccades, and/or due to the experimental design, which levies 25 points per fixation switch in the task (see Methods).

2.3.3 Model Predictions

In this section, we describe future experiments that can further differentiate the various eye movement models, by considering scenarios in which Infomax and C-DAC are predicted to have particularly divergent behavior. We showed in the previous section that the switch cost may play a key role in modulating human visual search behavior. We therefore use model simulations to predict how behavior should change depending on whether the switching between fixation locations is penalized or not (e.g. in our experiment design, it can be done by changing the number of points charged for fixation switch, which results in greater monetary penalty).

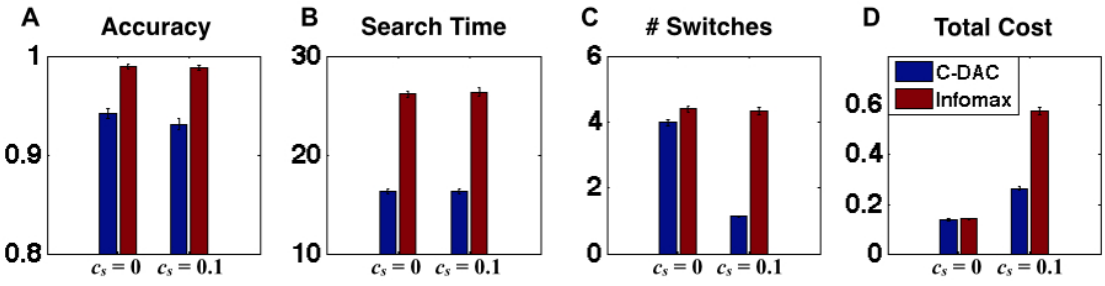


Figure 2.8. Behavioral predictions as a function of switching cost. (A) Accuracy, (B) search time, (C) number of switches, and (D) total behavioral cost, for C-DAC and Infomax for two environments $(c, c_s, \beta) = (0.005, 0, 0.68)$ and $(0.005, 0.1, 0.68)$. For $c_s > 0$, C-DAC reduces the number of switches and incurs overall less behavioral cost.

As can be seen in Fig. 2.6, explicit manipulating switch cost leads to rather different control policies. This should presumably lead to rather different behavioral outcomes as well. As shown in Fig. 2.8A, for the case $c_s = 0$, the accuracy of C-DAC is poorer as compared to Infomax, because the threshold used for Infomax (based on Theorem 1) is more conservative (thus stopping and declaration happens at higher confidence, leading to higher accuracy), but C-DAC takes less time (Fig. 2.8B) and

fewer number of switches (Fig. 2.8C) to reach the decision. Looking at the various behavioral costs, we see that although C-DAC loses in accuracy, it makes up through other measures, leading to slightly lower total cost (Fig. 2.8D). For the case when $c_s = 0.1$, the accuracy and search time for both policies are relatively comparable to the case with $c_s = 0$. However, C-DAC has a notable advantage in terms of number of switches, while the number of switches remains unchanged for Infomax (Fig. 2.8A and B). This case exemplifies the context-sensitivity of C-DAC, as it reduces number of switches when switching becomes more costly (Fig. 2.8C). When all these costs are combined, we see that C-DAC incurs a lower cost than Infomax (Fig. 2.8D). These observations can be used to aid the design of future experiments, to gain a fuller understanding of human eye movement behavior.

2.3.4 Visual Search with Peripheral Vision

Another way to extend our work is to include peripheral vision in future experiments. In the very simple three-location visual search problem we considered above, we did not incorporate the possibility of peripheral vision, or the more general possibility that a sensor positioned in a particular location can have distance-dependent, degraded information about nearby locations as well. We therefore consider a simple example with peripheral vision (see Fig. 2.9B), whereby the observer can saccade to intermediate locations that give reduced information about either two (sensing locations on the edges of the triangle) or three (sensing location in the center) stimuli. This is motivated by experimental observations that humans not only fixate most probable target locations but sometimes also center-of-gravity locations that are intermediate among two or more target locations (Findley, 1982; Zelinsky et al., 1997).

Formally, we need an *acuity map*, the notion that it is possible to gain information about stimuli peripheral to the fixation center (fovea), such that the quality of

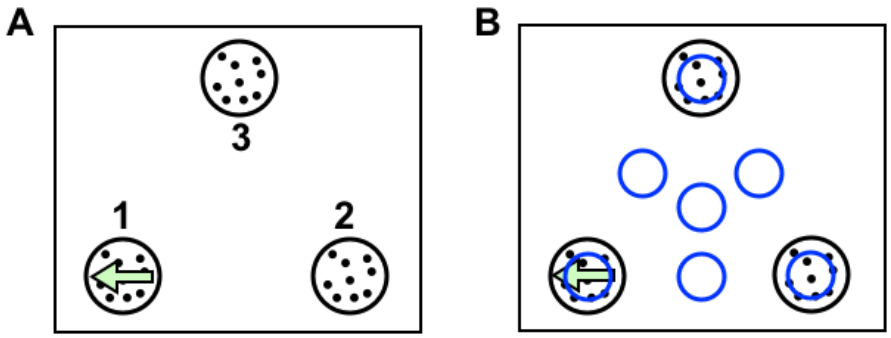


Figure 2.9. Schematics of visual search task. The general task is to find the target (left-moving dots) amongst distractors (right-moving dots). Not drawn to scale. (A) Task 1: agent fixates one of the target patches at any given time. (B) Task 2: agent fixates one of the blue circle regions at any given time.

that information decays at greater spatial distance away from the fovea. For example, the task of Fig. 2.9B would require a continuation action space of 7 elements, $L = \{l_1, l_2, l_3, l_{12}, l_{23}, l_{13}, l_{123}\}$, where the first three actions correspond to fixating one of the three target locations, the next three to fixating midway between two target locations, and the last to fixating the center of all three. We parameterize the quality of peripheral vision by augmenting the observations to be three-dimensional, (x^1, x^2, x^3), corresponding to the three simultaneously viewed locations. We assume that each x_i is generated by a Bernoulli distribution favoring 1 if it is the target, and 0 if it is not, and its magnitude (absolute difference from 0.5) is greatest when observer directly fixates the stimulus, and smallest when the observer directly fixates one of the other stimuli. We use 4 parameters to characterize the observations ($1 > \beta_1 > \beta_2 > \beta_3 > \beta_4 >= 0.5$). So, when the agent is fixating one of the potential target locations (l_1, l_2 or l_3), it gets an observation from the fixated location (parameter β_1 or $1 - \beta_1$ depending on whether it is the target or a distractor), and observations from the non-fixated locations (parameter β_4 or $1 - \beta_4$ depending on whether they are a target or a distractor). Similarly, for the midway locations (l_{12}, l_{23} or l_{13}), the observations are received for the closest locations (parameter β_2 or $1 - \beta_2$ depending on whether they are a target or a distractor), and from

the farther off location (parameter β_4 or $1-\beta_4$ depending on whether it is the target or a distractor). Lastly, for the center location (l_{123}), the observations are made for all three locations (parameter β_3 or $1-\beta_3$ depending on whether they are a target or a distractor). Furthermore, since the agent can now look at locations that cannot be target, we relax the assumption that the agent must look at a particular location before choosing it, allowing the agent to stop at any location and declare the target.

We first present the policies, and, similar to our discussion of simple visual search task, we only show the C-DAC policy looking at the first location (l_1) (the other fixation-dependent policies are rotationally symmetric). It is evident from Fig. 2.10 that now the C-DAC policy differs from the infomax policy even when no switch cost is considered, thus pointing to a more fundamental difference between the two. Note that for the parameters used here, C-DAC never chooses to look at the center l_{123}, but it does so for other parameter settings (not shown). Infomax, however, never even looks at the actual potential locations, favoring only midway locations before declaring the target location.

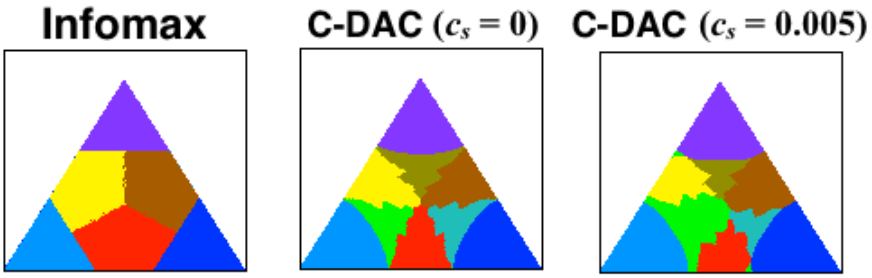


Figure 2.10. Decision policies. Azure: stop and choose location l_1. Blue: stop and choose l_2. Indigo: stop and choose l_3. Green: fixate l_1. Sea-green: fixate l_2. Olive: fixate l_3. Red: fixate l_{12}. Brown: fixate l_{23}. Yellow: fixate l_{13}. Environment $(c, \beta 1, \beta_2, \beta_3, \beta_4)$ = $(0.05, 0.62, 0.6, 0.55, 0.5)$. Threshold for infomax = 0.6

For performance comparison in terms of behavioral output, we again investigate two scenarios: (1) no switch cost, (2) with switch cost. The threshold for infomax is set so that the accuracies are matched to facilitate fair comparison. For all simulations, the

algorithm starts with uniform prior ($\mathbf{p} = (1/3, 1/3, 1/3)$) and initial fixation at the center (location l_{123}), while the true target location is uniformly distributed. Fig. 2.11 shows the accuracy, number of time steps, and number of switches for both scenarios. Now we notice that C-DAC outperforms infomax even when switch cost is not considered, in contrast to the simple task without peripheral vision (Fig. 2.8). Note however that C-DAC makes more switches for $c_s = 0$, which makes sense since switches have no cost, and search time can potentially be reduced by allowing more switches. However, when we add a switch cost ($c_s = 0.005$), C-DAC significantly reduces number of switches, whereas infomax lacks this adaptability to a changed environment.

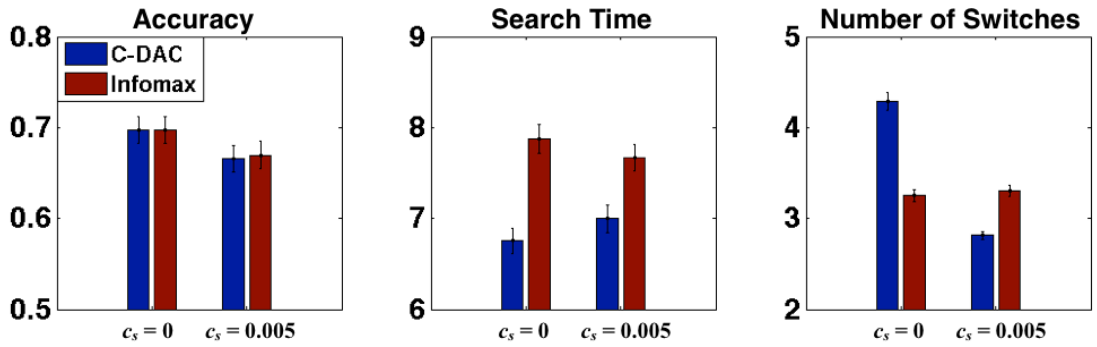


Figure 2.11. Comparison between C-DAC and Infomax on Task 2 for two environments $(c, \beta 1, \beta_2, \beta_3, \beta_4) = (0.05, 0.62, 0.6, 0.55, 0.5)$, $c_s = 0$ and $c_s = 0.005$. C-DAC adjusts the time steps and number of switches depending on the environment, taking a little longer but reducing number of switches when effort has cost.

2.4 Discussion

In this chapter, we proposed a Bayes-risk minimization framework for active sensing, C-DAC, which optimizes behaviorally relevant objectives in expectation, such as speed, accuracy, and switching efficiency, and can reproduce the confirmation bias. We also presented a novel visual search experiment that involves finding a target stimulus amongst distractors, whereby the spatial distribution of the target stimulus is biased toward certain locations. We found that the fixation and choice behavior of subjects

is modulated by top-down factors, specifically the likelihood of a particular location containing the target. Subjects were found to exhibit a certain confirmation bias – the tendency to systematically favor a location that is a priori judged more likely to contain the target, compared to another location less likely to contain the target, even in the face of identical sensory input and motor state. C-DAC was able to reproduce this confirmation bias. In contrast, Infomax, a policy that aims to optimize statistical objectives of task demands and ignores behavioral constraints (e.g. cost of time and switch), falls short. We augmented Infomax by introducing a stopping policy, which takes into account the time or sampling cost c, but that still did not sufficiently alleviate the context-insensitivity of Infomax. This is most likely due to both a sub-optimal incorporation of sampling cost and an intrinsic lack of sensitivity toward switching cost.

While C-DAC does a good job of reproducing human behavior, at least based on the experimental data presented here, we note that this does not necessarily imply that the brain implements C-DAC exactly. In particular, solving C-DAC exactly using dynamic programming requires a representational complexity that scales exponentially with the dimensionality of the search problem (i.e. the number of possible target locations), thus making it an impractical solution for more natural and complex problems faced daily by humans and animals. However, it is possible that the underlying computations can be approximated efficiently, making the formulation psychologically more plausible. We proposed different approximate algorithms that reduce the computational complexity of C-DAC in its original formulation. The low dimensional approximations provide a very good approximation to C-DAC, but these are still offline and global, requiring the computation of the policy for all possible belief states before engaging in the task. On the other hand, the myopic approximation to C-DAC we proposed scales linearly with search dimensionality, by eschewing a globally optimal solution that must be computed and maintained offline, in favor of an online, approximately and locally optimal solution.

