UNIVERSITY OF CALIFORNIA, SAN DIEGO

Decision-making and Motor control: Computational Models of Human Sensorimotor Processing

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

in

Cognitive Science

by

He Huang

Committee in charge:

Douglas Nitz, Chair
Martin Paulus, Co-Chair
Marian Bartlett
Andrea Chiba
Sarah Creel
Jeffery Krichmar
Javier Movellan

2014

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

The Dissertation of He Huang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2014

TABLE OF COTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my committee chair Douglas Nitz for his encouragements and support that give me the strength to complete this journey.

I would like to thank Javier Movellan for his continuous guidance and timely advice for my thesis project. He kindly took me when I was at a critical point in the PhD program. Since then he has been a great advisor and a mentor, not only for the research methods, but teaching me how to become a passionate, unbiased and thorough researcher.

I would like to thank Martin Paulus in guiding me in the thesis project. The work in this dissertation would have been impossible without his constant support and help. Being a great mentor, his suggestions and advice were very inspirational in designing and interpreting the results from this project.

I would also like to thank Marian Bartlett for guiding me in the facial expression analysis, and introducing me to Martin, and made this thesis project possible.

I would like to thank the rest of my committee: Andrea Chiba, Sarah Creel and Jeffery Krichmar for generously giving their time and sharing their knowledge to guide me through this dissertation.

I would also like to thank Angela Yu, my former advisor, who took me in this program and gave me inspirations from our enlightening discussions.

I would also like to thank my collaborators: Pradeep Shenoy and Shunan Zhang from Yu Lab who have provided their expertise in Bayesian modeling, Katia Harle from Palus Lab who guided me in the depression study, and Walter Talbott from Machine Perception Lab who gave me numerous help in my dissertation project.

Chapter 2, in full, is a reprint of the material as it appears in Decision 2014. Yu, A. J & Huang, H.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the Cognitive science Society Conference 2014. Zhang S, Huang H and Yu A.

Chapter 4, in full, is a reprint of the material as it appears in Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on. Talbot W, Huang H and Movellan J.

Chapter 5, in full, is submitted to the journal of Cognition. Huang H, Movellan J, Paulus M and Harlé K.

Chapter 6, in full, is currently being prepared for submission for publication of the material. Huang H, Harlé K, Movellan J and Paulus M.

VITA

2009 - 2014 Ph.D. in Cognitive Science, UC San Diego, La Jolla, CA

2007 - 2009 M.S. in Statistics, UC San Diego, La Jolla, CA

2003 - 2007 B.S. in Mathematics, Nanjing University of Finance and Economics
Jiang Su, China

Publications

Huang, H., Movellan, J., Paulus, M.P. & Harlé, K.M. The Influence of Depression on Cognitive Control: Disambiguating Approach and Avoidance Tendencies, *submitted*

Huang, H., Harlé, K.M., Paulus, M.P.& Movellan, J. (2014) Inverse Optimal Control Model of Driving Behavior in Depressed Individuals, Proceedings of the 36th Annual Conference of the Cognitive Science Society, pp. 2387-2392. Austin, TX: Cognitive Science Society, 2014

Yu, A.J. & **Huang, H.** Maximizing Masquerading as Matching: Statistical Learning and Decision-making in Choice Behavior, Decision, Vol 1(4), Oct 2014, 275-287

Ahmad, S., **Huang, H.,** & Yu, A.J. (2014). Cost-Sensitive Bayesian Control Policy in Human Active Sensing. Frontiers in Human Neuroscience, 8, 955.

Zhang, S., **Huang, H.** & Yu, A.J. (2014). Sequential Effects: A Bayesian Analysis of Prior Bias on Reaction Time and Behavioral Choice, Proceedings of the 36th Annual Conference of the Cognitive Science Society, pp. 1844-1849. Austin, TX: Cognitive Science Society

Ahamd, S., **Huang, H.** & Yu, A.J. (2013). Context-sensitivity in Human Active Sensing, Advances in Neural Information Processing Systems 26. MIT Press, Cambridge, MA.

Talbot,W., **Huang, H.** & Movellan, J. (2012). Infomax Models of Oculomotor Con- trol, Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on

ABSTRACT OF THE DISSERTATION

Decision-making and Motor control: Computational Models of Human Sensorimotor Processing

by

He Huang

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2014

Professor Douglas Nitz, Chair
Professor Martin Paulus, Co-Chair

To survive and effectively interact with the environment, human sensorimotor control system collects sensory information and acts based on the state of the world. Human behavior can be considered and studied at discrete time or continuous time. For

the former, human makes discrete categorical decisions when presented with different alternative choices (e.g. choose Left or Right at an intersection). For the later, humans plan and execute continuous movements when instructed to perform a motor task (e.g. drive to a destination). In this dissertation we examine human behavior at both levels. Part I focuses on understanding decision-making at discrete time using Bayesian Models. We start by investigating the influence of environmental statistics in a saccadic visual search ask, in which we use a dynamic belief model to describe subjects' learning process of the environment statistics cross-trials. Then we look at a special effect of decision-making, the sequential effect, and apply the dynamic belief model to explain subjects' cross-trial learning and a drift diffusion model to explain their within-trial decision-making process. Part II focuses on examining motor control at continuous time using Optimal Control Theory. We start by investigating the objective functions in oculomotor control (saccadic eye movement, smooth pursuit, and applications in eye-hand coordination) with an infomax model. Then we apply inverse optimal control model to study impaired motor behavior in depressed individuals. In particular, we present a framework based on optimal control theory, which can distinguish the effects of sensorimotor speed, goal setting and motivational factors in goal-directed motor tasks. Finally, we propose to use facial expression as another measure of the emotional state in depressed individuals, which can be used to provide further understanding of the behavior and model parameters estimated from the proposed inverse framework.

Chapter 1

Introduction

To effectively interact with the environment, humans make decisions and plan movements based on the sensory information collected from the world. This dissertation is going to focus on two types of behavior to gain better understanding of human sensorimotor processing: perceptual decision making (or action selection) at discrete time, and goal-directed motor control (or movement planning and execution) at continuous time.

Decision-making and motor control are not two independent behavior systems (Trommershauser, Maloney & Landy 2003; Rigoux & Guigon 2012). They both describe human sensorimotor processes, and can fit in reinforcement learning theory (Sutton & Barto, 1998; Wolpert & Landy 2012). That is, both decision-makers and movement controllers have the goal to achieve an objective function (maximize the reward or minimize the loss) in a given task. They do this by learning the dynamics of how the state updates (i.e. state transitions), predicting the state of the environment and taking actions that are based on sensory observations, and then using observed outcomes to optimize performance (Figure 1.1).

Given a certain decision-making task, for example a 2 Alternative-Forced-Choice (2AFC) task, a decision-maker evaluates benefits and costs associated with the given choices, uses perceived sensory information, and selects the action that will maximize the expected benefit. Then the decision-maker will use the observed outcomes to improve the

decision-making process. Thus it forms a sensorimotor loop between the decision-maker and the environment.

Given a goal-directed motor task, for example driving a car to a stop sign, the movement controller also assesses the reward and cost associated with the task, uses perceived sensory information to evaluate current state of the car (car position, velocity etc.), and based on which sends motor command to execute the action to interact with the environment. Then the movement controller will use the observed sensory information to update the state information at next time step to make continuous motor commands. Thus it also forms a sensorimotor feedback loop between the movement controller and the environment.

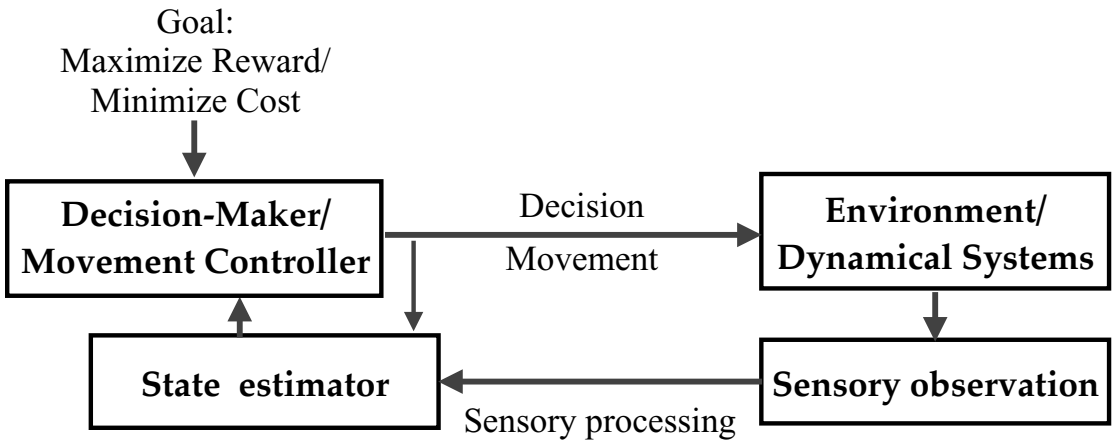

Figure 1.1: Decision-making & Motor control in sensorimotor feedback loop

1.1 Bayesian approach in Decision-making

Bayesian approach is one of the most important computational models in studying decision-making under uncertainty (Kording 2007; Yu 2007). It provides quantitative measures of how one's belief changes based on prior knowledge and new sensory

observation. Bayesian inference has been applied successfully in a variety of tasks (Yuille & Bulthoff 1994; Huys & Dayan 2009; Vilares & Kording 2011) and has been shown to be an effective model to explain human behavior. In particular, from a Bayesian viewpoint, we can now examine different factors that influence how new information is combined with past knowledge. For example, with different expectations of how stable the environment is (i.e. how frequently unexpected changes occurs in the environment), it may not be ideal to keep tracking of outcomes in the far past. That is, in a highly non-stationary environment, the prior knowledge may not contribute as effectively in the decision-making process as it would in a more stationary world, in which past knowledge can provide much insight in predicting the state in the future. Thus, one's internal measure of the environment stability can be one important factor in combining prior and new information. In this dissertation, we are going to investigate this issue by using a dynamic belief model (Yu & Cohen 2009), which takes into account of individual's belief of the stability of the environment in estimating the posterior belief of the underlying state.

It is worth noting that, recursive Bayesian inference can be considered as a problem of Kalman filter (Berniker & Kording 2011), which is a key concept in control theory. Kalman filter is a recursive process for inference about underlying state θ_t using observations $\boldsymbol{x}_t$, based on *an observation equation* that describes how observation $\boldsymbol{x}_t$ is generated from the underlying state θ_t with an observation noise, and *a state equation* that describes how state θ_t transitions over time with a state noise. It predicts the current state θ_t based on past observations and *the state equation* (i.e. $p(\theta_t|\boldsymbol{x}_{t-1})$, predict step), and then updates the state estimation (i.e. $p(\theta_t|\boldsymbol{x}_t)$, update step), by comparing observed

error between the observation prediction (based on $p(\theta_t|\boldsymbol{x}_{t-1})$ and *the observation equation*), and the actual observation $\boldsymbol{x}_t$, so to predict the state at next time step (i.e. $p(\theta_{t+1}|\boldsymbol{x}_t)$). Also note in this framework, it assumes the state of environment can change over time. A brief illustration is shown as follows.

Bayes' theorem: $p(state|data) \propto p(state)\, p(data|state)$

It can also be written as: $p(\theta_t|\boldsymbol{x}_t) \propto p(\theta_t|\boldsymbol{x}_{t-1})\, p(x_t|\, \theta_t)$

That is, the updated filter distribution $p(\theta_t|\boldsymbol{x}_t)$ is found by combining the current predictive distribution $p(\theta_t|\boldsymbol{x}_{t-1})$ (used as a prior) with the new likelihood $p(x_t|\, \theta_t)$ from the incoming observation.

1.2 Optimal Control Theory in Motor-control

While some tasks focus on discrete actions, most tasks in daily activities require continuous planning and execution of complex trajectories. Unlike the simple 2 AFC task in which the cost only depends on the outcome of the decision (final state), in a motor task (e.g. driving to a target location), the cost also includes the physical effort of the movement that can accumulate over the entire trajectory. Optimal control theory addresses this issue by using a cost function that is a weighted mixture of the inaccuracy of reaching task goal and the accumulated effort.

1.2.1 Forward model: Generate optimal movements

In tasks where the objective function (reward function or loss function) is specified by the experimenter, optimal control theory will generate the optimal action sequences that maximize the reward (or minimize the loss). This forward model has been widely applied in the robotic control field. Recently it also has gained attention in human biological movement research (Liu & Todorov 2007; Todorov 2004; 2009) and is shown to be a plausible model to explain the underlying computational principles of human movements.