However, this myopic algorithm does not approximate C-DAC as well as the low-dimensional Q-factor approximations. One way to improve the myopic algorithm is to find better approximations to the switching and stopping boundaries without having to undergo the computationally expensive value iteration procedure, since these together completely characterize any decision policy. Our simulations (not shown here) suggest that there might be a systematic, monotonic relationship between the decision boundaries and the different cost parameters. We proposed one such bound on the stopping boundary here, and other approximate bounds have been proposed for similar problems (Naghshvar and Javidi, 2012). Future work is needed to find better approximations that improve performance without incurring significantly more computational cost, and which is closer to neural computation in natural vision.

There have been several other related active sensing algorithms that differ from C-DAC in their state representation, inference, control and/or approximation scheme. We briefly summarize some of these here. In (Darrell and Pentland, 1996), the problem of active gesture recognition is studied, by using historic state representation and nearest neighbor Q-function approximation. Sensing strategies for robots in RoboCup competition is studied in (Kwok and Fox, 2004), which uses states augmented with associated uncertainty and model-free Least Square Policy Iteration (LSPI) approximation (Lagoudakis and Parr, 2003). Context dependent goals are considered in (Ji et al., 2007) and (Naghshvar and Javidi, 2010). The former concentrates on multi-sensor multi-aspect sensing using Point Based Value Iteration (PBVI) approximation (Pineau et al., 2006). The latter aims to provide conditions for reduction of an active sequential hypothesis testing problem to passive hypothesis testing. A Reinforcement Learning paradigm where reward is not dependent on information gain but on how close a saccade brings the target to the optical axis has also been proposed (Minut and Mahadevan, 2001). Other control strategies like random search, sequential sweeping search, “Drosophila-inspired”

search (Chung and Burdick, 2007) and hierarchical POMDPs for visual action planning (Sridharan et al., 2010) have also been proposed. We choose infomax to compare our C-DAC policy against because, as a human-vision inspired model, it not only explains human fixation behavior on a variety of tasks, but also has cutting edge computer vision applications (e.g. the digital eye (Butko and Movellan, 2010)).

A related problem domain, not typically studied as POMDP or MDP, is Multi-Armed Bandits (MAB) (Gittins, 1979). The classical example of a MAB problem concerns with pulling levers (or playing arms) in a set of slot machines. The person gambling is unaware of the states and reward distribution of the levers, and has to figure out which lever to pull next in order to maximize the cumulative reward. Noting a correspondence between the ideas of pulling arms and fixating location, and between rewards and observations, the MAB framework seems to describe the active sensing problem. Concretely, given the locations fixated (arms played) so far, and the observations (rewards) received, how to choose which location to fixate (which arm to play) next. However, there are certain characteristics of the active sensing problem that make it difficult to study in a MAB framework as yet. Firstly, the problem is an instance of restless bandits (Whittle, 1988), where the state of an arm can change even when it is not played. In active sensing, the belief about a location being the target does change even when it is not fixated. Whittles index is a simple rule that assigns a value to each arm in a restless setting, and the arm with the highest value is then played. The rule is asymptotically optimal only for a sub-class of problems (e.g. (Washburn and Schneider, 2008) and (Liu and Zhao, 2010)), but not optimal in general. Secondly, the states of the arms in the active sensing task are correlated (the elements of the belief-state have to add up to 1). There is some work on correlated arms for specific structure of correlation, like clustered arms (Pandey et al., 2007) and Gaussian process bandits (Dorard et al., 2009), but so far there is no general strategy for handling this scenario.

Active learning is another related approach, with hypothesis testing as a sub-problem that is related to the problem of active sensing. The setting involves an unknown true hypothesis, and an agent that can perform queries providing information about the underlying hypothesis. The task is then to determine which query to perform next to optimally reduce the number of plausible hypothesis (version space). In active sensing however, although the belief about a hypothesis (target location) can become arbitrarily low, the number of plausible hypothesis does not reduce. This problem is investigated in (Golovin et al., 2010), and a near-optimal greedy solution is proposed along with performance guarantees. Besides the sub-optimality of the approach, the same test cannot be performed more than once (whereas in active sensing, one location can be fixated more than once). The lack of this provision stems from the fact that the noisy observations considered are actually deterministic with respect to a hidden noise parameter. Thus, as of yet it is hard to cast the active sensing problem in this framework.

In conclusion, although there is a rich body of literature on related problems in the application domain, our formulation stands out for its principled approach to incorporating contextual elements into the problem of active sensing. In general, the framework proposed here has the potential for not only applications in visual search, but a host of other problems, ranging from active scene categorization to active foraging. Its flexibility and robustness to different environments makes the framework an appealing choice for a variety of active sensing and related problems.

Acknowledgements

This chapter contains material from S. Ahmad and A.J. Yu, Active sensing as bayes-optimal sequential decision-making, Uncertainty in Artificial Intelligence (UAI), 2013; S. Ahmad, H Huang, and A.J. Yu, Context-sensitive active sensing in humans, Advances in Neural Information Processing Systems (NIPS), 2013; and S. Ahmad, H Huang,

and A.J. Yu, Cost-sensitive bayesian control policy in human active sensing, Frontiers in Human Neuroscience, 2014. The dissertation author was the primary investigator and author of these papers.

Chapter 3

Competitive Foraging

Effective foraging of food and other resources is critical for survival in many animal species. Each agent must solve the problem of *where* to forage and *when* to move on from one potential food patch to the next. The task is further complicated when there are other foraging agents, potentially competing for the same limited resources. Previous models of foraging consider foraging agents either in isolation or in groups but without competition. Here, we present a novel Bayesian model for competitive foraging, Socially Aware Bayesian Agent (SABA), that takes into explicit account the presence of other agents for both learning and decision-making. In terms of learning, other agents' choices confer valuable information about the reward abundance in each foraging patch, in addition to any personal choices and rewards experienced by an agent. In terms of the foraging decision, given the same beliefs about reward abundance, knowledge about how other agents will be distributed in the near future can increase personal gain by avoiding crowding. We incorporate these elements into SABA in a principled manner, and consider two self-consistent decision policies, a deterministic one and a stochastic one. In addition, we also implement a simple Naive Agent model that completely ignores other agents for both learning and decision making. We find that when the reward structure is stationary, all models perform well in simulation, converging quickly to the optimal population-level solution (also a Nash equilibrium), whereby agent distribution matches the underlying

reward distribution. However, when the environment is non-stationary, only SABA with the stochastic foraging policy can readily adapt to the new reward distribution. These results represent a first step toward cognitively sophisticated representations for learning and decision-making in competitive foraging.

3.1 Motivation

Foraging is the general problem of deciding how to choose among different reward options (e.g. food patches) adaptively, based on sequentially observed, noisy data, and thus in the presence of uncertainty associated with reward values of the different options. In competitive foraging, there is both benefit, in terms of information, and cost, in terms of competition, due to the presence of other foraging agents. Daily life is rife with such competitive foraging examples, such as picking a restaurant, asking someone out on a date, bidding on a house, picking a vacation location, etc. In the experimental literature, this has most often been studied in the context of animals foraging in an environment containing patches of food (see (Pyke et al., 1977; Perry and Pianka, 1997) for review). The optimal solution for the population is to have the agents distributed proportional to the underlying reward rates at the different food patches, as that ensures maximal efficiency of food intake in the population. This distribution is also known as the Ideal Free Distribution (IFD), and has been observed in humans (Sokolowski et al., 1999; Kraft and Baum, 2001; Madden et al., 2002; Goldstone and Ashpole, 2004; Goldstone et al., 2005) and other species such as ducks (Harper, 1982), guppies (Dill, 1987), sticklebacks (Milinski, 1979), Bumblebees (Dreisig, 1995), etc.

Interestingly, IFD is also the individually optimal Nash equilibrium solution (Nash et al., 1950), in which no agent can do better by moving to another patch. Competitive foraging, therefore, presents an ecologically convenient case study for how individual-based, decentralized learning and decision-making, can lead to coherent and desirable

behavioral patterns at the group level. Surprisingly, despite the rich body of empirical literature on the subject, the cognitive and neural mechanisms leading to IFD are not well understood. Most of the models (see (Pyke, 1984) for a detailed review) are descriptive in nature, and make unrealistic assumptions about foraging agents' knowledge about the reward environment and their behavioral constraints. In particular, agents are assumed to have precise and unbiased knowledge about the reward distribution (the *ideal* assumption), and that they can freely travel from one patch to another without any limitation or cost (the *free* assumption). In more natural scenarios, humans and animals often have to forage without (initial) knowledge about the relative reward value of the different options, and there may be costs associated with switching between options (e.g. time, energy, opportunity cost, money). A more complete understanding of competitive foraging behavior needs to incorporate these factors.

Recently, there has been a shift toward models that relax these strong assumptions, such as EPICURE (Roberts and Goldstone, 2006), Polyworld (Griffith and Yaeger, 2011), Avidian agents (Walker, 2011), Omega agents (Seth, 2011), and Dynamic system (Krivan et al., 2008). While these models have demonstrated that IFD can be achieved even when the ideal assumption is relaxed, they still make the free assumption. More importantly, these models are overly simplistic in their assumptions about each agent's representation of the environmental reward distribution and other competing agents. Most of these assume that learning only proceeds based on *private information*, i.e. an agent's own history of choices and reward outcomes (Griffith and Yaeger, 2011; Walker, 2011; Seth, 2011). EPICURE (Roberts and Goldstone, 2006) does consider the effect of competition but it just fits an ad-hoc, descriptive parameter corresponding to agent density, so that the competition has a unidirectional (negative for their data) effect on agent choices, failing to account for the fact that depending on agent's knowledge, at times high agent density might make an option more lucrative. A large body of empirical ecology research shows

that animals readily use *public information*, i.e. other agents' choices and possibly also their outcomes, in foraging (e.g. (Krebs, 1973; Boyd, 1988; Templeton and Giraldeau, 1995)), as well as in other biologically important behavior, such as mate selection, predator avoidance, path selection, etc. (see (Galef Jr and Laland, 2005) for review). Thus a psychologically plausible model of foraging should take public information into account.

A separate line of empirical and modeling studies have considered group foraging scenarios in which competition is not a factor (Green, 1980; Lima, 1985; Valone, 1989; Livoreil and Giraldeau, 1997; Olsson and Holmgren, 1998; Sernland et al., 2003), either because rewards are so abundant that there is no need for competition, or because the species under study is by nature highly cooperative. So although these studies take public information into account for learning, they do not, and have no need to, take into account other agents' anticipated behavior in making foraging decisions. There is also the caveat that these models assume that other agents' choice *outcomes* are publicly observable as well, whereas in many natural situations, perhaps only others' choices are observable and not the outcomes. Nevertheless, others' choices are still useful for learning, as they represent crowd-sourced information about the relative reward values for the different options.

Thus, there are two broad families of existing models for foraging, one considering competition but not the use of public information, and the other using public information but sidestepping competition. In this paper, we consider both the cost (competition) and benefit (public information) of foraging among competing agents. We propose a family of models, which we call SABA (Socially Aware Bayesian Agents), that takes into account: (1) knowledge about the fact that more competition at a given patch leads to lower reward, and (2) public information in the form of other agents' choices (and not their outcomes). We also consider two self-consistent decision policies, one that

makes deterministic choices (based on the highest "patch profitability ratio") and another that makes stochastic choices (proportional to the inferred "patch profitability ratios" of the different options). We also present a Naive Agent model, a Bayesian analogue of the existing competitive foraging models (like EPICURE, Polyworld, Avidian and Omega agents), which is oblivious of other agents both for learning and decision-making.

The rest of the chapter is organized as follows. We first survey some of the existing models of human foraging behavior in a competitive scenario. We then present SABA and the Naive Agent model, followed by simulation results on a simple two patch foraging problem, both for a scenario in which reward distribution is stationary over time, and another in which it is non-stationary. We conclude with a discussion of implications and limitations of our work, as well as pointers for future research.

3.2 Background

There is a plethora of empirical studies on the foraging behavior of different animals where the experiment design generally involves animals in a constrained environment with patches of food. However, there are only a handful of studies with human subjects. Since the recent surge in the models of competitive foraging comes from these human experiments, we first present a brief summary of the experiments followed by a brief description of some of the representative models.

3.2.1 Human Experiments

We describe two sets of foraging experiments here. In (Sokolowski et al., 1999; Kraft and Baum, 2001; Madden et al., 2002), different colored cards (red and blue) represent a "patch" such that each patch has fixed number of points that get divided equally between the participants who choose that patch. The experiment has different conditions depending on whether the participants are allowed to change their choice

within a trial or not. Note that each participant knows the quality of the patch (point associated with a color) a-priori, separating the decision policy from reward learning.

In (Goldstone and Ashpole, 2004; Goldstone et al., 2005), the food is distributed on a 2-D plane such that there are two centers, and the amount of food available at a given distance from a center follows a Gaussian distribution. The experiment is presented as a game to the participants on a computer screen. Depending on experiment conditions, the participants can see themselves (yellow dot), the competitors (blue dots), food (green dots) or any combination of these. Note that in several conditions, the food remains in a location until consumption, so that the instantaneous quality of the patch is different from the one intended experimentally.

Both these sets of human experiments discover that humans exhibit IFD, sometimes with slight under-matching, such that the richer patch has fewer agents than predicted by IFD. The authors for the second set of experiments also noted that there were population oscillations between patches, which could potentially be due to the fact that if a patch is even slightly underutilized, food would accumulate and at some point several agents would migrate to that patch, leading to the whole process getting repeated periodically. The authors found out that the best fitting model to human data assigned a negative weight to the crowd-density at a patch, leading to the claim that higher crowd-density at a patch is always deterrent to its value. Like we previously argued, with limited individual experience, higher crowd-density at a patch may signify it to be a rich patch. Further experiments are needed to better understand how humans utilize social information.

3.2.2 Existing Models

A few models have been proposed recently to explain the behavior that emerges when animals forage in a competitive environment. In the EPICURE model (Roberts

and Goldstone, 2006), each agent maintains a "value" of every location as a weighted sum of factors like distance, food density and agent density, and chooses a location probabilistically using Luce's choice rule (Luce, 1959). Along similar lines, Polyworld (Griffith and Yaeger, 2011), has agents that can see the instantaneous food availability in a location, moving to one with more food. Avidian agents (Walker, 2011) operate in a continuous environment, having short and long range sensors for food, and always moving in the direction of higher food patch. An agent can consume the food at any location, deterring other agents from moving to the same location. Omega agents (Seth, 2011) estimate the resources at a patch using exponential averaging of all the past rewards, and choose the next patch using an ε-greedy policy.

These models relax the assumption about agents having to know the quality of each patch a priori, but they assume that agents are free to move about and often assume that the environment is stationary. Furthermore, these models do not have a cognitively sophisticated way to incorporate the effects of competition. One of the terms in the value estimator of EPICURE is the level of competition, but it is unclear how to set the weight of that component, and such a formulation also leads to a fixed effect: positive weight means the presence of crowd is deemed as an indicator of higher location value, negative weight means it lowers the value of a location. However, a more sensible formulation should allow the agent to vary the direction and magnitude of the competition interaction term as a function of dynamically evolving factors, such as an agent's internal uncertainty about the reward distribution. In particular, we should expect that an agent who is highly unfamiliar/uncertain about an environment to take advantage of the social information to infer what the most rewarding options are (follow-the-crowd), and one who is highly familiar/confident about an environment to maximize competition avoidance (avoid-the-crowd).

3.3 Socially Aware Bayesian Agents (SABA)

We consider a scenario in which each agent decides which patch to forage next based on the history of individual reward and public information (crowd-density at different patches). There are n patches, and patch k has a (hidden) abundance level $\theta_k \in \{0,1,2,\ldots,(R-2),(R-1)\}$, where R is the number of discrete abundance levels. Similarly, patch k has (observed) crowd-density $\phi_k \in \{0,1,2,\ldots,(C-2),(C-1)\}$, where C is the number of discrete levels of crowd-density. There are two reasons for considering quantized levels of crowd-density instead of exact number of competitors in the model: (1) numerosity judgment noise scales with magnitude in humans and animals (Dehaene et al., 1998), (2) having arbitrarily precise numerosity information can easily overwhelm private information and lead to informational cascade (Bikhchandani et al., 1992).

3.3.1 Bayesian Inference

We use a Bayesian generative model to capture the agent's knowledge about the statistical nature of how reward is generated at a patch, and how crowd-density serves as an indicator of patch richness. We assume that the agents use Bayesian inference in the recognition model to update their belief about the hidden state of the world (patch richness).

At patch k, reward $x_t \in \{0,1\}$ is generated as:

$$p(x^t|\theta_k,\phi_k^t) \sim Bernoulli(x^t; 1-exp(-\theta_k/\phi_k^t)) \tag{3.1}$$

This generative model captures the idea that the probability of getting a reward ($x^t = 1$) increases with increasing richness level and decreases with increasing crowd-density, but the dependence is non-linear in both cases.

Each agent also needs a model for predicting how other agents will move, but

an exact model of such dynamics is inherently intractable due to the infinite recursion when an agent, a_i, tries to model other agents' beliefs about the world including a_i. Thus we need a model of crowd-dynamics that is tractable and captures the behavior approximately. We propose:

$$p(\phi_k^t|\theta_k,\phi_k^{t-1}) = \alpha \mathbf{1}_{\{\phi_k^t=\phi_k^{t-1}\}} + (1-\alpha)p'(\phi_k^t|\theta_k) \tag{3.2}$$

which asserts that at each time instant, the agent distribution can remain the same with probability α, or a redistribution around θ_k can occur with probability $(1-\alpha)$. The distribution p' is peaked at the hidden reward abundance levels θ_k; in the following simulations, we assume p' to be a discretized Gaussian with peak at θ_k. Eq. 3.2 captures our assumption that each agent assumes that the population distribution gradually tends toward the IFD from the current configuration, with the parameter α parameterizing how fast this convergence is expected to happen. Note that the equation does not imply that agents have any additional information about the underlying abundance level θ_k, only that agents assume a certain behavior of how other agents move, specifically that agent-density is likely to be (noisily) matched to the hidden abundance level.

In the recognition model, based on the observed reward history $\mathbf{x}_t$ at the chosen patch and the observed crowd density history Φ_k at all patches, each agent updates its belief about patch richness θ_k by applying Bayes' rule. Thus at the chosen patch:

$$p(\theta_k|\mathbf{x}^t,\Phi_k^t) \propto p(x^t|\theta_k,\phi_k^t)p(\phi_k^t|\theta_k,\phi_k^{t-1})p(\theta_k|\mathbf{x}^{t-1},\Phi_k^{t-1}) \tag{3.3}$$

And at all other patches:

$$p(\theta_k|\Phi_k^t) \propto p(\phi_k^t|\theta_k,\phi_k^{t-1})p(\theta_k|\Phi_k^{t-1}) \tag{3.4}$$

3.3.2 Decision Strategies

Since the goal of each agent is to maximize its individual gain, a behaviorally relevant strategy is to move to a patch which has the highest expected ratio of patch-richness to crowd density. Furthermore, relaxing the free assumption would imply that moving between patches has some cost. With these considerations, a deterministic strategy for next patch selection is:

$$k^{t+1} = \underset{k}{\operatorname{argmax}} \left\{ \frac{\mathbb{E}(\theta_k)}{\mathbb{E}(\phi_k^{t+1})} - c_s \mathbf{1}_{\{k \neq k^t\}} \right\} \tag{3.5}$$

The deterministic policy bases the decision on not only the agent's belief about patch richness but also about its belief about how other agents would move (eq. 3.2). Since the crowd dynamics model is only approximate, a disagreement between true dynamics and assumed dynamics could potentially lead to misinformed decisions. To avoid this, an agent can also employ a stochastic policy that only depends on its belief about patch richness, so that:

$$p(k^{t+1} = k) \propto \mathbb{E}(\theta_k) \tag{3.6}$$

Another motivation for using a stochastic strategy comes from the animal behavior data (Harper, 1982) where the equilibrium is dynamic in nature, i.e. agents (ducks) move between patches so that the number of ducks on any patch still approximates the IFD. A stochastic strategy does lead to a dynamic equilibrium as agents choose a location randomly and hence switch often, whereas in a deterministic strategy the choice of location becomes fixed once the system is in equilibrium. As we will see in the result section, a deterministic strategy can get stuck in a suboptimal solution.

3.4 Naive Agents

Several existing models of foraging (like (Roberts and Goldstone, 2006; Griffith and Yaeger, 2011; Walker, 2011; Seth, 2011)) propose agents that maintain an estimate of patch-richness based on their individual experience and employ a simple strategy for selecting the next patch. In order to compare SABA to this class of model, we propose agents within our Bayesian framework, that are agnostic about social information and crowd's effect on the reward. Furthermore, these agents greedily choose the next patch as the one which they believe to be the best (has highest expected value of patch richness). Such agents have a wrong statistical model of the environment, attributing the reward outcome solely to (fixed) patch-richness instead of attributing it to the (fluctuating) presence of competitors:

$$\hat{p}(x^t|\theta_k) \sim Bernoulli(x^t;\theta_k) \tag{3.7}$$

Furthermore, since individual reward is the only source of information, these agents only update belief about the attended patch where reward can be observed. After observing reward x_t, the belief is updated via Bayes rule:

$$p(\theta_k|\mathbf{x}^t) \propto \hat{p}(x^t|\theta_k)p(\theta_k|\mathbf{x}^{t-1}) \tag{3.8}$$

And the control strategy is:

$$k^{t+1} = \underset{k}{\operatorname{argmax}}\{\mathbb{E}(\theta_k)\} \tag{3.9}$$

3.5 Results

For ease of visualization, we consider a simple foraging problem over 2 patches, with 100 agents. We address more complex scenarios in the Discussion. The (hidden) patch richness $\theta_k \in \{0,1,2,\ldots,9,10\}$, and the (observed) crowd-density $\phi_k \in \{0,1,2,\ldots,9,10\}$. The α parameter in the approximate crowd-density distribution, eq. 3.2, is 0.6. Fig. 3.1 shows the behavior of SABA (Deterministic and Stochastic strategies), as well as Naive Agents for a stationary environment [7 3], i.e. the richness level of patch 1 (blue) is 7, and that of patch 2 (green) is 3. The dotted lines show the IFD, which is matched to patch richness, and is therefore also [7 3]. The solid lines show the simulated distributions when all the agents follow a given strategy. We note that agents following any of the three strategies achieve IFD. However, only the stochastic policy achieves a dynamic equilibrium whereas for the other two, the number of agents on a given patch remains constant once equilibrium is attained. Note that although it seems like that both Deterministic and Naive strategies under- or over-match, this is not the case, since the discretization of richness and crowd density enforces that the observed crowd-density is always an integer. For example, for the deterministic strategy, although there are 68 agents at blue patch, as far as the agents are considered it is crowd-density level 7, which is matched to the patch richness. Thus, just by looking at the aggregate agent behavior for this environment, it may seem that all three strategies are equally good, leading to the optimal solution.

In order to further differentiate the models and investigate the scenarios where considering social information may potentially be useful, we look at the beliefs that agents on different patches infer. In other words, we wish to understand whether the agents, following these different strategies, also correctly figure out the underlying richness of different patches, in addition to just the observed IFD behavior. This is important,

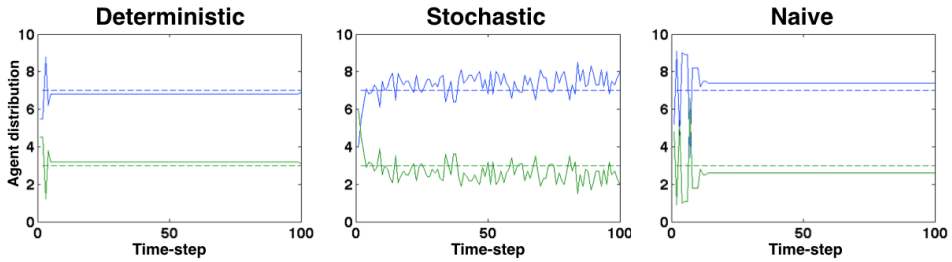


Figure 3.1. Stationary Environment: [7 3]. Dotted lines indicate the patch richness level (hidden). Solid lines indicate the normalized agent density on each patch, blue: patch 1 and green: patch 2.

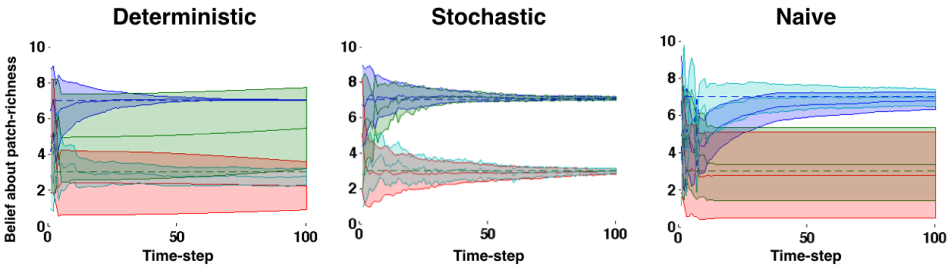


Figure 3.2. Inferred beliefs of agents at different patches. Dotted lines show the true richness level of the two patches. Blue: Belief about patch 1, of agent at patch 1. Green: Belief about patch 1, of agent at patch 2. Red: Belief about patch 2, of agent at patch 1. Cyan: Belief about patch 2, of agent at patch 2. The boundaries indicate the standard deviation of beliefs.

because an agent operating with incorrect belief (e.g. believing that all other patches are very low on resources) may never adapt to changes in the environment (e.g. always stays at the settled patch even when its own patch has deteriorated). Fig. 3.2 shows the average beliefs of agents in a given patch about the two patches. We note that the agents following the deterministic strategy infer the richness that is qualitatively accurate, i.e. both agents at patch 1 and patch 2 believe that patch 1 is richer, and similarly both believe that patch 2 is poorer, although the beliefs are not always strong. Agents following the stochastic strategy infer the quantitatively accurate richness, i.e. agents at both patches become confident that patch 1 has a richness level 7, and patch 2 has a richness level 3. On the other hand, agents following the Naive strategy fail to infer even the qualitative

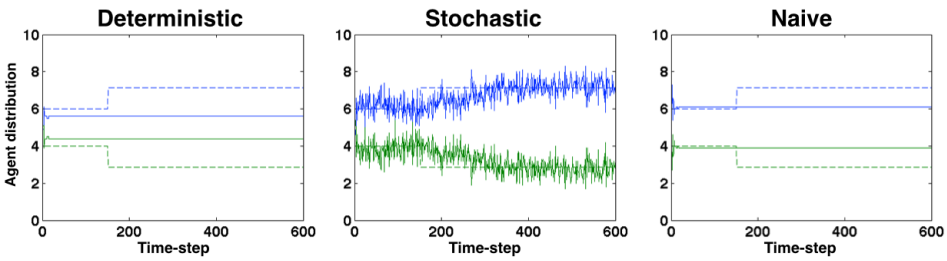


Figure 3.3. Non-stationary Environment: [6 4] → [7 3]. Dotted lines indicate the patch richness level (hidden) which changes at time-step 150. Solid lines indicate the normalized agent density on each patch, blue: patch 1 and green: patch 2.

trends, specifically agents at patch 2 settle on a belief that patch 2 is better than patch 1, opposite of the real richness levels. These observations motivated us to investigate whether the different strategies can perform well when the environment changes.