1.2.2 Inverse model: Estimating reward/loss function

In tasks where the objective function (reward function or loss function) is not specified, we can inverse the process (inverse optimal control), using observed sequence of actions to infer the underlying objective function that is used to optimize the movement (Ng & Russell 2000; Abbeel & Ng 2004). It is also known as inverse reinforcement learning, imitation or apprentice learning. Inferring subjects' reward (or loss) function has been used to understand sensorimotor learning (Kording & Wolpert 2004), urban navigation (Ziebart et al. 2008), and human goal inference (Baker, Saxe and Tenenbaum 2009). A brief illustration of forward/inverse model is shown in Figure 1.2.

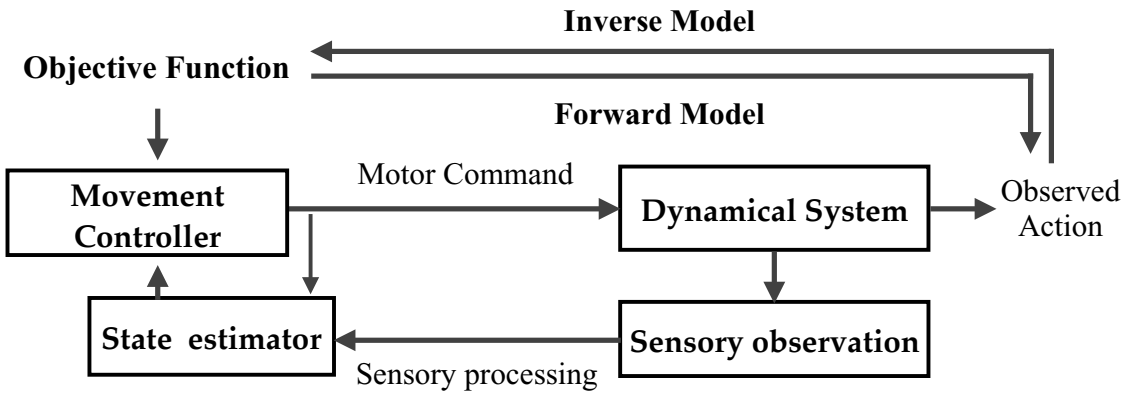

Figure 1.2: Forward/Inverse Model in Optimal Control

With different components (e.g. state estimation, objective function, etc.) in sensorimotor processes, studying discrete decision-making or continuous motor control can address different problems of interest.

1.3 Dissertation Outline

1.3.1 Decision-making at discrete time

Part I (Chapter 2 and Chapter 3) will be focused on perceptual decision-making process. First, we want to explore if human subjects can learn statistical regularities in the environment and utilize this information in decision-making process. In Chapter 2 (Figure 1.3), we present a visual search task in which there is an embedded target distribution among three alternative locations (i.e. 1:3:9 to be the target). Subjects were instructed to locate the target as fast as possible, and were rewarded by correct decision and fewer locations searched. Research questions here are, if subjects can learn the

underlying state of the environment (i.e. the target distribution embedded in the environment), and use the learned statistics to decide a search path (i.e. a sequence of fixations) to optimize the search strategy (e.g. based on the decreasing target probability at those locations). In addition, if we can use Bayesian models to explain how subjects' belief (of the state of the environment) changes over time, and to provide insights of the underlying computational principles of their decision-making process.

In chapter 3 (Figure 1.4), we want to investigate one of the special effects in human sensorimotor processing, the sequential effect (Yu & Cohen 2009). Sequential effect describes the fact that humans respond faster with a higher accuracy to a stimulus that is consistent with a strong local pattern (e.g. A coming after AAAA or ABABA), while they respond slower with a higher error rate to the stimulus that violates such local pattern (B coming after AAAA or ABAB). For this type of behavior, the research question is, if we can find a rationale based on Bayesian approach to explain this seemingly 'superstitious' behavior.

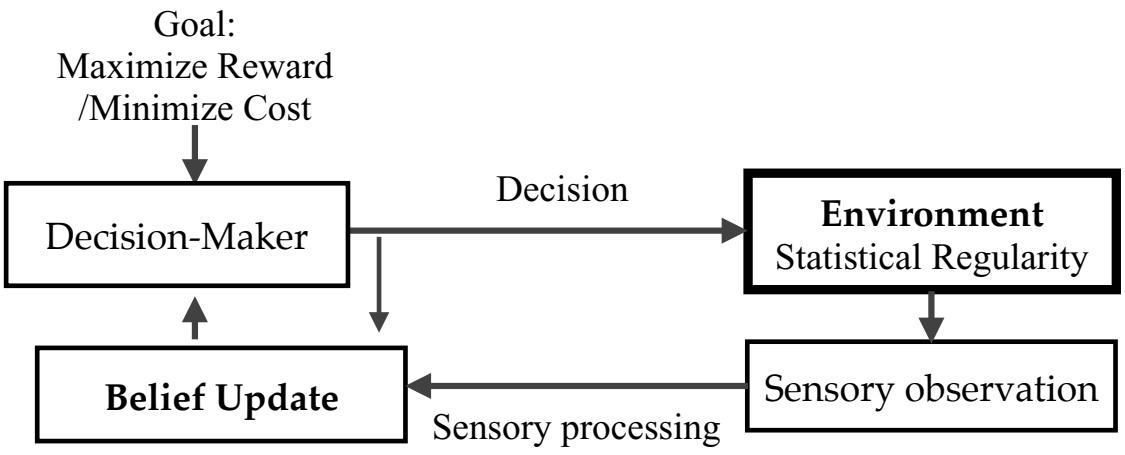

Figure 1.3: The Influence of Environmental statistics in Visual Search Tasks (Chapter 2)

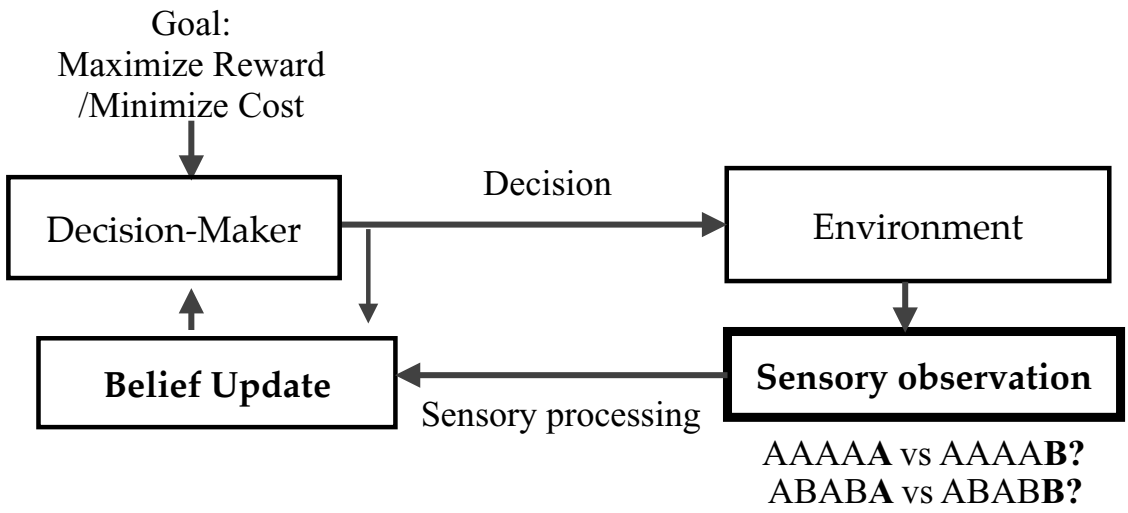

Figure 1.4: Sequential Effect in 2AFC Tasks (Chapter 3)

1.3.2 Motor-control at continuous time

Part II (chapter 4-chapter 7) will be focused on human movement at continuous time. In particular, it seeks to uncover the objective function that human subjects are using to generate the observed movements. In Chapter 4 (Figure 1.5), we want to explore how humans plan saccadic eye movements under uncertainty and noise (e.g. signal-dependent noise). For example, in a rapid reaching task under risk, in which the subject needs to point to a small target location within a limited amount of time, due to the uncertainty of the target location, subjects need to use visual information to provide the state of the environment (target location), and guide the hand movement at continuous time. Here, the research question is, how does the subject minimize the uncertainty of the target by coordinating eye–hand movements at continuous time, and can we find the underlying computational principles to explain observed eye movement and hand movement trajectories.

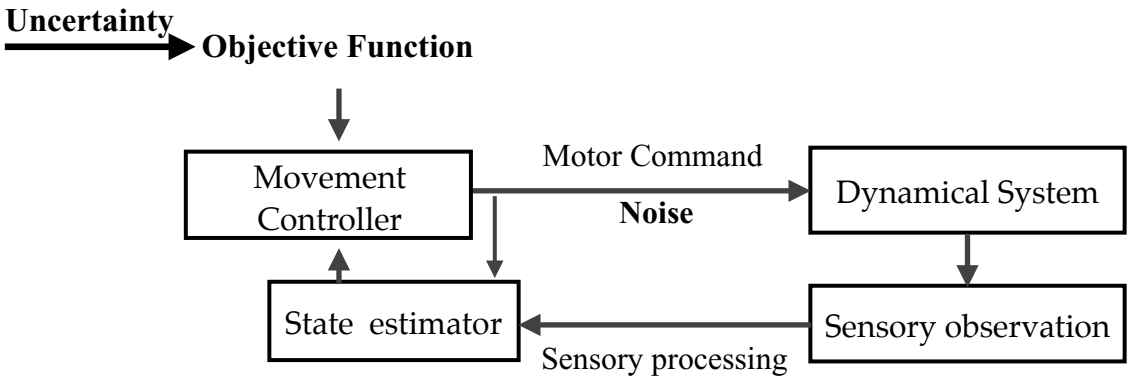

Figure 1.5: Informax model of oculomotor control (Chapter 4)

Another important facet of looking at motor control is to explain observed motor performance impairment in different populations, for example, depressed individuals. Chapters 5-6 (Figure 1.6) are focused on studying the motor deficits in depressed population. In Chapter 5, we proposed a simulated driving task to study impaired motor performance in depressed individuals. Chapter 6 presents an inverse optimal control framework, which can disentangle different components (sensorimotor speed, goal state, motivation) in this goal-directed motor task. We applied this framework in the simulated driving task, to study the influence of depression in sensorimotor system and reward-evaluation.

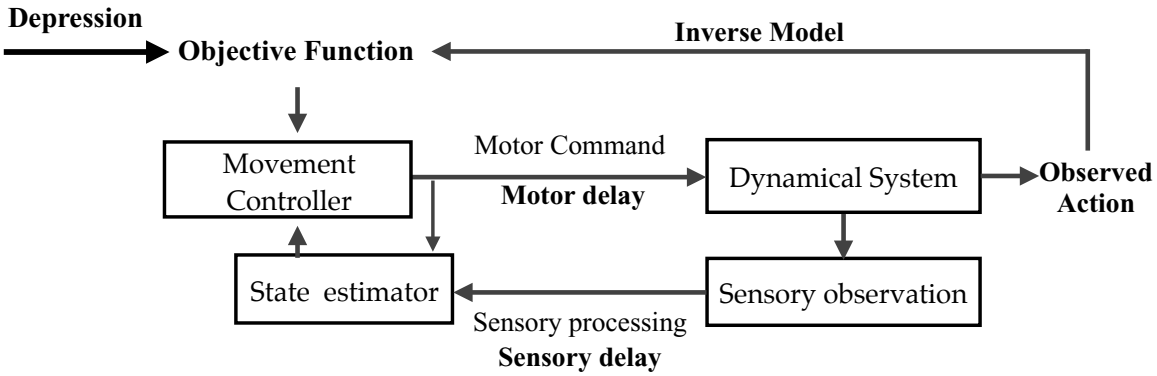

Figure 1.6: Inverse model of depressed behavior (Chapter 5-6)

References

Abbeel, P., & Ng, A. Y. (2004, July). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning* (p. 1). ACM.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329-349.

Berniker, M., & Kording, K. (2011). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 419-428.

Huys, Q. J., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition*, *113*(3), 314-328.

Körding, K. (2007). Decision theory: what" should" the nervous system do?. *Science*, *318*(5850), 606-610.

Körding, K. P., & Wolpert, D. M. (2004). The loss function of sensorimotor learning. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(26), 9839-9842.

Liu, D., & Todorov, E. (2007). Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *The Journal of Neuroscience*, *27*(35), 9354-9368.

Ng, A. Y., & Russell, S. J. (2000, June). Algorithms for inverse reinforcement learning. In *Icml* (pp. 663-670).

Rigoux, L., & Guigon, E. (2012). A model of reward-and effort-based optimal decision making and motor control. *PLoS computational biology*, *8*(10), e1002716.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.

Trommershauser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial vision*, *16*(3), 255-275.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature neuroscience*, *7*(9), 907-915.