Fig. 3.3 shows the behavior of different algorithms for a non-stationary environment [6 4] → [7 3], the change occurring after 150 time-steps. For this case, the stochastic strategy is the only one that adapts, and achieves IFD in the changed environment, owing to its ability to infer the hidden richness correctly. The deterministic strategy suffers because it puts too much trust on other agents, both during inference and decision making, which leads to self-information being inundated by less accurate public information. The naive strategy suffers because it does not incorporate public information at all. The results shown suggest that a strategy like the stochastic one that uses public information, without letting it overwhelm self information, works the best.

3.6 Discussion

We introduced a Bayesian competitive foraging model, SABA, as well as a Naive Agent model. SABA takes into account both the self and public information, which is in accordance with the findings that animals use both of these sources for learning and decision-making while foraging. Introduction of stochasticity leads to an adaptive policy with dynamic equilibrium, as observed for foraging animals. Our

model is principled in that it only makes assumptions about the statistical nature of the environment, in contrast to the ad hoc parameter fitting needed by existing approaches. In a comprehensive review on the use of public information (Valone and Templeton, 2002), the authors pointed out two shortcomings of the state-of-the-art research in the field stemming from these assumptions: (1) each agent can not only observe all the other foragers but also whether they received a reward, (2) public information serves as an additional sample information and is given equal weighting as the self-information. We relax assumption 1 by allowing each agent to only observe the crowd-density at other patches instead of the quantity of food available or each forager's individual choice outcome. We relax assumption 2 by modeling public information as having a constant level of informativeness (discretization) regardless of the absolute number of competing agents. These also alleviate the problem of informational cascade since self information contributes more to the inference. Although, we show results for 2 patches, the model extends naturally to more patches, and the increase in complexity is linear owing to the inference on patches being decoupled. Simulations for 3 and 4 patches show that the model achieves IFD without high computational demands.

We believe our work is a step toward more sophisticated models of decision making in social scenarios. However, it also has limitations that leave scope for future inquiries, some of which we discuss here. The model of crowd-dynamics, eq. 3.2, is only approximate and needs to be revisited for more complex problems. This also means that we need ways to manage how much trust to place in public information, since not only the model of crowd-dynamics is approximate, an agent cannot trust other agents to be behaving optimally. Towards this trust management goal, we tried an inference method, called Perturbed Bayes (Jadbabaie et al., 2012), which provides a way to combine high quality self-information with the low-quality public information. However, the technique is ad-hoc and there is no principled way of deciding how to combine the differential-

quality information sources, making it only moderately useful for the problem. There are few other investigations in this domain, but it is largely an open problem. Another approach we tried, is to start doubtful about public information, and put incrementally more trust in the crowd, since the crowd is expected to have learned more about the environment as time passes. We implement this idea by decreasing the variance of the distribution p' in eq. 3.2 with time, but it is not clear how to set the rate of this decrease in a principled manner.

Another way to improve the performance in a non-stationary environment, is to use a Dynamic Belief Model (DBM) (Yu and Cohen, 2009), which assumes non-stationarity of the environment in the generative model, making agent more *aware* during inference that resource levels can change abruptly. Using a model such as DBM would improve not only the Naive Agents but also SABA, especially when combined with a deterministic decision policy, although the parameter specifying the rate of change would have to be higher in Naive Agent compared to SABA, as it would need to account for reward variability not only due to true reward non-stationarity but also net profitability due to the presence/absence of other agents.

Another direction for extension is to consider a scenario in which the reward options are located (embedded) over a terrain, and the hidden reward abundance level is an unknown continuous function over that terrain – this would also allow the possibility of including patch-switching cost that depends on the distance traveled between two options. However, this scenario poses serious tractability issues since the number of patches is not finite (or even countable) any more. One way to get around this is to use a parametric reward surface or a Bayesian non-parametric formulation such as a Gaussian Process (Williams and Rasmussen, 1996). Then each sample can be used to update the agent's beliefs about the entire terrain.

Yet another direction for future enquiry is to consider a collection of heterogenous

agents who may differ in capability, costs, and goals. Indeed, heterogeneity has been observed in animals, in which more dominant members may "despotically" monopolize some food patches and push the population distribution away from IFD (Parker and Sutherland, 1986).

To conclude, the problem of competitive foraging in its entirety is highly complex. This model represents a first step toward combining self and public information rationally to infer the hidden richness of different patches and using this inferred knowledge, along with beliefs about other agents' intended actions, to inform future decisions. And this individually optimal, decentralized foraging strategy naturally leads to IFD, the socially optimal solution for competitive foraging.

Acknowledgements

This chapter, in full, is a reprint of the material from S. Ahmad and A.J. Yu, A socially aware bayesian model for competitive foraging, Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2014. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Multi-attribute Decision Making

Choice behavior in humans has been known to exhibit curious contextual effects, whereby the relative choice preference between two multi-attribute options (like two cars with different economy and safety ratings) can change significantly, sometimes even reversing, when a third option is introduced into the choice set. The three most prominent effects being attraction, compromise and similarity. Relatedly, the inclusion of an unavailable (or sold-out) option, has also been observed to affect choices, known as phantom decoy effect. Commonly proposed explanations for these effects include appealing to the human irrationality in the domain or idiosyncrasies of the underlying neural mechanisms – both fairly strong assumptions. Furthermore, the attempts at directly modeling phantom decoy effect have been sparse. In this chapter we propose a rational framework that can explain all four aforementioned contextual effects while only making some simple and justifiable assumptions, the central one being that the options present in the choice set, including the sold out ones, provide information about the relative importance of the option attributes. We also discuss the generalizability of our model to non-consumer domains and discuss how to potentially extend it to model individual differences. Lastly, we augment the model to explain individual differences in preference choice as well as the evolution (and potential divergence) of individual preference through the consumption of chosen options.

4.1 Motivation

Humans are often faced with a choice between different options. These options can have multiple attributes, making the choice harder since it may be unclear as to how these attributes contribute to the net attractiveness of the option. For example, choosing between two smartwatches, one of that has high processing power but is low on economy vs. the other that is fairly economical but with low processing power. The problem becomes harder with increasing number of options and when the individual lacks experience. Understanding how humans make decisions in such scenarios is an important line of study in several fields – cognitive science, psychology, economics, etc. Research shows that humans exhibit systematic shifts in their relative preference between two multi-attribute options, when a third option, called a 'decoy' is introduced into the choice set (Kahneman and Tversky, 1979; Kahneman et al., 1982; Tversky and Simonson, 1993).

Three prominent contextual effects have been observed when starting from two equi-preferable options (say X and Y), each with two attributes: (1) attraction effect (Fig. 4.1A) (Huber et al., 1982; Heath and Chatterjee, 1995), where introduction of an option Z, that is close to Y and is dominated by it on both attributes, leads to an increased preference of Y over X. (2) compromise effect (Fig. 4.1B) (Simonson, 1989), where introduction of an extreme option Z, that is better than Y on one attribute and inferior on the other, as well as is farther from X, again leads to an increased preference of Y over X. (3) similarity effect (Fig. 4.1C) (Tversky, 1972), where the introduction of an option Z, that is almost comparable to Y on both attributes, leads to an increased preference of X over Y. Furthermore, a related effect, called phantom decoy effect (Fig. 4.1D) (Pratkanis and Farquhar, 1992), has also been observed, where introduction of an unavailable option Z, that dominates Y, leads to an increased preference of Y over X. Note that in this case,

even the set of available options is unchanged, and the only difference is the knowledge about the existence of another option that has since become unavailable or sold-out.

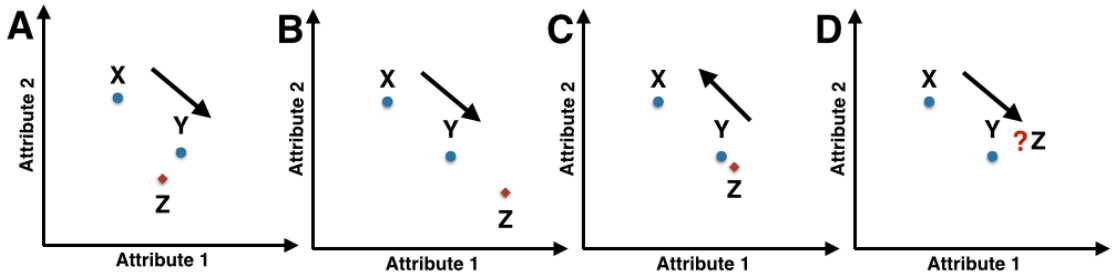


Figure 4.1. Contextual effects. Options X and Y are equi-preferable in the absence of a third option. Arrow denotes the direction of the preference shift when a decoy, Z, is introduced. (A) Attraction effect. (B) Compromise effect. (C) Similarity effect. (D) Phantom decoy effect – the option introduced is not available to be chosen, i.e. it's sold out.

Traditional models of human preference choice assumed that each option has a pre-assigned utility (or value) for the decision-maker, and as such, the presence/absence of a third option should not change the relative desirability of the two original options (Luce, 1959; Thurstone, 1954; Tversky, 1972). They therefore concluded that the existence of contextual effects implied irrationalities or sub-optimalities in human decision-making. A separate approach eschewed normative explanations, instead attributing contextual effects to idiosyncrasies in neural architecture or dynamics in brain areas that support preference choice (Busemeyer and Townsend, 1993; Roe et al., 2001; Usher and McClelland, 2004; Trueblood et al., 2013a). However, with either accounts, it seems superfluous to assume that irrationality or special neural architecture is needed to explain such behavior. Furthermore, neither of these assumptions have been verified empirically.

In this chapter, we propose a rational framework for multi-attribute decision-making, that can explain all four contextual effects, while making three simple and empirically justifiable assumptions: (1) options present in a choice set provide information about the relative importance of the attributes (Sher and McKenzie, 2014); (2)

different attributes are imperfect substitutes for one another (Hicks, 1932), in particular one unit of a scarce attribute is usually more valuable than an abundant one; (3) uncertainty in posterior belief about attribute trade-off contributes to stochasticity in preference on repeated presentations of the same options (see e.g. (Debreu, 1958)). We formalize these assumptions using a Bayesian generative model, and demonstrate that all four contextual effects are consequences of rational (Bayesian) inference of "value", conditioned on the available options. Unlike most models, our model is generalizable to non-economic domains and the "value" can refer to its usual economic sense or a more general quantity, like the "chance" of a suspect being a true culprit based on different strengths of eye-witness accounts (Trueblood, 2012), or the "area" of a rectangle given a visual presentation of different sized ones (Trueblood et al., 2013b). This ability to generalize becomes more important in light of the recent findings, like the preceding examples, that contextual effects are observed in several non-economic settings.

Besides the major contribution of developing a unified framework for explaining all four contextual effects, we make two other contributions: (1) we explain an expanded range of individual differences in human preference choice, by allowing for simultaneous but distinct representations of market-based and personal utility landscapes; (2) we show how the process of consuming choices and observing experienced satisfaction levels can be used by an individual to make inferences about personal utility landscape. We also discuss planned work that considers how the choices might be made when the overarching objective is not just maximization of immediate gain but also to gather information for potentially more advantageous future decisions. In other words, how an individual may choose which options to consume and gain information from, in order to potentially learn about the underlying value structure and learn about the market faster.

The rest of the chapter is organized as follows. First we describe the foundational option model as well as the augmented consumption model that can explain basic

contextual effects and individual differences respectively. Next we present simulation results that show how our model can replicate contextual effects as well as individual differences therein. We conclude with a discussion of future work in the direction of active preference discovery that allows an individual to collect information from an option when it is chosen. This ties in with the overarching goal of the thesis to model information control problems.

4.2 Methods

4.2.1 Option Model

There are two important assumptions that our generative model should incorporate – the options in the choice set provide information about attribute trade-off (e.g. if all cars are high on economy, then economy contributes less to the over-all decision); and the different attributes are not perfectly substitutable for each other, so that decreasing one and increasing the other by the same amount would not lead to the same utility.

We use a Cobb-Douglas utility function that combines the two attributes of an option (Douglas, 1976) using the preference parameter γ. Thus, the utility or value of an option (x, y) is $v = x^{\gamma} y^{1-\gamma}$. For this utility function, if x is decreased and y increased, each by ε, then the change in utilities is: $v' - v \sim v\varepsilon[-\gamma/x + (1-\gamma)/y]$, where v' is the new utility. Note that this captures the idea that the different attributes are not perfect substitutes for each other (if they were, there will be no net change in utility) and also that the differential change in an attribute affects the overall utility less if it was in abundance than if it was scarce (the change in utility has an inverse dependence on the attributes).

The generative model (Fig. 4.2) describes how different options are produced, so that the value of each option (v_i) comes from one set of underlying *market parameters* (k, μ) and the preference (γ) from another set (a, b). The two attributes of an option,

which constitute the observation (o_i), are derived by first parametrically inverting the utility function using θ_i, and then adding the observation noise (σ_0^2). The plate notation uses unshaded circular nodes to denote latent random variables, and the shaded ones to denote observables. Note that in this graphical model, the options do provide information about attribute trade-off via γ, and as more and more options are observed the belief about the hidden γ becomes increasingly peaked.

$$
\begin{aligned}
\gamma &\sim Beta(a, b) \\
v_i &\sim \Gamma(k, \mu/k) \\
\theta_i &\sim N(0, \sigma_\theta^2) \\
o_i &\sim N(f(v_i, \theta_i, \gamma), \sigma_0^2 I) \\
f(v_i, \theta_i, \gamma) &= \left(v_i^{\frac{1+\theta_i}{2\gamma}}, v_i^{\frac{1-\theta_i}{2(1-\gamma)}} \right)
\end{aligned}
$$

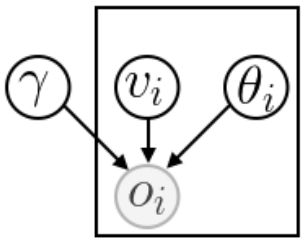


Figure 4.2. Graphical model showing how options are generated in the market. The square box (plate) represents repetitions, so that each *i* corresponds to a different option. γ is the market preference, v_i is the value of the i^{th} option, and θ_i is a parameter required to invert the Cobb-Douglas utility function to get the two attributes of the option. The unshaded nodes represent hidden variables and the shaded one represents observations.

4.2.2 Inference and Decision Making

In order to make meaningful judgements, the observers needs to make inference about the hidden value/utility of different options. This can be done in the recognition model, since having encountered all the options ($\mathbf{o} = \{o\}_i$), the individual can compute

the joint posterior belief on the hidden values ($\mathbf{v} = \{v\}_i$) as:

$$\begin{aligned} P(\mathbf{v}|\mathbf{o}) \propto P(\mathbf{o}|\mathbf{v})P(\mathbf{v}) &= \left\{ \int_{\gamma,\theta} P(\mathbf{o},\gamma,\theta|\mathbf{v}) d\gamma d\theta \right\} P(\mathbf{v}) \\ &= \left\{ \int_{\gamma,\theta} P(\gamma)P(\theta)P(\mathbf{o}|\mathbf{v},\gamma,\theta) d\gamma d\theta \right\} P(\mathbf{v}) \\ &= \left\{ \int_{\gamma} P(\gamma) \prod_i \left[\int_{\theta_i} P(\theta_i)P(o_i|v_i,\gamma,\theta_i) d\theta_i \right] d\gamma \right\} \left\{ \prod_i P(v_i) \right\} \end{aligned} \tag{4.1}$$

To investigate how the presentation of options affects the belief about the market preference parameter, γ, we can evaluate its posterior distribution in a similar fashion:

$$P(\gamma|\mathbf{o}) \propto \left\{ \prod_i \left[\int_{\theta_i,v_i} P(\theta_i)P(v_i)P(o_i|v_i,\gamma,\theta_i) d\theta_i dv_i \right] \right\} P(\gamma) \tag{4.2}$$

The calculations involved in evaluating the posterior are computationally expensive since there is no closed form solution. One way to reduce the computational cost is by iteratively updating the posterior as more and more options are encountered, so that if we have the posterior belief about the values of $n-1$ observed options, we can calculate the joint belief over the value of n options without having to do the entire calculation again:

$$\begin{aligned} \mathbf{p}_n \triangleq P(\mathbf{v}_n|\mathbf{o}_n) &\propto P(o_n, \mathbf{v}_{n-1}|\mathbf{o}_{n-1}, v_n)P(v_n) \\ &= \left\{ P(v_n) \int_{\gamma} P(\gamma|\mathbf{v}_{n-1},\mathbf{o}_{n-1}) \left[\int_{\theta_n} P(\theta_n)P(o_n|v_n,\gamma,\theta_n) d\theta_n \right] d\gamma \right\} \mathbf{p}_{n-1} \end{aligned} \tag{4.3}$$

where

$$P(\gamma|\mathbf{v}_n, \mathbf{o}_n) \propto \left\{ \int_{\theta_n} P(\theta_n) P(o_n|v_n, \gamma, \theta_n) d\theta_n \right\} P(\gamma|\mathbf{v}_{n-1}, \mathbf{o}_{n-1})$$

Similarly, for γ:

$$P(\gamma|\mathbf{o}_n) \propto \left\{ \int_{\theta_n, v_n} P(\theta_n) P(v_n) P(o_n|v_n, \gamma, \theta_n) d\theta_n dv_n \right\} P(\gamma|\mathbf{o}_{n-1}) \tag{4.4}$$

Another way, and the one we use for our simulations, is to use Gibbs sampling (Geman and Geman, 1984) to directly generate samples from the required posterior distribution. As will become evident shortly, this technique goes well with our decision policy too since the policy uses samples from the posterior belief about the values to make choices.

We now look into how the posterior belief about the values could be used to make a choice decision between the available options. Two empirical observations motivate the decision policy we propose – human choice is stochastic upon repeated presentation of same set of options (Camerer, 1989), and humans are known to sample from the posterior on some decision-making tasks. Therefore, we propose that the decision-making phase involves generating a sample, $\hat{\mathbf{v}}$, from the posterior $P(\mathbf{v}|\mathbf{o})$ and choosing an option using the policy:

$$\pi(\hat{\mathbf{v}}) = \underset{j}{\operatorname{argmax}}\, \hat{\mathbf{v}}_j \tag{4.5}$$

The model as presented, along with the inference and decision strategy, incorporates all the information available from just the presentation of the options, and can explain attraction, compromise and similarity effects. However, with the phantom decoy effect, there is also information available about which option was sold out and hence

chosen by others. In terms of the model, this signifies that the value of the sample from the sold out option's posterior was higher than the competing option's. We can incorporate the extra order relationships in our model without any addendum, and hence the phantom decoy effect arises naturally. Formally, if we are given a set of relationships $S = \bigcap_{(i,k)\in I}\{v_i < v_k\}$ where $I \subseteq \{1,\ldots,n\}\text{x}\{1,\ldots,n\}$, s.t. I is irreflexive and anti-symmetric, we can evaluate the posterior belief that incorporates this information as:

$$
\begin{aligned}
P(\mathbf{v}|\mathbf{o},S) &\propto P(\mathbf{o},S|\mathbf{v})P(\mathbf{v}) \\
&= P(\mathbf{o}|\mathbf{v})P(S|\mathbf{v})P(\mathbf{v}) \\
&= \left\{\prod_{(i,k)\in I} \mathbf{1}_{v_i<v_k}(\mathbf{v})\right\} P(\mathbf{o}|\mathbf{v})P(\mathbf{v}) \\
&= \left\{\prod_{(i,k)\in I} \mathbf{1}_{v_i<v_k}(\mathbf{v})\right\} \left\{\int_{\gamma} P(\gamma)\prod_i \left[\int_{\theta_i} P(\theta_i)P(o_i|v_i,\gamma,\theta_i)d\theta_i\right] d\gamma\right\} \left\{\prod_i P(v_i)\right\}
\end{aligned}
\tag{4.6}
$$

Once we have this posterior distribution, we can use the usual decision policy (eq. 4.5). Unfortunately, the integral that appears in the posterior is hard to evaluate but Gibbs sampling can be used again to generate samples from this posterior. In order to do that we need to introduce an extra node that compares the two values. If v_1 denotes the value of the sold out option, and v_2 that of the competing option, then Fig. 4.3 shows how addition of one more node c can encode that extra information.

To reiterate, the explanation of phantom decoy does not require a new model, and the extra node serves as a computational trick to evaluate a complex integral and generate samples from the required posterior.

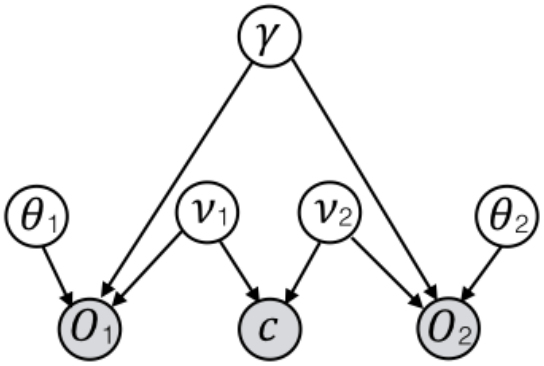


Figure 4.3. Augmented model to incorporate the additional information provided by phantom decoys. $c = \mathbf{1}(v_1 > v_2)$ is the variable encoding which option was consumed, where $\mathbf{1}(.)$ is the indicator function.

4.2.3 Consumption Model

In this model, there are two parameters that contribute to the attribute trade-off, in turn affecting the multi-attribute utility function. The market utility, $v_m = x^{\gamma}y^{1-\gamma}$, is parametrized by market tradeoff parameter γ, and the personal utility, $v_p = x^{\lambda}y^{1-\lambda}$, is parametrized by the personal tradeoff parameter λ. We propose the following way to combine these towards a net utility: $v = v_m^w v_p^{1-w}$, since it leads to a simplified form: $v = x^{\zeta}y^{1-\zeta}$ ($\zeta \triangleq w\gamma + (1-w)\lambda$), and intuitively provides an "average" of the two utilities (in this case, a generalized geometric mean). The parameter w can be interpreted as a personality trait dictating how much an individual values uniqueness as opposed to conforming to the market. For example, consider a consumer buying a car; even though she may be more price-conscious herself, she might still want to buy a trendy, more expensive car because of external considerations like status symbol, peer pressure or resale value ($w > 0.5$). A more rebellious consumer may value uniqueness more and make a decision primarily based on her own preference, giving little consideration to market preferences ($w < 0.5$).

The generative model (Fig. 4.4A) also assumes that depending on whether the

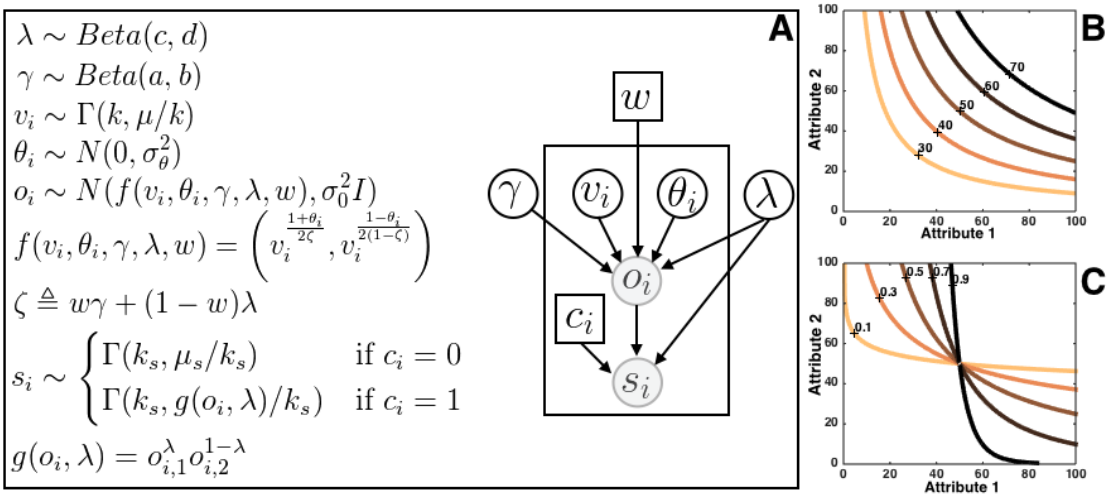


Figure 4.4. (A) Graphical model showing how options (o_i) are generated in the market, based on the hidden value (v_i), market preference (γ), individual preference (λ), weight (w) and the parameter θ_i. Also describes the *satisfaction* as a noisy representation of the personal value, depending on the individual preference (λ) and whether the option is consumed (c_i). (B) For a particular setting of the attribute trade-off parameter (ζ), all options lying on the same curve have the same value/utility as denoted on the curve; different curves signify different values. (C) For a fixed value (v_i), all options lying on the same curve have the same attribute trade-off (ζ), as denoted on the curve; different curves signify different trade-offs.

option is consumed or not ($c_i \in \{0,1\}$), the individual gets different levels of satisfaction. If the option is not consumed ($c_i = 0$), a small level of stochastic satisfaction is received, which can be interpreted to be resulting from mental simulation (thinking about the consumption of the option) but which yields no information about the hidden preference λ. On the other hand, if the option is consumed, the satisfaction derived is a noisy version of the personal utility, $v_p = x^\lambda y^{1-\lambda}$ for an option (x,y), which in turn is informative about λ. For example, after accepting a few of the PhD applicants, the professor may decide, after all, that a student's grades are a better predictive of success for conducting research in her lab than recommendation letters; while another professor may decide just the opposite.