Todorov, E. (2009). Efficient computation of optimal actions. *Proceedings of the national academy of sciences*, *106*(28), 11478-11483.

Vilares, I., & Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22-39.

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current opinion in neurobiology*, *22*(6), 996.

Yu, A. J. (2007). Adaptive behavior: Humans act as Bayesian learners. *Current Biology*, *17*(22), R977-R980.
Yu, A. J., & Cohen, J. D. (2009). Sequential effects: superstition or rational behavior?. In *Advances in neural information processing systems* (pp. 1873-1880).

Yuille, A. L., & Heinrich, H. B. (1994). Bayesian Decision Theory and Psychophysics.

Ziebart, B. D., Maas, A. L., Bagnell, J. A., & Dey, A. K. (2008, July). Maximum Entropy Inverse Reinforcement Learning. In *AAAI* (pp. 1433-1438).

Chapter 2

Maximizing masquerading as matching: Statistical learning and decision-making in choice behavior

Abstract: There has been a long-running debate over whether humans *match* or *maximize* when faced with differentially rewarding options under conditions of uncertainty. While maximizing, i.e. consistently choosing the most rewarding option, is theoretically optimal, humans have often been observed to match, i.e. allocating choices stochastically in proportion to the underlying reward rates. Previous models assumed matching behavior to arise from biological limitations or heuristic decision strategies; this, however, would stand in curious contrast to the accumulating evidence that humans have sophisticated machinery for tracking environ- mental statistics. It begs the questions of why the brain would build sophisticated representations of environmental statistics, only then to adopt a heuristic decision policy that fails to take full advantage of that information. Here, we revisit this debate by presenting data from a novel visual search task, which are shown to favor a particular Bayesian inference and decision-making account over other heuristic and normative models. Specifically, while subjects' first-fixation strategy appears to indicate matching in aggregate data, they actually maximize

on a finer, trial-by-trial timescale, based on continuously updated internal beliefs about the spatial distribution of potential target locations. In other words, matching-like stochasticity in human visual search is neither random nor heuristics-based, but due specifically to fluctuating beliefs about stimulus statistics. These results not only shed light on the matching versus maximizing debate, but also more broadly on human decision-making strategies under conditions of uncertainty.

Motivation

The cognitive process involved in decision-making leading to complex actions is difficult to investigate. However, over the past decade, remarkable progress has been made in studies of visual-saccadic decision-making (Hanes & Schall 1996; Shadlen & Newsome 1996, 2001), an experimentally accessible approach to understand decision making in general.

Features of Decision-Making Process

Decision-making has some important features: active, sequential and goal-driven.

Active: To select the best choice among a set of alternatives, we always actively explore the resources to collect valuable information and use those 'as evidence' to support or against existing alternatives.

Sequential**:** Any decision is a consequence of a series of decisions following the goal, which sequentially affects each other over time. Also, experience and prior knowledge will affect one's decision-making process as well.

Goal-driven: Any decision-making process coexists with a goal, which is used to guide

the information collection, information process and action selection. This ‘gain’ is measured as a reward from selecting one action from a set of alternatives, and the goal of decision-making is to maximize the reward.

Features of saccadic-visual system

Our vision system also shares those features.

Active*:* Saccadic eye movement in visual-search is an active process, because what we see is influenced by where we fixate, through actively using fovea vision to focus on the interesting area to get the sharp vision.

Sequential: Our eyes explore the world in a series of fixations connected by saccadic eye movements. In a visual-search task, prior experience and knowledge of the "visual target" will affect the planning of the saccades’ sequence. For example, if you are looking for a cup in a room, most likely you will start from the desk, instead of the ceiling or floor.

Goal-driven: The world is filled with noisy sensory information. In order to efficiently interact with the environment, our eye movement needs to quickly gather useful information in the noisy world.

Decisions involved in saccadic eye movement

With limited sensing abilities from degraded vision of the periphery, it is extremely important for our brain to efficiently acquire sensory information in a goal-driven manner. Actions in a visual-search task are the result of two basic decisions: *where to look at* and *when to look away* from one location to the other. The first action of *where to look* entails motor planning, which is affected by observer's goal, environment, and expectation. The second action of *when to saccade away* from one location to the other entails the interaction of motor planning and sensory processing, in which the visual

information collected during sensory processing will determine if the observer successfully achieves his goal or need to go to next location.

Choice of fixations (*where to look at*) is not a trivial process in visual-search, but reflects how we actively decide where to acquire sensory information to efficiently achieve the goal. Extensive studies of eye fixations have been done in past decades. It has been proposed that eye movements are directed to maximize the information gain about the target (Najemnik & Geisler 2005; Butko & Movellan, 2010), which for both a static but also a moving target (Land & McLeod 2000; Talbott, Huang & Movellan 2012). It also has been shown the spatial information of the target can be used to facilitate the search process (Zelinsky & Sheinberg 1997; Walthew & Gilchrist 2006), and subjects may use a search strategy that maximize the current posterior probability of the target location (MAP), thus favoring the most likely location (Najemnik & Geisler 2008). However, so far, it is still unclear how human learn the spatial statistics and use them in planning saccadic search sequence. For example, Araujo, Kowler & Pavel 2001 showed that majority of the subjects in their study did not use the probability cue of the target and did not select the high-probability target as the first fixation. Studies are also indicating subjects match their response probabilities to the payoff probabilities instead of choosing the most rewarding action that maximizes the probability (probability matching vs. maximization; Herrnstein 1961; Koehler & James 2009). Recently, it has been proposed (Gaissmaier & Schooler 2008) that probability matching, a seemingly irrational and sub-optimal strategy may be a ‘smart’ strategy that facilitates subjects to ‘search’ the underlying statistical pattern. Alternatively, there are also studies shown that, due to limited memory, subjects may use ‘maximization’ strategy but only for a limited time

window, which is known as melioration theory (Herrnstein & Prelec 1991). Indeed, using spatial statistics in eye movement planning still remains largely unexplained.

Choice of fixation duration (*when to look away),* another decision in saccadic eye movement planning, is influenced by the sensory strength of the stimulus. In a random-dot direction discrimination task (Gold & Shadlen 2000), it has been shown that with increasing uncertainty of the sensory information (i.e. decreasing coherence level of the primary motion) in the stimulus, there will be increasing view time and thus longer reaction time to the identification of the target. With the presence of sensory uncertainty, our current knowledge of how human subjects plan eye movement under both target spatial uncertainty and stimulus sensory uncertainty is extremely limited.

Here we propose to use a visual-search task with random-dot stimulus to investigate if human subjects can internalize spatial statistics and use this information in guiding eye movement planning. In particular (Chapter 2), we examine the matching vs. maximizing strategy with the assumption that matching behavior may be due to subject's internal belief of the stability of the spatial statistics in the environment, which is coupled but not dissociated with the maximizing decision process. We achieve this by using two modified versions of Bayesian models (Dynamic Belief Model/DBM and Fixed Belief Model/FBM; Yu & Cohen 2009) and compare human fixation choices to different models (DBM + match, DBM + max, FBM + match, and FBM + max, melioration). We also use this task to investigate if/how fixation duration is influenced by the spatial statistics and sensory uncertainty (see Ahmad, Huang & Yu 2014), in particular, if subjects' fixation duration is biased by their prior belief of the target probability (i.e. confirmation bias).

2.1 Introduction

There has been a long history of debate over whether humans and animals *match* (Herrnstein, 1961) or *maximize* (Hall-Johnson & Poling, 1984; Blakely, Starin, & Poling, 1988), when choosing among options with unequal rates or probabilities of reward. While maximizing, or consistently choosing the most rewarding option, is theoretically optimal (greatest cumulative accuracy or reward in the long-term), there is a substantial body of literature indicating a curious tendency for humans and animals to match (e.g. Herrnstein, 1961; Sugrue, Corrado, & Newsome, 2004), or to allocate their choices in approximate proportion to the underlying reward rates.

This apparently sub-optimal stochasticity in choice behavior has been interpreted as either a con- sequence of biological limitations or, relatedly, a fast and frugal heuristic for coping with hard decision problems. For example, one prominent account is *melioration theory* (Herrnstein, 1970) and the formally equivalent "Take-the-Best" (TTB) heuristic (Gigerenzer & Goldstein, 1996): it posits that humans and animals have a limited memory buffer and that they choose the best (maximizing) option based on only a few recent data points, so that on average the choice policy will appear stochastic due to fluctuations in empirical statistics. An even simpler proposal is that sub- jects base each decision entirely on the last trial, by persisting with the same choice when met with success, and switching otherwise; it is known as the "Win-Stay-Lose-Shift" (WSLS) algorithm (Rapoport & Chammah, 1965; Nowak & Sigmund, 1993; Randall & Zentall, 1997; Warren, 1966; Steyvers, Lee, & Wagenmakers, 2009; Lee, Zhang, Munro, & Steyvers, 2011) for binary choices, and Take-the-Last (TTL) when there are more than

two choices (Gigerenzer & Goldstein, 1996). Separately, it has been suggested that matching can serve as a heuristic exploration strategy in a noisy and changeable environment, so that the decision-maker does not persevere with outdated choices (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

The shared assumption of these heuristic accounts, that humans and animals are incapable of utilizing sophisticated decision strategies, stands in curious contrast with the converging behavioral and neurophysiological evidence that the brain possesses the machinery to near-optimally track evolving environmental statistics (Yu & Dayan, 2005b; Daw et al., 2006; Behrens, Woolrich, Walton, & Rushworth, 2007; Nassar et al., 2012; Ide, Shenoy, Yu*, & Li*, 2013). It begs the questions why the brain would build sophisticated representations of environmental statistics, only then to adopt a heuristic decision policy that fails to take full advantage of that information.

Here, we re-examine this matching vs. maximizing debate within a Bayesian ideal observer framework (Green & Swets, 1966), specifically proposing the hypothesis that humans continuously track environmental statistics with an implicit assumption that the world can change at any moment; consequently, they choose where to search at any given time using an optimal, maximizing strategy but which is based on their dynamically evolving beliefs about environmental statistics. In other words, we hypothesize that the matching-like choice behavior is not due to random exploration or limitations of memory, but due to specific fluctuations in internal beliefs about environmental statistics, coupled with an optimal (maximizing) decision process. The hypothesis that humans have a natural tendency to extract statistical patterns while assuming such patterns are changeable over time is motivated by our previous work (Yu & Cohen, 2009), showing

that a similar hypothesis can explain a classical sequential effect in 2-alternative forced choice (2AFC) tasks, in which responses are faster and more accurate if a stimulus extends a recent run of repetitions or alternations, and conversely slower and less accurate when a stimulus violates such a run, even if these runs arise purely by chance (Soetens, Boer, & Hueting, 1985). Subjects appear to act with the implicit assumption that the world is potentially changeable – giving more recent observations greater emphasis in predicting future outcomes, instead of giving uniform weights to the entire history of data (Yu & Cohen, 2009; Wilder, Jones, & Mozer, 2010). Here, we adopt a similar Bayesian modeling framework to examine whether subjects depend more on the recent trials to predict next target location (Dynamic Belief Model; DBM) or treat all previous data equivalently when making that prediction (Fixed Belief Model; FBM). These two candidate models for prior learning/updating reflect differing statistical assumptions that either do (DBM) or do not (FBM) allow the possibility of un-signaled, discrete changes in the statistical regularities in the environment.

To examine this hypothesis, we obtain behavioral data from a novel visual search paradigm, in which subjects can exploit statistical regularities in target location in order to improve the accuracy and efficiency of their search strategy. The key behavioral measure is how subjects allocate their first fixation choice: do they simply follow the last trials target location (WSLS/TTL), do they choose stochastically in proportion to the underlying target distribution (matching), or do they systematically choose the most probable target location (maximizing)? We adopt a visual search task because it has long been known that saccadic eye movements are influenced by various cognitive factors (Yarbus, 1967), such as prior knowledge about target location (He & Kowler, 1989;

Einha¨user, Rutishauser, & Koch, 2008), temporal onset (Oswal, Ogden, & Carpenter, 2007), reward probabilities (Roesch & Olson, 2003), and general scene context (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009). In daily tasks, saccadic patterns have been observed to be different among visual search, scene memorization (Henderson, 2007), reading (Rayner, 1998), tea and sandwich making (Land & Hayhoe, 2001), and driving (Land & Lee, 1994). It is poorly under- stood how such contextual knowledge is acquired and how it precisely modulates saccadic choices and perceptual decisions – a scientific lacuna we address here. Compared to the rather abstract or artificial stimuli more commonly used in choice tasks, we expect human subjects to be particularly adept at internalizing and utilizing the spatial statistics of visual targets.