For inference, when the options are not consumed ($c_i = 0 \; \forall \; i$), the node s_i does

not depend on o_i or λ, thus the directed edges from these to s_i are effectively removed and the model reduces to our previous model (for $w = 1$). More generally, the posterior belief about λ can be updated based on the observed options ($\mathbf{o} = \{o\}_i$) and satisfaction ($\mathbf{o} = \{o\}_i$) as:

$$
\begin{aligned}
&P(\lambda|\mathbf{o},\mathbf{s}) \propto P(\mathbf{s}|\mathbf{o},\lambda)P(\mathbf{o}|\lambda)P(\lambda) \\
&= \left\{\prod_i P(s_i|o_i,\lambda;c_i)\right\}\left\{\int_{\gamma,\theta,\mathbf{v}} P(\gamma)P(\theta)P(\mathbf{v})P(\mathbf{o}|\mathbf{v},\gamma,\theta,\lambda;w)d\gamma d\theta d\mathbf{v}\right\}P(\lambda) \\
&= \left\{\prod_i P(s_i|o_i,\lambda;c_i)\right\} \\
&\quad \left\{\int_\gamma P(\gamma)\prod_i\left[\int_{\theta_i,v_i} P(\theta_i)P(v_i)P(o_i|v_i,\gamma,\theta_i,\lambda;w)d\theta_i dv_i\right]d\gamma\right\}P(\lambda)
\end{aligned} \tag{4.7}
$$

Similarly, the joint posterior on the hidden values can be updated as:

$$
\begin{aligned}
&P(\mathbf{v}|\mathbf{o},\mathbf{s}) \\
&\propto \left\{\int_{\gamma,\lambda} P(\gamma)P(\lambda)\prod_i\left[\int_{\theta_i} P(\theta_i)P(o_i|v_i,\gamma,\theta_i,\lambda;w)d\theta_i\right][P(s_i|o_i,\lambda;c_i)]\,d\gamma\,d\lambda\right\} \\
&\quad \left\{\prod_i P(v_i)\right\}
\end{aligned} \tag{4.8}
$$

As in our previous model, we assume that the decision policy involves sampling from the posterior $P(\mathbf{v}|\mathbf{o})$; the sample $\hat{\mathbf{v}}$ is then used to choose an option: $\pi(\hat{\mathbf{v}}) = \text{argmax}_j\,\hat{\mathbf{v}}_j$. Thus, stochasticity in choice upon presentation of the same set of options is expected because of the residual uncertainty in the posterior distributions of the option values (Debreu, 1958; Blavatskyy, 2008).

4.3 Results

We first briefly sketch our model, before moving on to show how its predictions compare against the observed contextual effects. We propose that the net value of an observed option is a generalized geometric mean of its attributes, parametrized by a hidden weight. In the generative model, it is assumed that the net values are random variables and the observed options are produced using the net value and the hidden weight parameter. Thus, as options are observed, the individual updates his belief about the weight parameter as well as the net value of each option using Bayes rule. The details of the model can be found in the Model section.

We now apply the proposed model to different multi-attribute choice tasks, in which sometimes subjects have to choose among just two options, and sometimes among three options whereby a "decoy" is added to the two original options (see Fig. 4.1) [1]. We use options with two attributes (consistent with empirical studies), both of which have been scaled to the same units. The two starting options are X = $(40,60)$, Y = $(60,40)$, and the different decoys are Z = $(50,30)$ for the attraction effect, Z = $(80,20)$ for the compromise effect, Z = $(65,35)$ for the similarity effect, and Z = $(65,45)$ for the phantom decoy effect. For the readers interested in the implementation details, the model parameters are: $a = 2$, $b = 2$, $k = 20$, $\mu = 50$, $\sigma_\theta = 20$ and $\sigma_0 = 2$. We find that no simple closed form expressions exist for the different posteriors and since approximations based on discretization of continuous variables are inexact and inefficient, we use a program called JAGS (Plummer, 2003) that uses Gibbs sampling (Geman and Geman, 1984) to generate samples from the posterior distribution of the desired model parameters.

We note that our model reproduces all three contextual effects, so that even though the two options are equi-preferable by themselves (Fig. 4.5A), the preference

[1] For the phantom decoy effect, the decoy is unavailable so the choice is still between the original two options

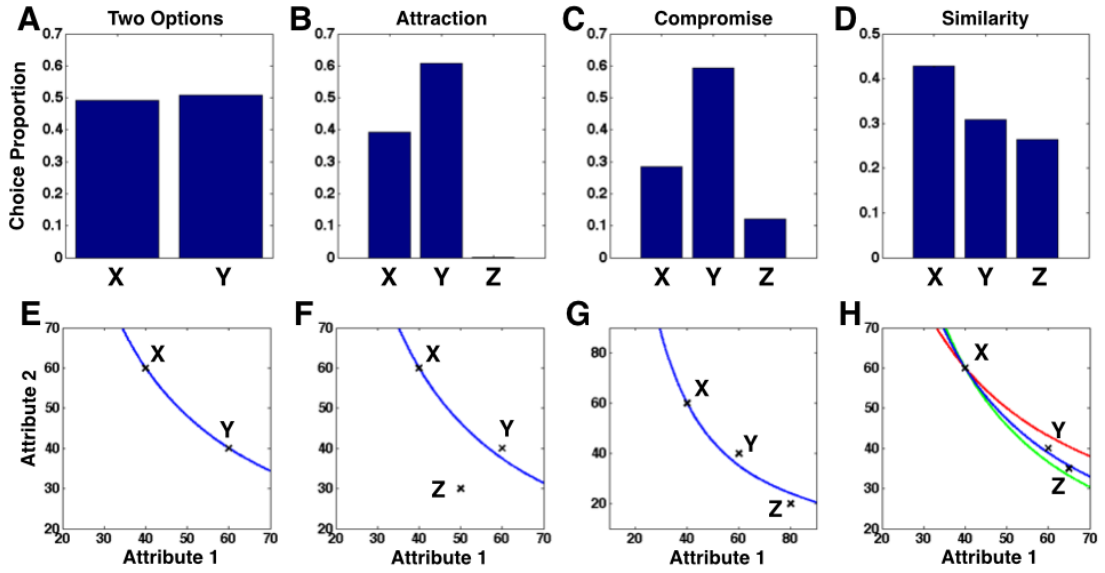


Figure 4.5. Proportion of choices for different effects and their explanation. (A) X and Y are equi-preferable. (B) Adding an option Z inferior to Y increases the preference for Y relative to X. (C) Adding an option Z even more extreme than Y, in relation to X, increases the preference for Y. (D) Adding an option Z very similar to Y, but not clearly more or less preferable to Y, decreases the preference for Y relative to X. (E) X and Y lie on the equi-preference curve which represents a fixed value; any option above this curve would be considered more valuable, and below would be considered less valuable. (F) The attraction decoy makes Y appear better on average. (G) The compromise decoy make Y appear better on average. (H) The valuation of Y and Z are highly correlated, such that they tend to be considered both better than X or both worse than X, for different settings of the hidden variable γ. When they are both worse than X, X is chosen; when they are both worse than X, Y is chosen half the time (Z the other half). Thus, on average, X is chosen more often than Y, even though they are on average considered about equally valuable.

shifts towards Y when an attraction or compromise decoy is added to the choice set (Fig. 4.5B and C), and towards X when a similarity decoy is added (Fig. 4.5D). Next, in order to understand how the model works for all the contextual effects, we show the resulting equi-preference curves, the locus of options that all have the same value for a given preference γ (in other words, all the points on the curve $x^{\gamma}y^{1-\gamma} = v$). Clearly, there are infinitely many equi-preference curves, one corresponding to each value of v, but we only plot the ones, for visual simplicity, that always pass through the option X.

When there are only two options, such an equi-preference curve with γ set to its posterior mean also passes through Y, making the two options equally attractive (Fig. 4.5E). For the attraction effect, option Y lies above such an equi-preference curve , making it appear relatively more lucrative (Fig. 4.5F); however option X is still selected owing to the stochasticity in the option values. Similar explanation holds for the compromise effect (Fig. 4.5G). For the similarity effect, the equi-preference curve with γ set to its posterior mean is not particularly informative (Fig. 4.5H, blue curve), since this effect arises due to the close correlation between the valuation of the option Y and Z in the model. As can be seen, the two options are likely to appear better or worse than X together (green and red curves respectively). Therefore, when they appear better, the choice gets split between them; when they appear worse, X gets chosen, leading to overall higher frequency of choosing X than Y.

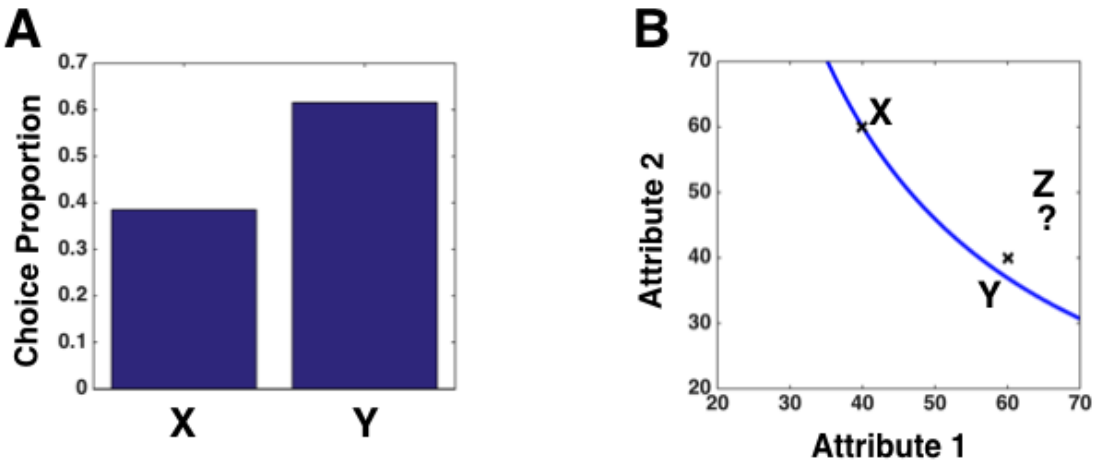


Figure 4.6. (A) Information about an unavailable, or sold out, option increases the preference for Y relative to X. (B) The phantom decoy makes Y appear better on average.

Similarly, our model can also reproduce the phantom decoy effect, so that information about a superior, unavailable option makes the preference shift toward the dominated option (Fig. 4.6A). The model now not only has the information about the options in the market but also additional information about the phantom decoy, which

must have been previously consumed by others for it to become unavailable. Thus, there is additional signal affecting the posterior of γ, in such a way that option Y lies above the equi-preference curve, making it appear more attractive on average (Fig. 4.6B).

We also investigate scenarios where individual exhibit different behavior based on their previous experience. Experiments have show that contextual effects are not always robust, with individual differences emerging when subjects value attributes differently (Malaviya and Sivakumar, 1998; Müller et al., 2012). To show that our model can capture such deviations, we simulate the scenario where an individual prefers attribute y more over attribute x ($\lambda = 0.35$). Furthermore, we assume that the individual relies equally on self and market preference ($w = 0.5$). With these settings, we observe that compromise effect becomes insignificant (Fig. 4.7A), which is what has been observed for consumers who are less quality conscious (attribute x) and more price conscious (attribute y) when making brand choices (Müller et al., 2012).

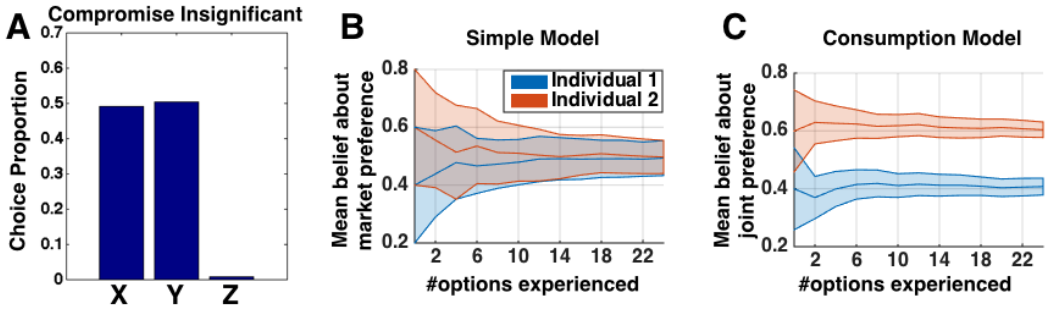


Figure 4.7. (A) Individual difference in contextual effects. Compromise effect becomes insignificant in the consumption model for individuals that value attribute 1 less ($\lambda = 0.35$). (B) & (C) Evolution of belief about market preference with experience for two individuals starting with different priors, using the previous simple model and our consumption model respectively. The translucent bounded region shows the standard deviation around the mean belief.

Lastly, we show how consuming the options can help an individual discover self-preference, and how the the process can lead to divergence in the choice behavior of two individual who learn their preference (from consumption) along with the market

preference (from options). In our original model, the only way the individuals could differ is if they have different priors over the market preference parameter (γ). However, with increase experience, the individual beliefs would converge, consequently leading to the same choice for all individuals who are experienced. In the simulations, we consider two individuals starting with different priors on γ, $Beta(2,3)$ and $Beta(3,2)$, respectively. When both of these individuals experience the same set of options, we note that their beliefs eventually converge (Fig. 4.7B). This contradicts the every day observation, where individuals with very similar experience with a product category, can still converge on buying different brands. Next, we consider the same two individuals and simulate their beliefs under our consumption model. We additionally note that individual 1 prefers attribute 2 more ($\lambda = 0.3$), whereas individual 2 prefers attribute 1 more ($\lambda = 0.7$), both individuals giving equal weight to market-preference and self-preference ($w = 0.5$). The extra parameters are $k_s = 100$ and $\mu_s = 5$. Now the individual beliefs about joint preference diverge, in accordance with their hidden preference (Fig. 4.7C). Thus, the model provides an explanation for why two individuals with the same experience may end up preferring different options, and more generally a framework for explaining different framing effects, as well as variations thereof, based on the individual's experience, personality, and other causes of individualized preferences.

4.4 Discussion

In this chapter, we proposed a Bayesian model of human choice preference that can account for four of the most studied contextual effects – attraction, compromise, similarity and phantom decoy effect. This is the first unified computational account of all the effects to the best of our knowledge and is grounded on empirically motivated assumptions about subjective utility, use of options as information and stochasticity leading to choice differences across trials. The model not only reproduces all the

contextual effects, it also provides insights as to how they can emerge in a rational framework, without having to appeal to sub-optimality of human learning and decision making. The model is also generalizable beyond the obvious consumer domain, since the value in our model can correspond to different metrics, and thus represents a general way to combine attributes of an option. For example, if the option is a potential suspect and the defining attributes the strength of witness accounts, the value metric can be the net confidence that a suspect is guilty (Trueblood, 2012). In this sense, the model can be easily extended to novel scenarios where decisions are made in a multi-attribute setting.

However, there are some empirical findings and common world observations, that pose challenges to our model. The biggest such challenge is capturing individual differences, wherein a set of consumers exposed to same set of options can converge on different choices in the long run. For example, two car salesman who have seen an identical array of cars, can still diverge and choose to drive different cars. This is difficult to capture by our current model since each option provides some information about the hidden parameters, and once a large number of observations have been made, the belief about the value of the parameters would become very strong (the posterior distribution will approach a delta function). Consequently, we would expect the samples drawn by two individuals from their posterior to be very close, converging to the same values as more and more options are observed, and hence leading to same choices. One way to accommodate individual differences in our model is by letting different individuals start with different priors but even this approach falls short in the limit of large number of observation as the effect of priors wear off.

To that end, we proposed an extension that takes both individual preference as well as effects of consuming options into account. This model reduces to our original model in the absence of consuming chosen options (and experiencing corresponding satisfaction levels) along with appropriate parameter settings, and can explain not only all

three contextual effects but also observed individual differences arising from experience and personal preference. This framework provides a general way for explaining how consumption combines with "window-shopping" to inform one's beliefs and preferences, thus leading to diversification of individual preferences.

In the future, we also wish to investigate whether an individual has a fixed relative weight for personal and market preferences, or whether this trade-off can change either with experience or context. One possibility is, as the individual gains more experience, the trade-off starts to favor self-preference, thus requiring a more sophisticated model where the trade-off parameter (w) is dependent on experience (perhaps tied to internal uncertainty/confidence). Another important direction of enquiry is the field of preference discovery, wherein instead of passively or non-strategically consuming options, an individual can adopt a policy of consuming options that maximize a joint objective of highest value option and options that provide the most information about the market or individual landscape. One way to go about this is by using a Knowledge Gradient policy (Frazier et al., 2008), which is a heuristic policy shown to perform close to optimal for Multi-armed Bandits and in fact providing a better explanation to human behavior than even the optimal policy (Zhang and Angela, 2013). In conclusion, our work provides foundations for understanding human decision making in multi-attribute choice scenarios, as well as leads the way for building more complex models of decision making that can capture the exploration vs. exploitation trade-off inherent in such choices.

Acknowledgements

This chapter contains material from S. Ahmad and A.J. Yu, A rational model for individual differences in preference choice, Proceedings of the 37th Annual Conference of the Cognitive Science Society, 2015.The dissertation author was the primary investigator and author of this paper.

This chapter also contains material being prepared for submission. The dissertation author is the primary investigator and author of this material.

Chapter 5

Conclusion

This dissertation focuses on developing principled models of learning and decision-making in different scenarios, with a focus on the problems of active cognition where the agent can exert control over the source of information. The major contribution of this work is the use of a well-substantiated Bayesian framework to model different behavior that arises when humans or animals are faced with certain problems, like target search, foraging and consumer choices. Often models assume irrationality of the learner or optimization of proxy goals as a first step towards understanding complex behavior. However, we argue that the seemingly irrational behavior is often an outcome of the imperfect state of knowledge of the individual. Therefore, we eschew the approach of using irrationality as an explanation and build our models so that they can keep track of the knowledge state of the individual, who then acts on her beliefs based on her current experience which may or may not correspond to the ground truth. This leads to principled, normative models of behavior based on justifiable assumptions. More concretely, we looked at three different problems and made novel contributions towards each.

The problem of active sensing deals with how to allocate sensory resources like saccades to informative aspects of the environment. We proposed a Bayesian framework that incorporates the noisy nature of observations and bases sampling decisions to optimize a joint behavioral metric consisting of speed, effort and accuracy. We also

presented a visual search experiment that produces behavioral data which can be used to validate different models. We compared our model with another recent model that optimizes for entropy and noticed that not only can our model trivially outperform it on engineering tasks where speed and accuracy are indeed the concerned metrics, it can also explain human behavior much better. Specifically, we observe that it leads to the emergence of confirmation bias which is consistent with human behavior where there is a bias towards the most likely target location both in terms of choice proportion and response time. We also presented several approximations to enable computational savings and potentially generalize our model to more complex scenario. Lastly, we introduced an extension to incorporate peripheral vision which is known to be used by humans during visual search, making our model more amenable for understanding real world scenario where there are no constraints on fixations.

The problem of competitive foraging deals with how an agent, say an animal looking for food in a patchy environment, must trade-off between the information it gets from personal experience with the information from social input, in order to maximize its reward gain. The problem is motivated by animal behavior literature where numerous empirical studies observe how animals distribute themselves between locations in the proportion of food availability, thus reaching the optimal solution. However, when it comes to modeling such behavior, stringent constraints on state of information or cooperation have been traditionally made. In our work, we relax these assumptions and show that a model that incorporates imperfect state of information as well as competition can still lead to the optimal solution. We explore different scenarios to show how this model can even adapt to non-stationary scenarios where a naive model fails to. Therefore, this works provides insights into how optimal crowd behavior can emerge in competitive scenarios.

Finally, the multi-attribute decision making problem is a computational study of

how the mere addition of an option can affect the relative preferences between pre-existing multi-attribute options. This phenomenon has been well studied empirically but poorly understood theoretically since it violates certain regularity principles that were deemed foundational in economics. We present a model that makes justifiable assumptions about the world and can reproduce the most prominent contextual effects. We also present an extension that can reproduce observed individual differences in humans by allowing an individual to not just observe but also consume options. Lastly, we ruminate on how our work can be possibly augmented to understand the scenarios where an individual has control over which options to consume and the objective is to not just consume the best option greedily but potentially consume options that can tell us more about the utility landscape. Our model is one of the first computational models, and to the best of our knowledge the first Bayesian model, that can explain different contextual effects in a single framework, and has potential to be extended to another complex problem of preference discovery.

To conclude, this dissertation makes contribution towards enriching our understanding of how humans and animals learn and make decisions in several scantily explored problems. It opens up venues to potentially gather a deeper understanding of brain functioning as well as provides ways to approach engineering applications motivated by how the brain solves these problems. We believe that an interdisciplinary approach such as this is a promising way forward to tackle some of the most complex problems encountered both in cognitive science and in computer science.

Appendix A

A.1 Greedy MAP Policy: Analytical Solution

The Greedy MAP policy suggests that agents should try to maximize the expected one-step look-ahead probability of finding the target. So, the reward associated with fixating location j at time $t+1$ is:

$$
\begin{aligned}
R^g(\mathbf{p}_t, j) &= \mathbb{E}_{x_{t+1}}[\max_i P(s=i|\mathbf{x}_t, x_{t+1}, \lambda_t, \lambda_{t+1}=j)] \\
&= \sum_{x_{t+1}\in\{0,1\}} P(x_{t+1}|\lambda_{t+1}=j, \mathbf{x}_t, \lambda_t,)[\max_i P(s=i|\mathbf{x}_t, x_{t+1}, \lambda_t, \lambda_{t+1}=j)] \\
&= \sum_{x_{t+1}\in\{0,1\}} \left[\sum_k P(x_{t+1}|\lambda_{t+1}=j, \mathbf{x}_t, \lambda_t, s=k) P(s=k|\lambda_{t+1}=j, \mathbf{x}_t, \lambda)\right] \left[\max_i (p^i_{t+1})\right] \\
&= \sum_{x_{t+1}\in\{0,1\}} \left[\sum_k P(x_{t+1}|\lambda_{t+1}=j, s=k) P(s=k|\mathbf{x}_t, \lambda)\right] \left[\max_i \frac{f_{i,j}(x_{t+1}) p^i_t}{\sum_l f_{s,l}(x_{t+1}) p^l_t}\right] \\
&= \sum_{x_{t+1}\in\{0,1\}} \left[\sum_k f_{k,j}(x_{t+1}) p^k_t\right] \left[\max_i \frac{f_{i,j}(x_{t+1}) p^i_t}{\sum_l f_{l,j}(x_{t+1}) p^l_t}\right] \\
&= \sum_{x_{t+1}\in\{0,1\}} \max_i f_{i,j}(x_{t+1}) p^i_t \\
&= \sum_{x_{t+1}\in\{0,1\}} \max_i \left\{ \left[\mathbf{1}_{\{i=j\}} \beta^{x_{t+1}} (1-\beta)^{1-x_{t+1}} + \mathbf{1}_{\{i\neq j\}} (1-\beta)^{x_{t+1}} \beta^{1-x_{t+1}}\right] p^i_t \right\}
\end{aligned}
$$

where the last step follows if we assume Bernoulli observations. Furthermore, for the belief state, $\mathbf{p}_t$, the Greedy MAP policy is given by: $\lambda^g(\mathbf{p}_t) = \text{argmax}_j R^g(\mathbf{p}_t, j)$

In order to demonstrate the policy plot, we consider the region where $p_t^1 > p_t^2 > p_t^3$, since these observations can then be generalized based on symmetry. Further, consider the three sub-regions:

Sub-region I. $(1-\beta)p_t^1 > \beta p_t^2$

For all $\mathbf{p}_t$ in this region, we get:

$$R^g(\mathbf{p}_t, 1) = R^g(\mathbf{p}_t, 2) = R^g(\mathbf{p}_t, 3) = p_t^1$$

Since all three rewards are same, there is no preference for the next fixation.

Sub-region II. $(1-\beta)p_t^1 < \beta p_t^2$ and $(1-\beta)p_t^1 > \beta p_t^3$

The rewards are:

$$R^g(\mathbf{p}_t, 1) = R^g(\mathbf{p}_t, 2) = \beta(p_t^1 + p_t^2)$$
$$R^g(\mathbf{p}_t, 3) = p_t^1$$

So, $R^g(\mathbf{p}_t, 1) = R^g(\mathbf{p}_t, 2) > R^g(\mathbf{p}_t, 3)$, and Greedy MAP policy is to fixate either location 1 or 2.

Sub-region III. $(1-\beta)p_t^1 < \beta p_t^2$ and $(1-\beta)p_t^1 < \beta p_t^3$

The rewards are:

$$R^g(\mathbf{p}_t, 1) = R^g(\mathbf{p}_t, 2) = \beta(p_t^1 + p_t^2)$$
$$R^g(\mathbf{p}_t, 3) = \beta(p_t^1 + p_t^3)$$

Again, $R^g(\mathbf{p}_t, 1) = R^g(\mathbf{p}_t, 2) > R^g(\mathbf{p}_t, 3)$, and Greedy MAP policy is to fixate either location 1 or 2.

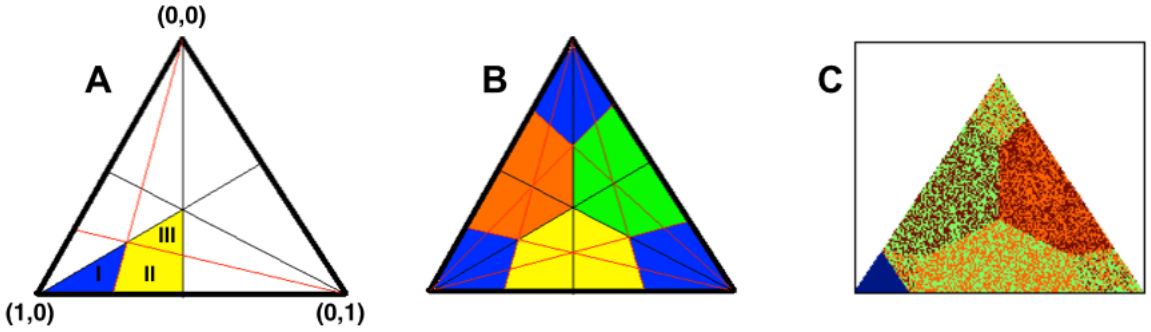


Figure A.1. (A) Analytical Greedy MAP policy plot for the region $\mathbf{p}_t^1 > \mathbf{p}_t^2 > \mathbf{p}_t^3$. (B) Analytical Greedy MAP policy plot for the complete simplex. Blue: fixate location 1, 2 or 3. Yellow: fixate location 1 or 2. Green: fixate location 2 or 3. Orange: fixate location 3 or 1. (C) Simulated Greedy MAP policy. Blue: stop. Green: fixate location 1. Orange: fixate location 2. Brown: fixate location 3. When more that one actions is optimal in a region, the action is chosen randomly leading to the blotchy appearance.

Fig. A.1(A) shows the policy plot for sub-regions I, II and III where blue denotes fixate 1, 2 or 3 and yellow denotes fixate 1 or 2. Also, based on the symmetry, the policy plot on the simplex is shown in Fig. A.1(B) such that green corresponds to fixate 2 or 3 and orange fixate 3 or 1. This corresponds closely with simulations, as can be seen in Fig. A.1(C), which shows one instance of the Greedy MAP policy – there are regions where it recommends that all locations are equally good, and similarly regions where two of the locations are equally good. The simulated policy also shows the stopping region (blue) which is an augmentation so that the algorithm has a confidence threshold about when to stop.