In this work, we compare human fixation choice behavior to the predictions of various models. We consider four Bayesian models, which differ in the assumptions they make about *statistical learning* and *decision-making*. Statistical learning refers to the observers' internal representation of the target location statistics, and the sequential updating of the prior distribution over where the target lies based on experienced outcomes. We examine two variants of Bayesian learning models, DBM and FBM. We also examine two different decision processes: (1) *Match*, which produces saccade fixation locations in proportion to the internal predictive distribution of target location, and (2) *Max*, which always chooses to first search the currently most probable target location. Thus, there are four Bayesian models altogether, DBM+Match, DBM+Max, FBM+Match, FBM+Max. In addition, we also consider a heuristic algorithm, *melioration*, which does not require a sophisticated statistical representation. We note that melioration is equivalent to TTB, and subsumes TTL as a special case, where the

memory buffer is just the last trial. In the following, we first describe the experimental design and present some basic data. We then use a series of data-model comparison to narrow down the best model for explaining human data.

2.2 Results

We first briefly describe the visual search task (see Methods for more details), before delving into the experimental findings and comparison to the various models. In the task (Figure 2.1a), 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 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).

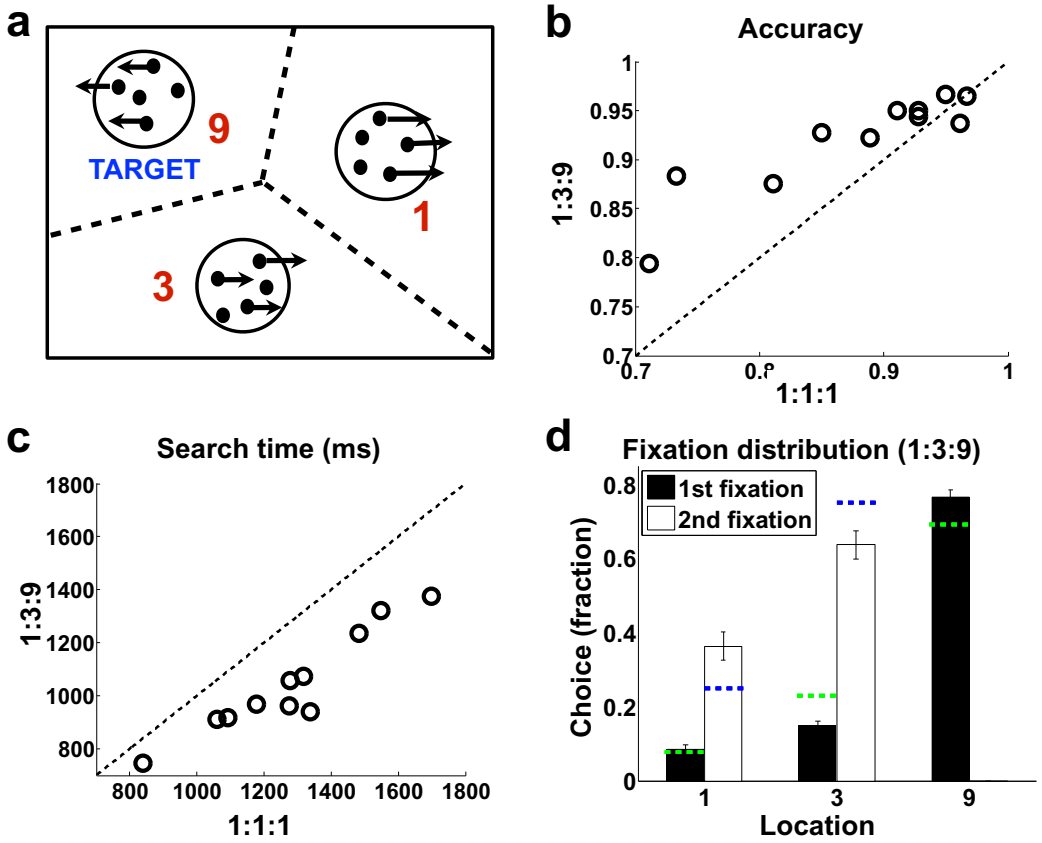

Figure 2.1: Experimental design and data. (a) On each trial, two of the random-dot stimuli are distractors, one is the target; subjects must find the target (see Methods). (b) Subjects are more accurate in finding the target in the 1:3:9 condition than the 1:1:1 condition, and (c) faster. (d) 1:3:9 condition, 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.

We found that human subjects indeed internalized and exploited the spatial statistics to locate the target stimulus more accurately (Figure 2.1b) and rapidly (Figure 2.1c). 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 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. For the first fixation, subjects preferentially fixated the 9 location over the 3 location ($p < 0.0001$), which in turn was favored over the 1 location ($p < 0.01$). Similarly, for the 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 ($p = 0.005$). Altogether, these results indicate that subjects not only knew where the most probable target location was, but had a graded representation of target probabilities at the different locations.

Aggregate statistics are coarse by nature, and much information is lost by averaging all trials together. In particular, it ignores the potential role of statistical learning. Subjects could be learning about spatial statistics based on each experienced trial, in a manner similar to DBM or FBM, and thus a Match or Max strategy should be defined with respect to their internal representation at each moment in time, instead of with respect to the "true" generative statistics (1:3:9 or 1:1:1), which they have no direct access to. Details of DBM and FBM can be found in Methods; here, we briefly describe their assumptions and properties. The versions of DBM and FBM used here are mult-alternative extensions of simpler models we previously developed for 2AFC tasks (Yu & Cohen, 2009). Although the true configuration of most probable, medium probable, and least probable target location does not change within a block (it is pseudo-randomized across blocks), and the relative odds for those locations remain at 1:3:9 for all blocks, DBM allows the possibility that subjects assume the underlying statistics to be changeable within a block. We entertain this hypothesis here, because we previously showed that, in 2AFC tasks, subjects act as though they assume the relative probability of

stimulus type is predictable from recent trial history, consistent with DBM (Yu & Cohen, 2009), even though the true experimental statistics are constant (and random) throughout the experiment. Figure 2.2a shows the generative model for DBM and FBM. Figure 2.2b shows a sample run of DBM on an actual experienced sequence of trials for a subject. The predictive probability DBM assigns to each of the potential target locations on each trial fluctuates with the recent history of experienced trials. DBM+max produces fixation predictions that closely correspond to this subject's actual choices (top panel). The rare discrepancies occur when there are unexpected observations and the underlying probabilities are close in magnitude: for example, trial 80. Most of the time, even when the underlying probabilities are similar due to unexpected observations, DBM+max and the subject concur in switching (57, 78) or staying (64, 71, 75, 76, 77, 79, 83, 85).

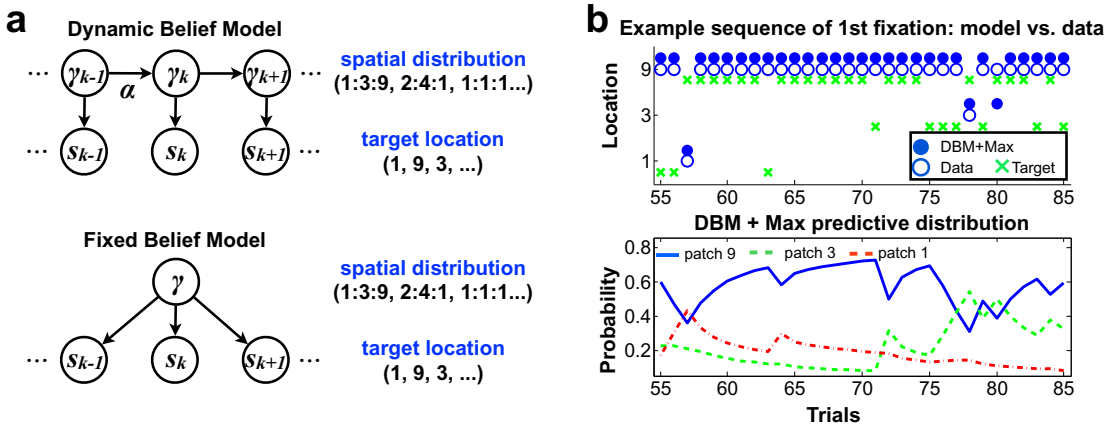

Figure 2.2: Dynamic Belief Model (DBM) and Fixed Belief Model (FBM) model architecture and comparison to behavioral data. (a) graphical model for DBM and FBM. FBM can be thought of as a special case of DBM, with $\alpha = 1$. (b) An example trial sequence (subject 2, block 4, trials 55-85) and corresponding DBM inference/prediction behavior ($\alpha = 0.92$). Given the sequence of target location experienced by a subject (green crosses in top panel), DBM computes the *predictive* probability on each trial k of each location containing the target (bottom panel), and the maximum is taken as the model prediction of choice (filled blue circle, top panel); compared to the subject's actual first fixation location (open blue circle, top panel).

In addition to the Bayesian models, we entertain the possibility that subjects employ melioration (Herrnstein, 1970), which keeps a limited and fixed memory of the last k trials, and picks the most frequent target location among these k trials as next trial's first fixation choice (ties are broken randomly). It is clear that subjects would show average "matching" behavior even if they always first searched in the last target's location (k = 1), because this choice distribution would exactly track the empirical target distribution, with a one-trial lag that would not be apparent in aggregate data. We fit the best memory size (number of recent trials kept in the buffer) for the melioration model, which was left as a free parameter in the original model (Herrnstein, 1970). Note that this subsumes the TTL/WSLS model as a special case with the memory size being 1 trial. By simulating the melioration model with different memory sizes (from 1 to 10), and comparing the choice predicted by the model and subjects' actual first fixation location, we found that the a memory size of 3 trials was the best at predicting subjects' choice.

To distinguish among the models, we examined the evolution of fixation choice pattern of humans, in comparison to the various models. Over the time course of a block, we found that subjects gradually learn to favor the more probable locations, in a manner well matched by both FBM+Match and DBM+Max (Figure 2.3a;b). In contrast, FBM+Max over-matched (Figure 2.3a, solid) and DBM+Match (Figure 2.3b, dashed) under-matched subjects' fixation choice distribution. Note that the fact that the human behavior and model traces reach asymptotic values pretty early (Figure 2.3a;b) is not necessarily indicative of the cessation of learning after the curves flatten out. *Where* the curves asymptote reflect more on learning, such that if human behavior flattened out at global maximizing (as in FBM+Max in Figure 2.3a), then that would indicate cessation

of learning. But in fact, human behavior flattens out at a level well below global maximizing, indicating either significant asymptotic learning (DBM+max) or asymptotic random stochasticity (FBM+match).

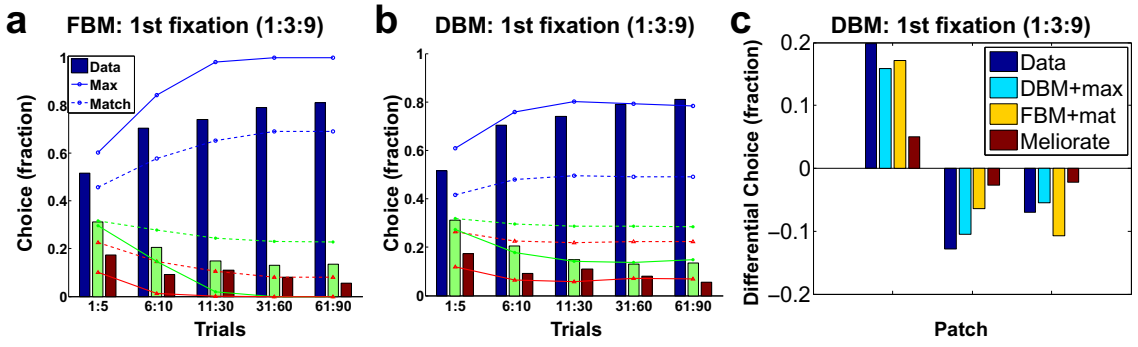

Figure 2.3: Model comparison of fixation distributions. (a) Averaged over subjects (n = 11) and blocks (6), first fixation location (colored bars: blue=9, green=3, red=1) increasingly favor the 9 location over 3, and in turn over 1, for different segments of the 90-trial blocks. FBM+max (solid line) predicts faster learning/over-matching compared to subjects; FBM+match produces predictions more similar to subjects' data. (b) Human data (colored bars) same as in (a). DBM+max produces predictions similar to subjects' data; DBM+match produces slower learning/under-matching compared to subjects. All predictions based on actual sequences of trials experienced by subjects. Errorbars = s.e.m. over blocks and subjects. (c) The average distribution of first-fixation choice, among the three patches, for the last ten trials of each block minus average choice distribution in the first ten trials.

This data also rule out melioration/TTB/TTL heuristic strategies, since they would produce choice statistics that are constant over the block (because they track empirical statistics, which are on average constant over the block). Fig. 3c shows the model comparison in a different way. When we look at the difference in subjects' average fixation distribution between the last ten trials and the first ten trials of each block, we find that subjects favor the 9 location much more relative to the 3 and 1 locations, toward the end of the block than at the beginning of the block. This trend is best captured by DBM+max (on average 20% different from data in absolute units, as

shown in Figure 2.3c), second best by FBM+match (39% absolute difference from data), and very poorly by melioration (74% absolute difference from data, memory size = 3, best-fitting parameter setting).

Left with two remaining model candidates, DBM+Max and FBM+Match, we next examine subjects' fixation choice distribution *conditioned* on the last target location. In FBM, new data have less and less capacity to shift the posterior as the total amount of data builds and the posterior becomes more rigid (Yu & Cohen, 2009), thus predicting that last trial's target location to have little to no effect on current trial fixation choice when averaged over the whole experimental session (Figure 2.4c). In contrast, DBM continuously entertains the possibility of change, and thus allows its posterior to shift according to new data (Yu & Cohen, 2009), thus predicting that last trial's target location should have asymptotically non-trivial effect on current trial fixation location (Figure 2.4b). Fig 2.4a shows that human subjects behave as predicted by DBM: while the 9 location is most frequently the first fixation location in all cases, its advantage is much reduced if the last target was in location 1 or 3, and increased if the last target was in location 9. More specifically, following the target being in location 1, first fixation percentage is boosted in location 1 on the subsequent trial; similarly, following the target in in location 3, first fixation percentage is boosted in location 3 on the subsequent trial. This indicates that subjects are not adopting the same stochastic "matching" policy on every trial, but are exquisitely sensitive to recent trial history. While DBM+max yields conditional distributions statistically indistinguishable from subjects' first fixation distributions (one-sided t-test of average Kullback-Leibler Divergence between subjects' conditional distributions and that of DBM+max, one-sided paired t-test, $p = 0.076$),

FBM+match produces conditional distributions that are significantly different (one-sided paired t-test, $p < 0.001$).

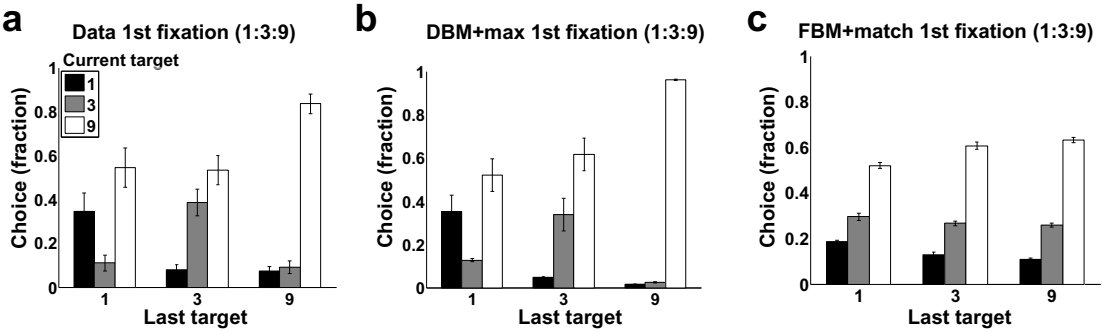

Figure 2.4: Data vs. model prediction of first fixation distribution conditioned on last target location. (a) In the 1:3:9 condition, human subjects generally prefer 9 (white) over 1 (black) and 3 (gray) regardless of last trial target location, but location 1 is particularly favored if the last target location was also 1 (left group), location 2 is particularly favored if the last target location was also 3 (middle group), and location 9 is particularly favored if the last target was also 9 (right group). (b) DBM+max produces a conditional distribution very similar to the behavioral data in (a). (c) FBM+match produces a conditional distribution dissimilar to the behavioral data in (a), in particular it is relatively insensitive to last trial target location. All model predictions based on actual sequences of stimuli subjects experienced. Errorbars = s.e.m. over blocks and subjects.

A different analysis shows the impressive advantage DBM+max has over other models not only in characterizing gross empirical statistics, but also in predicting trial-to-trial first fixation choice. For each model, we compute on each trial the prior probability each model assigns to the subject's *actual* first-fixation choice. It is 1 or 0 for deterministic models (Follow-Last-Trial and Maximization), for correct and incorrect predictions, and somewhere in between for stochastic models (Match). For DBM, we first find the best-matching α parameter for each subject, which was on average 0.87, with a standard deviation of 0.15. We find that DBM+max has an average predictive accuracy of 0.81, far outperforming TTL or melioration with a 1-trial buffer (0.54), FBM+Match

(0.58), DBM+match (0.59), FBM+max (0.78), $p < 0.01$ in each case (two-sided t-test). This big advantage is not due to the extra parameter, α, in DBM, which specifies an individual's belief about the probability of the target location statistics *not* changing from trial to trial (in contrast to FBM, which has no free parameters), as leave-one-block-out cross validation yields a predictive accuracy of 0.80, which is statistically indistinguishable from the training data predictive accuracy (two-tailed t-test, $p = 0.728$). That DBM is not over-fitting is hardly surprising, as we are in the realm of many more data points (540 trials per subject) than parameters (1 per subject).

While DBM+max can predict an individual's first fixation choice about 80% of the time, i.e. subjects choose the most probable target location 80% of the time, subjects do choose the other two locations about 20% of the time. We find that subjects favor the more probable of the remaining two options instead of choosing among equally often (12.2% versus 6.7%, $p = 0.016$). This raises the possibility that subjects may be applying some sort of softmax decision policy that is in be- tween matching and maximizing. We therefore fit a choice distribution (q_1, q_2, q_3) that is a softmax function of the belief distribution (p_1, p_2, p_3):

$$q_i = \frac{p_i^\beta}{\sum_{j\in\{h,m,l\}} p_j^\beta} \,. \tag{1}$$

We use a polynomial form of softmax instead of an exponential form, seen in related work, (e.g. as in Daw et al., 2006), because the polynomial form has a natural interpretation in terms of matching ($\beta = 1$), under-matching ($\beta < 1$), and over-matching ($\beta > 1$). We find that the best-fitting β values across subjects are significantly greater than 1

($p < 0.00002$), with a mean value of 3.27 (std = 1.00). $\beta = 3.27$ suggests a very strong tendency to maximize, as it would, for example, turn (p1, p2, p3) = (9/13, 3/13, 1/13) into (q1, q2, q3) = (0.973, 0.027, 0.001).

Finally, we note that a thorough analysis of the uniform condition behavior is omitted, because behavior in those blocks is completely unconstrained. Subjects can choose to employ whatever idiosyncratic strategy they like (including, for example, always starting from the same location, or systematically rotate through them, A, B, C, A, B, C, ...), and they would on average perform exactly equally well. We find that, indeed, there is quite a bit of variability of first-fixation strategy among subjects on the uniform blocks.

2.3 Methods

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, Shadlen, Newsome, & Movshon, 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. 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) was 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 coherences 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.

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 which stimulus was shown at any given time are as shown in Fig 2.1a. 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) even after the red bar reaches the edge of the display. 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 × (seconds taken to respond) + 25×(number of fixation switches, 0 if only one patch fixated) ± 50 (+ if final response correct, - if incorrect). Subjects were told about the reward scheme at the start of the experiment, in addition to receiving detailed feedback, as explained above, during the experiment. 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.

Subjects were excluded from the study if they did not have normal or corrected vision, 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).

Theoretical Modeling

To distinguish the two hypotheses that stochasticity in subjects' saccade choices arise from stable beliefs about target location statistics plus matching-like randomness in action selection, versus fluctuating beliefs about target location statics plus a maximizing strategy, we implement two distinct Bayesian models of trial-by-trial statistical learning: one that assumes that subjects use the entire history of observed data in the current block to infer about the target location statistics (Fixed Belief Model, FBM), and another that assumes that subjects use only recent history to make inferences about target location

(Dynamic Belief Model, or DBM). The versions of DBM and FBM used here are multi-alternative extensions of simpler models we previously developed for 2AFC tasks (Yu & Cohen, 2009).

We first describe DBM, and then explain how FBM differs. In the DBM, we model subjects' trial-by-trial inference using a hidden Markov model. In the generative model (Fig. 2.2a), the target location on trial k, denoted by $s_k \in \{1, 2, 3\}$, depends on the configuration b_k and multinomial parameters $\gamma_k := (\gamma_h, \gamma_m, \gamma_l)$, where $\gamma_h + \gamma_m + \gamma_l = 1$:

$$P(s_k|b_k, \boldsymbol{\gamma}_k) = \begin{cases} (\frac{1}{3}, \frac{1}{3}, \frac{1}{3},) & b_k = 1 \\ (\gamma_h, \gamma_m, \gamma_l), & b_k = 2 \\ (\gamma_h, \gamma_l, \gamma_m), & b_k = 3 \\ (\gamma_m, \gamma_h, \gamma_l), & b_k = 4 \\ (\gamma_m, \gamma_l, \gamma_h), & b_k = 5 \\ (\gamma_l, \gamma_h, \gamma_m), & b_k = 6 \\ (\gamma_l, \gamma_m, \gamma_h), & b_k = 7 \end{cases} \tag{2}$$

At the start of each experimental block, the prior distribution over b_1 and γ_1 on the first trial are $p_0(\gamma)p_0(b)$, where $p_0(\gamma)$ is a Dirichlet distribution Dir(γ;9/13, 3/13,1/13), and

$$p_0(b) = \begin{cases} 1/4 & b = 1 \\ 1/8 & b = 2 \dots 7 \end{cases} \tag{3}$$

In order to capture the assumption of a non-stationarity, we assume (b_k , γ_k) to be subject to change: on each trial, it has probability α of remaining the same as the last trial, and probability $1 - \alpha$ of being redrawn from the prior distribution p_0 (b_k , γ_k).

$$P(b_k, \boldsymbol{\gamma}_k | b_{k-1}, \boldsymbol{\gamma}_{k-1}) = \alpha\delta((b_k, \boldsymbol{\gamma}_k) - (b_{k-1}, \boldsymbol{\gamma}_{k-1})) + (1-\alpha)p_0(b_k, \boldsymbol{\gamma}_k) \quad (4)$$

The recognition model inverts the above generative process to *infer* the current parameter values γ_k and configuration b_k based on observed target locations $s_k := (s_1 \ldots s_k)$. This inference of the joint distribution over (b_k, γ_k) can be computed iteratively as follows:

$$P(b_k, \boldsymbol{\gamma}_k | \mathbf{s}_k) \propto P(s_k | b_k, \boldsymbol{\gamma}_k) P(b_k, \boldsymbol{\gamma}_k | \mathbf{s}_{k-1}) \quad (5)$$

$$P(b_k, \boldsymbol{\gamma}_k | \mathbf{s}_{k-1}) = \alpha P(b_{k-1}, \boldsymbol{\gamma}_{k-1} | \mathbf{s}_{k-1}) + (1-\alpha)p_0(b_k, \boldsymbol{\gamma}_k) \quad (6)$$

The *predictive distribution* over the upcoming target location can be computed by marginalizing out model variables:

$$P(s_k | \mathbf{s}_{k-1}) = \sum_{b_k} \int_{\gamma_k} P(s_k | b_k, \boldsymbol{\gamma}_k) P(b_k, \boldsymbol{\gamma}_k | \mathbf{s}_{k-1}) d\boldsymbol{\gamma}_k \quad (7)$$

FBM differs from DBM in that it assumes that the target statistics are stable over time. It can bethought of a special case of DBM, in which $\alpha = 1$.

Data analysis: model predictive accuracy

We use the average predictive probability of subjects' first fixation under each model for comparing among the various models. For DBM+max and FBM+max, the predictive probability on a trial is 1 if the model successfully predicted the subject's first fixation choice on that trial, and 0 otherwise. For DBM+match and FBM+match, the predictive probability refers to the probability assigned by each model to the subject's choice. These predictive probabilities are then averaged across subjects and blocks of trials.

For DBM, we use leave-block-out cross-validation performance to assess any potential over-fitting due to the α parameter, the sole free parameter in the model. For the 1:3:9 condition, for each subject, the average predictive probability on 5 nonuniform blocks was minimized by selecting α from a range of values covering the range [0, 1]. The model's performance was then evaluated, for the chosen value of α, on the held-out 6th nonuniform block. This procedure was repeated with each block as hold-out set, and the average "test-data" predictive accuracy is reported in the main text.

2.4 Discussion

The experimental and theoretical results in this study shed new light on the debate of matching versus maximizing in choice behavior. As evidenced by our careful model-based analysis of human fixation choice behavior in a novel visual search task, apparently matching-like behavior actually arises from a dynamically evolving Bayesian learning process, combined with a maximizing decision policy. We find that humans readily internalize spatial statistics after just a handful of exemplars, and use that information to improve accuracy and efficiency in target search by biasing saccadic choice in a systematic manner. This work shows that the control of eye movements is not only sensitive to low-level sensori-motor factors previously identified, such as saliency (Itti & Koch, 2000) or long-term ones such as visual acuity map (Najemnik & Geisler, 2005), but also to dynamically changing contextual factors such as evolving spatial knowledge about the spatial distribution of target location.