Bibliography

Araya-López, M., Thomas, V., Buffet, O., and Charpillet, F. (2010). A closer look at momdps. In *Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on*, volume 2, pages 197–204. IEEE.

Bellman, R. (1952). On the theory of dynamic programming. *PNAS*, 38(8):716–719.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, pages 992–1026.

Blavatskyy, P. R. (2008). Stochastic utility theorem. *Journal of Mathematical Economics*, 44(11):1049–1056.

Boyd, R. (1988). *Culture and the evolutionary process*. University of Chicago Press.

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.*, 12:4745–65.

Buhmann, M. (2003). *Radial basis functions: theory and implementations*, volume 12. Cambridge university press.

Busemeyer, J. and Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432–459.

Butko, N. J. and Movellan, J. R. (2008). I-POMDP: An infomax model of eye movement. In *Proceedings of the International Conference on Development and Learning (ICDL)*.

Butko, N. J. and Movellan, J. R. (2010). Infomax control of eyemovements. *IEEE Transactions on Autonomous Mental Development*, 2(2):91–107.

Camerer, C. F. (1989). An experimental test of several generalized utility theories.

Journal of Risk and uncertainty, 2(1):61–104.

Chung, T. H. and Burdick, J. W. (April 2007). A decision-making framework for control strategies in probabilistic search. *Intl. Conference on Robotics and Automation. ICRA*.

Darrell, T. and Pentland, A. (1996). Active gesture recognition using partially observable markov decision processes. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 3, pages 984–988. IEEE.

Debreu, G. (1958). Stochastic choice and cardinal utility. *Econometrica*, 26:440–444.

Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in neurosciences*, 21(8):355–361.

Dill, L. M. (1987). Animal decision making and its ecological consequences: the future of aquatic ecology and behaviour. *Canadian Journal of Zoology*, 65(4):803–811.

Dorard, L., Glowacka, D., and Shawe-Taylor, J. (2009). Gaussian process modelling of dependencies in multi-armed bandit problems. In *Int. Symp. Op. Res*.

Douglas, P. (1976). The Cobb-Douglas production function once again: its history, its testing, and some new empirical values. *The Journal of Political Economy*.

Dreisig, H. (1995). Ideal free distributions of nectar foraging bumblebees. *Oikos*, pages 161–172.

Findley, J. M. (1982). Global processing for saccadic eye movements. *Vision Research*.

Frazier, P. I., Powell, W. B., and Dayanik, S. (2008). A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439.

Galef Jr, B. G. and Laland, K. N. (2005). Social learning in animals: empirical studies and theoretical models. *Bioscience*, 55(6):489–499.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 6:721–741.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *J. Royal Stat. Soc.*, 41:148–77.

Goldstone, R. L. and Ashpole, B. C. (2004). Human foraging behavior in a virtual environment. *Psychonomic Bulletin & Review*, 11(3):508–514.

Goldstone, R. L., Ashpole, B. C., and Roberts, M. E. (2005). Knowledge of resources and competitors in human foraging. *Psychonomic Bulletin & Review*, 12(1):81–87.

Golovin, D., Krause, A., and Ray, D. (2010). Near-optimal bayesian active learning with noisy observations. *arXiv preprint arXiv:1010.3091*.

Green, R. F. (1980). Bayesian birds: a simple example of oaten's stochastic model of optimal foraging. *Theoretical Population Biology*, 18(2):244–256.

Griffith, V. and Yaeger, L. S. (2011). Ideal free distribution in agents with evolved neural architectures. *arXiv preprint arXiv:1112.3574*.

Harper, D. (1982). Competitive foraging in mallards:ideal freeducks. *Animal Behaviour*, 30(2):575–584.

He, P. Y. and Kowler, E. (1989). The role of location probability in the programming of saccades: implications for "center-of-gravity" tendencies. *Vision Res*, 29(9):1165–81.

Heath, T. and Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, 22:268–284.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, 16(4):219–22.

Hicks, J. R. (1932). *The theory of wages*.

Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of consumer research*, pages 90–98.

Itti, L. and Baldi, P. (2006). Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems, Vol. 19*, pages 1–8, Cambridge, MA. MIT Press.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10-12):1489–506.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-bayesian

social learning. *Games and Economic Behavior*.

Ji, S., Parr, R., and Carin, L. (2007). Nonmyopic multiaspect sensing with partially observable markov decision processes. *Signal Processing, IEEE Transactions on*, 55(6):2720–2730.

Kahneman, D., Slovic, P., and Tversky, A., editors (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47:313–27.

Kaplow, R. (2010). *Point-based POMDP solvers: Survey and comparative analysis*. PhD thesis, McGill University.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.*

Kraft, J. R. and Baum, W. M. (2001). Group choice: The ideal free distribution of human social behavior. *Journal of the Experimental Analysis of Behavior*, 76(1):21–42.

Krebs, J. R. (1973). Social learning and the significance of mixed-species flocks of chickadees (parus spp.). *Canadian Journal of Zoology*, 51(12):1275–1288.

Krivan, V., Cressman, R., Schneider, C., et al. (2008). The ideal free distribution: a review and synthesis of the game-theoretic perspective. *Theoretical population biology*, 73(3):403–425.

Kwok, C. and Fox, D. (2004). Reinforcement learning for sensing strategies. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 4, pages 3158–3163. IEEE.

Lagoudakis, M. and Parr, R. (2003). Least-squares policy iteration. *The Journal of Machine Learning Research*, 4:1107–1149.

Land, M. F. and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26):3559–65.

Land, M. F. and Lee, D. N. (1994). Where we look when we steer. *Nature*, 369(6483):742–4.

Lee, T. S. and Yu, S. (2000). An information-theoretic framework for understanding

saccadic behaviors. In *Advance in Neural Information Processing Systems*, volume 12, Cambridge, MA. MIT Press.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature neuroscience*, 5(4):356–363.

Lima, S. L. (1985). Sampling behavior of starlings foraging in simple patchy environments. *Behavioral Ecology and Sociobiology*, 16(2):135–142.

Liu, K. and Zhao, Q. (2010). Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access. *Information Theory, IEEE Transactions on*, 56(11):5547–5567.

Livoreil, B. and Giraldeau, L.-A. (1997). Patch departure decisions by spice finches foraging singly or in groups. *Animal behaviour*, 54(4):967–977.

Lovejoy, W. (1991). Computationally feasible bounds for partially observed markov decision processes. *Operations research*, 39(1):162–175.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley, New York.

Madden, G. J., Peden, B. F., and Yamaguchi, T. (2002). Human group choice: Discrete-trial and free-operant tests of the ideal free distribution. *Journal of the Experimental Analysis of Behavior*, 78(1):1–15.

Malaviya, P. and Sivakumar, K. (1998). The moderating effect of product category knowledge and attribute importance on the attraction effect. *Marketing Letters*, 9(1):93–106.

Milinski, M. (1979). An evolutionarily stable feeding strategy in sticklebacks1. *Zeitschrift für Tierpsychologie*, 51(1):36–40.

Minut, S. and Mahadevan, S. (2001). A reinforcement learning model of selective visual attention. In *Proceedings of the Fifth International Conference on Autonomous Agents, Montreal*.

Müller, H., Vogt, B., and Kroll, E. B. (2012). To be or not to be price consciousa segment-based analysis of compromise effects in market-like framings. *Psychology & Marketing*, 29(2):107–116.

Naghshvar, M. and Javidi, T. (2010). Active m-ary sequential hypothesis testing. In

Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on, pages 1623–1627. IEEE.

Naghshvar, M. and Javidi, T. (2012). Active sequential hypothesis testing. *arXiv preprint arXiv:1203.4626*.

Najemnik, J. and Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, 434(7031):387–91.

Nash, J. F. et al. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49.

Olsson, O. and Holmgren, N. M. (1998). The survival-rate-maximizing policy for bayesian foragers: wait for good news. *Behavioral Ecology*, 9(4):345–353.

Ong, S. C., Png, S. W., Hsu, D., and Lee, W. S. (2010). Planning under uncertainty for robotic tasks with mixed observability. *The International Journal of Robotics Research*, 29(8):1053–1068.

Oswal, A., Ogden, M., and Carpenter, R. H. S. (2007). The time course of stimulus expectation in a saccadic decision task. *J Neurophysiol*, 97:2722–30.

Pandey, S., Chakrabarti, D., and Agarwal, D. (2007). Multi-armed bandit problems with dependent arms. In *Proceedings of the 24th international conference on Machine learning*, pages 721–728. ACM.

Parker, G. and Sutherland, W. (1986). Ideal free distributions when individuals differ in competitive ability: phenotype-limited ideal free models. *Animal Behaviour*, 34(4):1222–1242.

Perry, G. and Pianka, E. R. (1997). Animal foraging: past, present and future. *Trends in Ecology & Evolution*, 12(9):360–364.

Pineau, J., Gordon, G., and Thrun, S. (2006). Anytime point-based approximations for large pomdps. *Journal of Artificial Intelligence Research*, 27(1):335–380.

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Sympo Quant Biol*, 55:899–910.

Powell, W. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. Wiley-Interscience.

Pratkanis, A. R. and Farquhar, P. H. (1992). A brief history of research on phantom alternatives: Evidence for seven empirical generalizations about phantoms. *Basic and Applied Social Psychology*, 13(1):103–122.

Pyke, G. H. (1984). Optimal foraging theory: a critical review. *Annual review of ecology and systematics*, 15:523–575.

Pyke, G. H., Pulliam, H. R., and Charnov, E. L. (1977). Optimal foraging: a selective review of theory and tests. *Quarterly Review of Biology*, pages 137–154.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., and Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Res*, 42(11):1447–63.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025.

Roberts, M. E. and Goldstone, R. L. (2006). Epicure: Spatial and knowledge limitations in group foraging. *Adaptive Behavior*, 14(4):291–314.

Roe, R. M., Busemeyer, J. R., and Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review*, 108(2):370–392.

Roesch, M. R. and Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J Neurophysiol*, pages 1766–89.

Sernland, E., Olsson, O., and Holmgren, N. M. (2003). Does information sharing promote group foraging? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1520):1137–1141.

Seth, A. K. (2011). Optimal agent-based models of action selection.

Sher, S. and McKenzie, C. R. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General*, 143(3):1127.

Simoncelli, E. P. and Heeger, D. J. (1998). A model of neuronal responses in visual area mt. *Vision research*, 38(5):743–761.

Simonson, I. (1989). Choice based on reasons: the case of attraction and compromise effects. *Journal of Consumer Research*, 16(158-74).

Sokolowski, M., Tonneau, F., and i Baqué, E. F. (1999). The ideal free distribution in humans: An experimental test. *Psychonomic bulletin & review*, 6(1):157–161.

Sridharan, M., Wyatt, J., and Dearden, R. (2010). Planning to see: A hierarchical approach to planning visual actions on a robot using pomdps. *Artificial Intelligence*, pages 704–725.

Templeton, J. J. and Giraldeau, L.-A. (1995). Patch assessment in foraging flocks of european starlings: evidence for the use of public information. *Behavioral Ecology*, 6(1):65–72.

Thurstone, L. (1954). The measurement of values. *Psychological Review*.

Trueblood, J. S. (2012). Multialternative context effects obtained using an inference task. *Psychonomic bulletin & review*, 19(5):962–8.

Trueblood, J. S., Brown, S. D., and Heathcote, A. (2013a). The multi-attribute linear ballistic accumulator model of decision-making. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

Trueblood, J. S., Brown, S. D., Heathcote, A., and Busemeyer, J. R. (2013b). Not just for consumers: context effects are fundamental to decision making. *Psychological science*, 24(6):901–908.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological review*, 79(4).

Tversky, A. and Simonson, I. (1993). Context dependent preferences. *Management Science*, 39:1179–89.

Usher, M. and McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review*, 111(3):757.

Valone, T. J. (1989). Group foraging, public information, and patch estimation. *Oikos*, pages 357–363.

Valone, T. J. and Templeton, J. J. (2002). Public information for the assessment of quality: a widespread social phenomenon. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1427):1549–1557.

Walker, J. C. (2011). The evolution of optimal foraging strategies in populations of digital organisms. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 203–210. ACM.

Washburn, R. and Schneider, M. (2008). Optimal policies for a class of restless multi-armed bandit scheduling problems with applications to sensor management. *Journal of Advances in Information Fusion. v3*.

Whittle, P. (1988). Restless bandits: activity allocation in a changing world. *J. App. Probability*, 25A:287–98.

Williams, C. K. I. and Rasmussen, C. E. (1996). Gaussian processes for regression. In D. S. Touretzky, M. M. and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems 8*, pages 514–20. MIT Press, Cambridge, MA.

Wilson, R. I. and Mainen, Z. F. (2006). Early events in olfactory processing. *Annu. Rev. Neurosci.*, 29:163–201.

Yarbus, A. F. (1967). *Eye Movements and Vision*. Plenum Press, New York.

Yu, A. J. and Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, 21:1873–80.

Yu, A. J. and Huang, H. (2014). Maximizing masquerading as matching in human visual search choice behavior. *Decision*. To appear.

Zelinsky, G. J., Rao, R. P., Hayhoe, M. M., and Ballard, D. H. (1997). Eye movements reveal the spatio-temporal dynamics of visual search. *Psychol. Sci.*

Zhang, S. and Angela, J. Y. (2013). Forgetful bayes and myopic planning: Human learning and decision-making in a bandit setting. In *Advances in neural information processing systems*, pages 2607–2615.