Our results contrast with most previous work on choice behavior under conditions

of uncertainty, which attributed matching-like behavior of subjects solely to inherent stochasticity in decision policy (Sugrue, Corrado, & Newsome, 2005; Daw et al., 2006; Vul, Goodman, Griffiths, & Tenenbaum, 2009); it builds on recent work suggesting that humans are continuously learning about environmental statistics due to an implicit assumption of a changing world (Yu & Cohen, 2009) and that matching-like behavior may arise from mistuned prior probabilistic beliefs rather than a truly sub-optimal decision policy (Grene, Benson, Kersten, & Schrater, 2010). The specific contribution we make here is that matching-like behavior in part arises from a maximizing (and optimal) choice strategy based on stochastic beliefs about stimulus statistics, which are driven by chance fluctuations in empirical statistics over the experimental session. While the non-stationary assumption seems sub-optimal in our experimental context, it would be a valuable asset in natural environments where statistical regularities do change over time, such as seasonal weather patterns, rise and fall in predator and prey populations, financial and economic markets, and so on. We hypothesize that the apparently irrational matching behavior is an adaptive response to the inherent non-stationarity in natural environments, and that the variability in how close subjects act like a "matcher" versus a "maximizer" may arise from implicit assumptions about the stability of environmental statistics in a particular behavioral context.

We found in the study that as a population, human subjects tend to act as though they believe environmental statistics could be changing on the order of once every seven to eight trials (corresponding to the population mean of the best-fittingα value of 0.87), though there was significant variability across the population. We also found that subjects' choice distribution significantly over-matched with respect to their prior

probability distribution (with an average polynomial ex- ponent of 3.27), again with significant variability across the population. It is possible that subjects may not be exactly maximizing, but injecting a certain amount of stochasticity into their choice pol- icy that is still highly over-matching. However, with the current experimental design, it is difficult if not impossible to distinguish whether there is true stochasticity in their choice policy, or whether there is *apparent* noise due to DBM still not being quite the correct learning model. Indeed, it has been shown in a two-alternative forced choice task that a modification to the DBM learning model, termed DBM2 (Dynamic Belief Mixture Model), captures certain minor but systematic aspects of sequential adjustment not captured by DBM (Wilder et al., 2010). Future experimental and modeling work will certainly be helpful to further unravel the precise nature of learning and decision-making in these tasks.

While subjects behave as bounded rational observers, operationalized as iterative Bayesian inference, we note that this work does not necessarily imply that doing the task requires explicit representation of probabilities. The brain evolved under the selective pressure to approach statistical optimality, but may do so mechanistically without any explicit representation or understanding of probabilities. Indeed, previously we have shown that the predictive probabilities yielded by the Bayesian Model, DBM, can be well approximated by simple leaky accumulating neuronal dynamics, as long as the parameters of the dynamics are tuned just right to reflect the statistics of the problem (Yu & Cohen, 2009). Relatedly, although at first glance, FBM appears to need to keep track of all past observations at all times, and thus may impose an unrealistic demand on an arbi- trary large memory, it can be actually be implemented exactly by keeping track of a

running total of the number of times the target appears in each of the stimulus locations, as these provide the sufficient statistics for the Dirichlet posterior distribution (Bishop, 2006). Thus, the implementation of neither FBM nor DBM requires an explicit representation of probabilities or particularly complex computations, and the issue of computational complexity and neural plausibility is not a particularly pertinent one for dismissing one model in favor of the other, or both altogether.

While not explicitly addressed here, our results do not preclude the possibility that humans can learn about the true nature of the stability, or lack thereof, of statistical regularities in the environment. However, as we showed previously (Yu & Cohen, 2009), it can take even an ideal Bayesian learner a surprising number of trials to exchange a prior bias toward changeable, statistical regular world for a random, stationary world – thus, even if humans are capable of adapting to the rate of statistical change in the world, the length of our experimental session may be insufficient for that type of learning. Future work is needed to clarify the extent to which humans can adapt their internal assumptions about the rate of change of the world in different behavioral contexts.

Our results are also relevant for the study of attention. Our findings demonstrate that overt attention, mediated by purposeful eye movements, complement covert attention to play a critical role in the brain's selection and filtering process. While traditionally attentional selection was thought of arising from limited neuronal resources at perceptual, cognitive, and motor levels (Broadbent, 1958; Deutsch & Deutsch, 1963; Treisman, 1969; Eriksen & St James, 1986), more recently for- mal Bayesian statistical models have suggested covert attentional selection to be computationally desirable beyond any resource limitation considerations (Dayan & Yu, 2002; Dayan & Zemel, 1999; Yu &

Dayan, 2005a; Yu & Cohen, 2009). The current work adds to this "selection-for-computation" principle of attentional selection (Dayan & Zemel, 1999; Yu & Dayan, 2005b; Yu, Dayan, & Cohen, 2009), the concept that attentional processes support the optimization of the inductive process (Helmholtz, 1878) inherent in sensory and perceptual processing (see Yu, 2013 for a longer discussion). Specifically, this work demonstrates that eye movements contribute to sensory processing efficiency by specifically favoring sensing locations in a manner that is sensitive to environmental statistics and task objectives. Understanding the precise manner in which covert and overt attention interact to mediate efficient sensory processing is an important direction of future research.

2.5 Acknowledgements

We thank the assistance from R. Jackson and C. Carper, and J. Schilz in writing the experimental code, J. Schilz for help with running subjects, members of H. Poizner's lab for help in operating the eyetracker, and P. Shenoy for help with comparing per-trial model accuracy. We also thank Peter Dayan for helpful comments on the paper. This work was in part supported by NSF to the Temporal Dynamics of Learning Center, by a UCSD academic senate research grant to A.J.Yu, and by ARL to a CAN CTA consortium including A.J.Yu.

Chapter 2, in full, is a reprint of the material as it appears in Decision 2014. Yu, A. J & Huang, H. Dissertation author was involved in experimental design, coding the experiments in matlab, collecting and analyzing behavior data, and implementing models

under the supervision of the PI (Prof Yu) of this project.

References

Amsterdam, Netherlands. Warren, J. M. (1966). Reversal learning and the formation of learning sets by cats and rhesus monkeys. *Journal of Comparative and Physiological Psychology*, *61*(3), 421-8.

Araujo, C., Kowler, E., & Pavel, M. (2001). Eye movements during visual search: The costs of choosing the optimal path. *Vision research*, *41*(25), 3613-3625.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neurosci*, *10*(9), 121421.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blakely, E., Starin, S., & Poling, A. (1988). Human performance under sequences of fixed-ratio schedules: Effects of ratio size and magnitude of reinforcement. *Psychological Record*, *38*,111-20.

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

Broadbent, D. (1958). *Perception and communication*. Elmsford, New York:Pergamon.

Butko, N. J., & Movellan, J. R. (2010). Infomax control of eye movements.*Autonomous Mental Development, IEEE Transactions on*, *2*(2), 91-107.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876-9.

Dayan, P., & Yu, A. J. (2002). ACh, uncertainty, and cortical inference. In T. G.

Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances In Neural Information Processing Systems 14* (p. 189-196). Cambridge, MA: MIT Press.

Dayan, P., & Zemel, R. S. (1999). Statistical models and sensory attention. *ICANN Proceedings*.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *87*, 272-300.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*, 1366-1378.

Einha¨user, W., Rutishauser, U., & Koch, C. (2008). Task demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 1-19.

Eriksen, C., & St James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225-40.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, *109*(3), 416-422.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650-669.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776), 390-394.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing.

Grene, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proc Natl Acad Sci USA*, *107*, 16401-6.

Hall-Johnson, E., & Poling, A. (1984). Preference in pigeons given a choice between sequences of fixed-ratio schedules: Effects of ratio values and duration of food delivery. *Jounral of the Experimental Analysis of Behavior*, *42*, 127-35.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427-430.

He, P. Y., & 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.

Helmholtz, H. L. F. v. (1878). The facts of perception. In R. Kahl (Ed.), *Selected writings of hermann von helmholtz.* Middletown, CT: Wesleyan University Press, 1971. (Translated from German original Die Tatsachen in der Wahrnehmung.)

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

Herrnstein, R. J. (1961). Relative and absolute strength of responses as a function of

frequency of reinforcement. *Journal of the Experimental Analysis of Behaviour*, *4*, 267-72.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behaviour*, *13*, 243-66.

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *The Journal of Economic Perspectives*, 137-156.

Ide, J. S., Shenoy, P., Yu*, A. J., & Li*, C.-R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *Journal of Neuroscience*, *33*, 2039-2047. (*Yu and Li contributed equally as senior authors)

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

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113*(1), 123-127.

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

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

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature neuroscience*, *3*(12), 1340-1345.

Lee, M. D., Zhang, S., Munro, M., & Steyvers, M. (2011). Psychological models of human and optimal performance in bandit problems. *Cogn Syst Res*, *12*, 164-74.

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

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

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040-1046.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms Tit-for-Tat in the Prisoner's Dilemma game. *Nature*, *364*, 56-8.

Oswal, A., Ogden, M., & Carpenter, R. H. S. (2007). The time course of stimulus

expectation in a saccadic decision task. *J Neurophysiol*, *97*, 2722-30.

Randall, C. K., & Zentall, T. R. (1997). Win-stay/lose-shift and win-shift/lose-stay learning by pigeons in the absence of overt response mediation. *Behavioural Processes*, *41*(3), 227-36.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma*. Ann Arbor, MI: University of Michigan Press.

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

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

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*(2), 628-633.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of neurophysiology*, *86*(4), 1916-1936.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? separating sequential effects in two-choice reaction time. *J. Exp. Psychol.: Human Perception & Performance*, *11*, 598-616.

Steyvers, M., Lee, M. D., & Wagenmakers, E. J. (2009). A bayesian analysis of human decision-making on bandit problems. *J Math Psychol*, *53*, 168-79.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782-7.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, *6*, 263-75.

Talbott, W., Huang, H., & Movellan, J. (2012). Continuous Time Infomax Models of Oculomotor Control. *Journal of Vision*, *12*(9), 1019-1019.

Treisman, A. (1969). Strategies and models of selective attention. *Psychol. Rev.*, *76*, 282-99.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? optimal decisions from very few samples. In *Proceedings of the 31st annual conference*

of the cognitive science society.

Walthew, C., & Gilchrist, I. D. (2006). Target location probability effects in visual search: an effect of sequential dependencies. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1294.

Wilder, M. H., Jones, M., & Mozer, M. C. (2010). Sequential effects reflect parallel learning of multiple environmental regularities. In P. P. J. Zemel R. S. Bartlett & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 24, p. 1791-9). La Jolla, CA: NIPS Foundation.

Yarbus, A. F. (1967). *Eye movements and vision*. New York: Plenum Press.

Yu, A. J. (2013). Bayesian models of attention. In S. Kastner & K. Nobre (Eds.), *Handbook of attention.* Oxford, UK: Oxford University Press.

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

Yu, A. J., & Dayan, P. (2005a). Inference, attention, and decision in a Bayesian neural architecture. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17.* Cambridge, MA: MIT Press.

Yu, A. J., & Dayan, P. (2005b). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681-92.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 700-17.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel–serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 244.

Chapter 3

Sequential effects: A Bayesian analysis of prior bias on reaction time and behavioral choice

Abstract- Human subjects exhibit "sequential effects" in many psychological experiments, in which they respond more rapidly and accurately to a stimulus when it reinforces a local pattern in stimulus history, compared to when it violates such a pattern. This is often the case even if the local pattern arises by chance, such that stimulus history has no real predictive power, and therefore any behavioral adjustment based on these erroneous pre- dictions essentially amounts to superstition. Earlier, we proposed a normative Bayesian learning model, the Dynamic Belief Model (DBM), to demonstrate that such behavior reflects the engagement of mechanisms that identify and adapt to changing patterns in the environment (Yu & Cohen, 2009). In that earlier work, we assumed a monotonic relationship between prior bias and response time (bias toward a stimulus was assumed to result in faster reaction time when that was the actual stimulus; conversely, when the other stimulus was present, it was assumed to result in a slower response). Here, we present a more detailed and quantitative analysis of the relationship between prior bias and behavioral outcome, in terms of response time and choice accuracy. We also present novel behavioral data, along with a framework for jointly

identifying subject-specific parameters of the trial-by-trial learning (Dynamic Belief Model, DBM) and within-trial sensory processing and decision-making (Drift-Diffusion Model, DDM) based on the behavioral data. Our results provide strong evidence for DBM, and reveal potential individual differences, in their differential beliefs about the timescale at which local patterns persist in sequential data.

Motivation

Decisions are influenced by outcomes in the past. It is true when there is a statistical structure in the stimulus (Chapter 2), in which subjects can learn information of the underlying state of the environment from the recent history to facilitate the decision-making process. It also exists when there is no embedded rewarding structure in the stimulus, for example if the stimulus in each trial is randomly ordered (i.e. no special statistical structure in target appearance). For instance, it has been shown that in a 2AFC task with A/B stimulus, subjects responded faster with higher accuracy if the current trial follows a local pattern (e.g. AAAA followed by A or ABAB followed by A), but are slower with higher error rate for the contrary situation (e.g. AAAA followed by B or ABAB followed by B). This sequential effect has been observed and reported from a variety of studies (Soetens, Boer & Hueting 1985; Hogarth & Einhorn 1992; Cho et al. 2002; Jones, Curran, Mozer & Wilder 2013).

In a 2AFC choice task (Choe et al. 2002), a Dynamic Belief Model (DBM) proposed by Yu & Cohen 2009 was used to account for its influence of reaction time and error rate, by assuming subject's belief of the correct choice is a combination of the result

from recent stimulus history and the internal belief of the stability of the environment. They showed that prior belief of a non-stationary environment could produce the observed sequential effects reported in the experiment. This model is also shown to be an approximated linear-exponential filtering of past observations. That is, the prior bias inferred from the model has an approximately linear effect on RT.

On the other hand, Drift Diffusion model (Ratcliff & Smith 2004; Bogacz, Brown, Moehlis, Holmes & Cohen, 2006; Ratcliff & Mckoon 2008) received increasing attention in past decade, for its accountability to explain reaction time and error rate in 2 choice discrimination tasks. Here (Chapter 3), in addition to use DBM for *cross-trial* learning, we propose to further investigate the influence of prior belief on decision-making *within a trial* by using a Drift-Diffusion Model. In particular, we propose to use the prior belief inferred from DBM as the starting point (bias) of the DDM, and subjective sensing difficulty in stimulus information (sensory noise in the stimulus) as the drift rate, thus to provide a more explicit model for both cross-trial learning and within-trial decision-making process.

3.1 Introduction

In a variety of behavioral experiments, human subjects display “sequential effects”, a modulation of response time and/or accuracy by recent trial history (e.g. Bertelson, 1961; Laming, 1968; Kornblum, 1973; Soetens, Boer, & Hueting, 1985; Cho et al., 2002; Jones, Curran, Mozer, & Wilder, 2013). For example, in two-alternative forced choice experiments, in which subjects discriminate between two types of stimuli

(A or B), subjects respond more accurately and rapidly if a trial is consistent with the recent pattern (e.g. AAAAA followed by A, ABABA followed by B), than if it is inconsistent (e.g. AAAAA followed by B, ABABA followed by A). This sequential effect depends on the length of the run (Cho et al., 2002). For instance, an alternation following four repetitions affects responding more than one following only two repetitions. Figure 3.1 illustrates a robust finding of the dependence of RT and error rate on recent trial his- tory, both being largest when a relative long run of repetitions or alternations are broken by the current observation (middle two trial types), and smallest when such runs are extended (left and right end).

Previously, we proposed a Bayesian learning model, the Dynamic Belief Model (DBM), to account for sequential effects, via a human learning mechanism that assumes the potential for discrete, un-signaled changes in the environment. Consequently, DBM repeatedly modifies internal estimates of the relative probability of one stimulus type versus another occurring, based on recent stimulus history (Yu & Cohen, 2009). By assuming reaction time and error rate to be monotonically and inversely correlated with the estimated prior probability of observing the actual stimulus prior to stimulus onset, DBM can qualitatively reproduce the empirically observed sequential effects shown by Cho et al. (2002).

In this work, we give a more precise and quantitative treatment of the influence of prior expectations on sensory processing and decision-making within a trial, by assuming an evidence-integration-to-bound process (Gold, 2002), which is formally similar to the Drift-Diffusion Model (DDM) (e.g. Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) and appears to explain activities of parietal cortical neurons during primate perceptual

decision- making (Roitman & Shadlen, 2014). We present a Bayesian method for simultaneously identifying subject- specific parameters of DBM and DDM based on an individual's choice accuracy and reaction times, and apply it to behavioral data collected in a simple 2-alternative choice perceptual discrimination task. Using this quantitative method, we will compute the relative support, measured in Bayes factors, the data lend to DBM versus a competing model, the Fixed Belief Model (FBM) (Yu & Cohen, 2009), which assumes that human subjects do not believe the task statistics to be changeable over time. We will also characterize the population distributions of subject-specific Bayesian model parameters, which cor- respond to semantically readily interpretable variables, such as subjects' beliefs about the rate of change in the environmental statistics, the overall relative frequency of repetition and alternation trials, and the subjective difficulty (or sensory/perceptual noise) associated with processing the sensory stimulus.

The paper is organized as follows. First, we will describe the experiment and the data collected. Second, we will review DBM and FBM, showing their qualitative differences in trial-wise behavior. Next, we will introduce the quantitative model of prior bias on choice RT and ac- curacy. Subsequently, we will analyze the empirical data using the models. And finally, we will conclude with a discussion of the results, their implications, and directions for future inquiries.

3.2 Experiment

42 college students in UCSD participated in this experiment, all with normal or

corrected vision. In this 2AFC discrimination task, subjects were instructed to decide if the coherent motion of a patch of moving dots on the computer screen was toward left or right by pressing the corresponding arrow keys. Subjects were seated in a chair approximately 60 cm from the computer screen. When a trial started, a patch (diameter: 5 deg visual angel) with coherent moving dots would appear on the screen, moving at a speed of 5 deg/second. The density of the dots was 16.7 dots per sq deg/second, with 3 pixels per dot.

All subjects needed to complete two practice sessions and reach an accuracy of at least 80% to proceed to the main experiment. In the main experiment, there were 7 coherence levels, ranging from 0 to 100%. There were 14 blocks in total, with 120 trials/block, and 2 blocks/coherence. As our first modeling attempt, we only considered data collected from the blocks with 100% coherence. There was no time limit on each trial. Stimuli repetition or alternation was 650 ms. Subjects received feedback at the end of each trial with a beep to indicate if the response was correct. A crisp beep indicated a correct response, while a low-frequency beep indicated an incorrect response. There was a 4-second black screen penalty for premature response (respond within 100 ms after stimuli appears).

Figure 3.1 C;D demonstrate that very similar sequential effects were observed in this experiment as in (Cho et al., 2002).

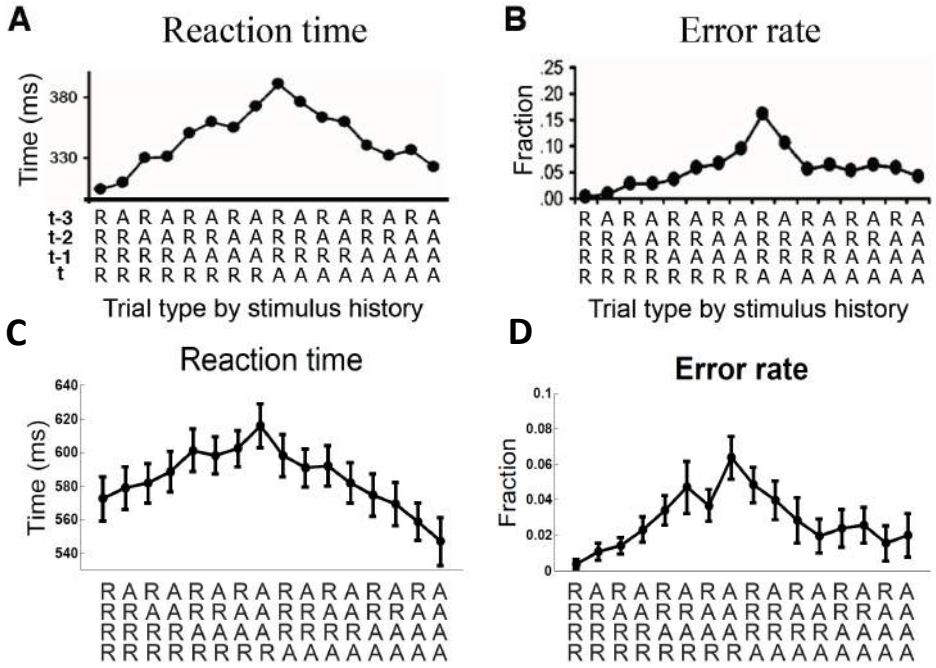

Figure 3.1: Sequential effects in 2AFC tasks manifested in previous and current experiments. (A) Median reaction time (RT) and (B) error rate (ER) averaged across six subjects, adapted from Figure 1 of (Cho et al., 2002). Along the abscissa are all of the possible five-stimulus- long sequences, where R stands for repetition, and A stands for alternation. Each sequence, read from top to bottom, proceeds from the earliest stimulus progressively toward the present stimulus. (C) Median RT and (D) error rate from the current experiment show similar pat- terns; error bars: s.e.m. (Each error bar shows the standard error of subjects' median RT's for the corresponding sequence)

3.3 Learning Models

We give a brief summary of the two Bayes-optimal, ideal observer models in (Yu & Cohen, 2009), which have different assumptions about the temporal persistence of statistical contingencies in the world.

Fixed Belief Model (FBM): Learning about a Fixed World

Suppose the subject has an internal model that on each trial t, there is a fixed probability γ of seeing a repetition ($x_t = 1$), and therefore a probability $1 - \gamma$ of seeing an alternation ($x_t = 0$). $P_0(\gamma)$ is the generic prior to capture the subject's belief about γ before any observations. The prior is modeled as a Beta distribution, Beta (a, b). Over time, the subject can use the number of observed repetitions versus alternations to gain an increasingly precise estimate of the underlying γ. After t observations, the posterior belief of γ is

$$p(\gamma|x_t) \sim p(x_t|\gamma)p_0(\gamma) \qquad (1)$$

which is simply Beta($a + r_t$, $b + t - r_t$), where r_t is the number of repetitions observed up to trial t; x_t is shorthand for the vector of observed sequence (x_1,...,x_t). The probability of observing a repetition on trial t + 1 is the mean of the posterior distribution over γ:

$$p(x_{t+1} = 1|x_t) = \frac{r_t + a}{t + a + b} \qquad (2)$$

Dynamic Belief Model (DBM): Learning about a Changing World

Suppose the subjects believe that the relative frequency of repetition (versus alternation) can undergo discrete changes at unsignaled times during the experimental session (see Figure 3.2B for graphical representation of the model), the subject's implicit task, then, is to track the evolving frequency of repetition versus alternation over the course of the experiment. The crucial assumption is that γ_t has a Markovian dependence on γ_{t-1}, such that there is a large probability α of them being the same, and a small probability $1 - \alpha$ of γ_t being redrawn from the generic prior distribution, p_0. As with the

FBM, the observer would then need to combine the sequentially developed prior belief about stimulus identity, with the incoming stream of sensory inputs, $x_1, x_2, ..., x_t, ...$, to infer the identity of the stimulus in each trial in an iterative fashion:

$$p(\gamma_t = \gamma | \mathbf{x}_{t-1}) = \alpha p(\gamma_{t-1} = \gamma | \mathbf{x}_{t-1}) + (1 - \alpha) p_0(\gamma) \quad (3)$$

and the posterior distribution is:

$$p(\gamma_t = \gamma | \mathbf{x}_t) \propto p(x_t | \gamma_t = \gamma) p(\gamma_t = \gamma | \mathbf{x}_{t-1}) \quad (4)$$

The probability of seeing a new repetition is thus:

$$\begin{aligned} p(x_t = 1 | \mathbf{x}_{t-1}) &= \int p(x_t = 1 | \gamma_t) p(\gamma_t | \mathbf{x}_{t-1}) d\gamma_t \\ &= (1 - \alpha) \langle \gamma \rangle_{p_0(\gamma)} + \alpha \langle \gamma_t | \mathbf{x}_{t-1} \rangle \end{aligned} \quad (5)$$

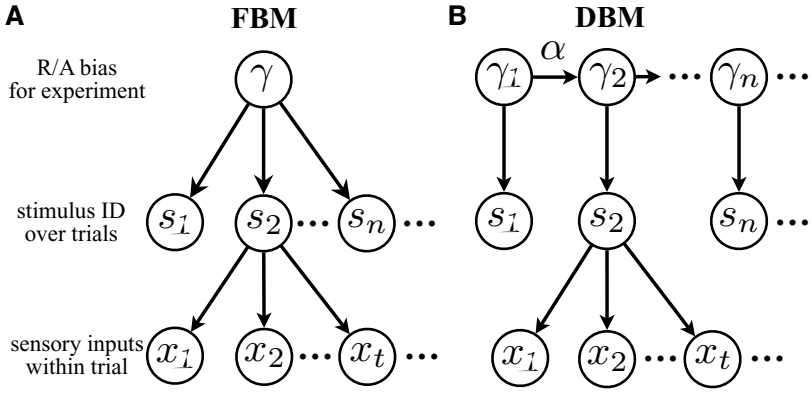

Figure 3.2: Generative models of the fixed and dynamic belief models. (A) Fixed Belief Model (FBM): a hidden bias parameter specifies the frequency of repetitions (and alternations) in the experiment. (B) Dynamic Belief Model (DBM): the hidden bias parameter can change from trial t to trial t+1.

One important consequence of the diminishing uncertainty in the FBM versus the persisting uncertainty in the DBM is that the influence of individual observations persist

indefinitely in FBM, but decreases over time for DBM, with the parameter α determining the effective time window over which individual events exert predictive influence on future events. Figure 3.3 graphically illustrates analyses of the consequences of the different assumptions made by the two models. These simulation results support our hypothesis that the trial-to-trial adjustments seen in subjects' behavior in 2AFC tasks reflect a (perhaps implicit and unconscious) assumption that statistical regularities, such as runs of repetitions or alternations, exist and persist on a characteristic timescale. Such a strategy is useful for a truly volatile environment but inappropriate for the experimental environment, in which stimulus statistics are held fixed.

3.4 Bayesian Model of Prior Belief on Reaction Time and Choice

Previously (Yu & Cohen, 2009), we suggested within-trial perceptual inference and decision-making to be analogous to the sequential hypothesis ratio test (SPRT), and made a loose argument for the prior bias (inferred by DBM) to have an approximately linear effect on mean RT. Separately, we have found that the approximately linear relationship to RT hold for a wide range of α values (data not shown). Here, we take a similar approach but explicitly model the relationship between prior bias and RT. The general idea is to specify a within trial mapping from the belief state (prior probability impending stimulus type), and the stimulus strength/sensory uncertainty, to choice and RT of a given trial.

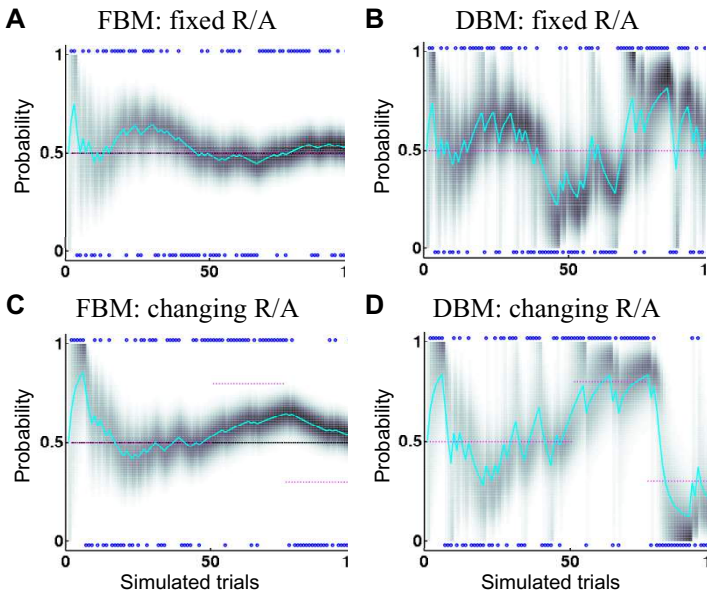

Figure 3.3: Sequential effects transient in FBM, persistent in DBM. (A) illustrates the trial-wise predictive probability of repetition by FBM (cyan), to a sequence of 100 data points, random with = .5 for a repetition. An ideal observer implementing FBM is superior: it becomes immune to the noisy fluctuations in the sequence of observations (the darkening and narrowing band) and converges to the true value. (B) illustrates the trial-wise predictive probability by DBM under the same process. As DBM is (erroneously) applied to learn about a stationary (and random) process, it is strongly and persistently influenced by spurious local patterns. (C) When the underlying environment is truly volatile, FBM cannot easily adapt to new values of, whereas as (D) DBM negotiates these changes adroitly.

Within-Trial Processing

We introduce our method in the context of our experiment, whereby the decision maker needs to decide whether the coherent motion of the dots is toward left or right. For simplicity, and similar to the signal-detection- theory (SDT) formulation (Green & Swets, 1966), we assume each of the two possible stimuli generates normally distributed noisy neural responses at some intermediate stage of the visual pathway, based on which the

subsequent brain region(s) must decide which stimulus was present (and thus which response is required) and when to respond. We assume that the perceived strength and uncertainty of the motion does not depend on its direction, thus the two distributions under hypothesis H_1 (motion toward the left) and H_2 (motion toward the right) have means $-\mu$ and μ, and equal variance, σ^2. The distribution under the true hypothesis is termed the target distribution. Figure 3.4A illustrates SDT.

SPRT solves the problem of deciding between H_1 and H_2, based on an ongoing stream of independent series of sensory signals from the stimulus (the target distribution), $y_1,...,y_t,...$, perceived at discrete steps. The total length (sample size) of sensory signals is also under the observer's control. SPRT says that the observer should keep tracking the relative likelihoods of the two hypotheses being true, and choose the more likely one as soon as the likelihood ratio crosses some decision threshold Z_1 (in which case, stop and decide H_1) or Z_2 (stop and decide H_2). Suppose the prior probability of H_1 being true is p; the probability of the sensory signals up to time t, $\mathbf{y}_t := \{y_1, \ldots, y_t\}$, conditioned on hypothesis H_1 being true, is $f_1(y_t)$, and the probability of the same sequence of sensory signals being generated by hypothesis H_2 is $f_2(y_t)$, then SPRT says to stop as soon as $S_n := \frac{pf_1(y_t)}{(1-p)f_2(\mathbf{y}_t)} \geq Z_1$ or if $S_n \leq Z_2$, and continue otherwise (i.e. if $Z_2 < S_n < Z_1$). Suppose ε is the type I error to be controlled for deciding on either hypothesis, then $Z_1 = \frac{1-\epsilon}{\epsilon}$, and $Z_2 = \frac{\epsilon}{1-\epsilon}$.

It has long been noted that SPRT is formally equivalent to a bounded random-walk model (Laming, 1968; Bogacz et al., 2006). When the observations have

statistically independent noise, we have $f_j(y_1, \ldots, y_t) = \prod_{i=1}^{t} f_j(yi)$ for j = 1,2, and thus

$$I_n := \log S_n = \log \frac{p}{1-p} + \sum_{i=1}^{t} \log \frac{f_1(y_i)}{f_2(y_i)}$$

Notice that f_1 and f_2 are density functions of Gaussian distributions $N(-\mu, \sigma^2)$ and $N(\mu, \sigma^2)$. The increment of information gained from a sensory signal y_i, and its mean and variance are

$$\begin{aligned} \delta I_i &= \log \frac{f_1(y_i)}{f_2(y_i)} = -\frac{2\mu}{\sigma^2} y_i \\ m &:= \mathrm{E}(\delta I_i) = \pm \frac{2\mu^2}{\sigma^2} \\ s^2 &:= \mathrm{Var}(\delta I_i) = \frac{4\mu^2}{\sigma^2} \end{aligned} \tag{6}$$

Since $\log S_n$ is strictly monotonically related to S_n, the decision policy is equivalent to stopping as soon as $I_n \geq Z$ (and choose H_1) or $I_n \leq -Z$ (and choose H_2), for $Z := \log(1 - \varepsilon) - \log \varepsilon$. In other words, the sensory signal accumulation in SPRT is equivalent to a bounded random walk with noisy increments that have a mean drift rate of m per time step (positive if H_1 is true, negative if H_2 is true).

We can then rewrite the total sensory evidence accrued with n steps as

$$I_n = I_0 + \sum_{i=1}^{n} \delta I_i \tag{7}$$

where $I_0 := \log p - \log(1 - p)$ is the starting bias toward H_1. Assuming sensory signals are obtained at small time intervals, we consider the continuous-time limiting

process, I(t), which satisfies the stochastic differential equation

$$dI = mdt + sdW, \quad I(0) = I_0 \qquad (8)$$

where m and s^2 are defined by Equation 6. Equation 8 is a drift diffusion process. The thresholds for I_n, with respect to the thresholds in the original SPRT form, are just ±Z. Figure 3.4 B illustrates DDM.

Cross-Trial Processing

Both DBM and FBM infer the identity of the stimulus (repetition vs. alternation) for each trial, based on the previous observed sequence (Equation 2 and Equation 5). At trial t, the prior probability of seeing a repetition $p(x_t = 1|x_t{-}1)$ can be readily translated to the bias in SPRT: $I_0 = \pm (\log p(x_t = 1|x_t{-}1) - \log(1 - p(x_t = 1|x_t{-}1)))$ which takes positive (negative) sign if H_1 (H_2) was true on the previous trial.

Graphical Model Implementation

Figure 3.4C shows a graphical representation of our join inference of DBM, SDT, and DDM. Unshaded nodes represent model parameters to be inferred from data. Because only the ratio of μ and σ (relative discriminability of the two hypotheses) matters in determining the drift of DDM, we can fix σ at 1. We model the generic prior of the probability of repetition in DBM using Beta(a, b), and denote its mean by e := a/(a + b). To reduce computational complexity, we fix a + b at 2, assuming subjects have (equally) low certainty of the environment before observing any stimulus. We also use the simplest treatment for non-decision time by subtracting the smallest RT for each subject. In fitting

the joint model, we first generated the sequential predictions by DBM, given the true sequence of stimuli observed by each subject, using a grid (.02 increment) of α and e values ranging between 0 and 1.

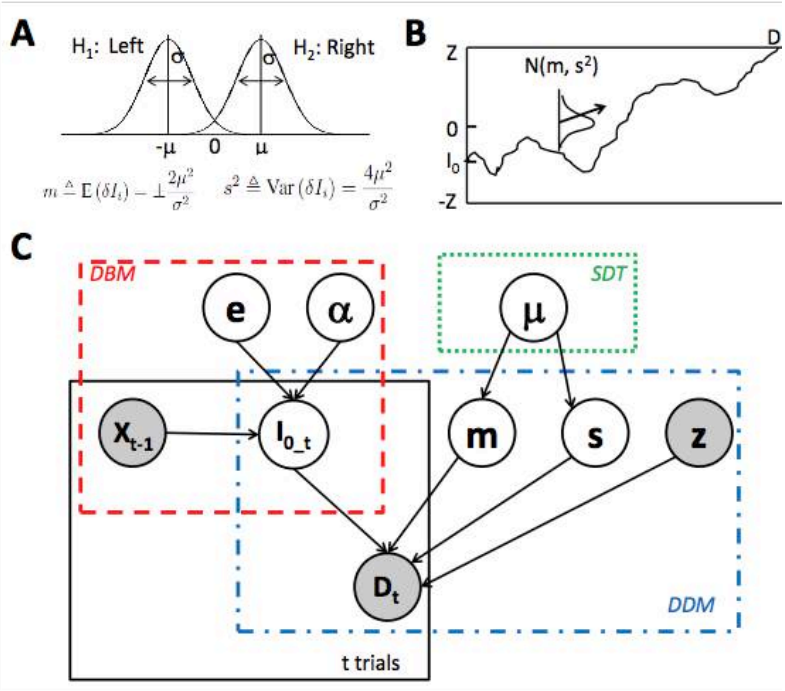

Figure 3.4: Graphical representation of the joint inference of the cross-trial learning (DBM) and within-trial decision making (DDM). Discriminability of the hypothetical distributions in SDT (captured by μ, since we can fix at 1 without loss of generality) determines the rate (m) and variability (s^2) of sensory evidence accumulation in DDM within a trial. The bias of DDM (I_0) is determined by the prior probability of repetition as inferred by DBM from cross-trial learning, conditioned on the parameter associated with a stationary environment and the generic prior belief of the probability of a repetition trial (e). Shaded nodes represent variables that are observable to the experimenter, or can be calculated in a model-free fashion, such as the stimulus sequence (x_t), the observed choices and RT's (D_t), and the log odds of overall accuracy of decision making (Z). The black solid frame indicates t repeated plates of the cross-trial variables. Colored broken frames illustrate different components in the composite model.

We then used MCMC sampling for the graphical model inference using a uniform prior between 0 and 10 for μ, and "discretized uniform" priors for α and e. We fit the

model to each individual subject, and conducted a formal model comparison between DBM and FBM by examining the Bayes factors (e.g. Kass & Raftery, 1995).

3.5 Results

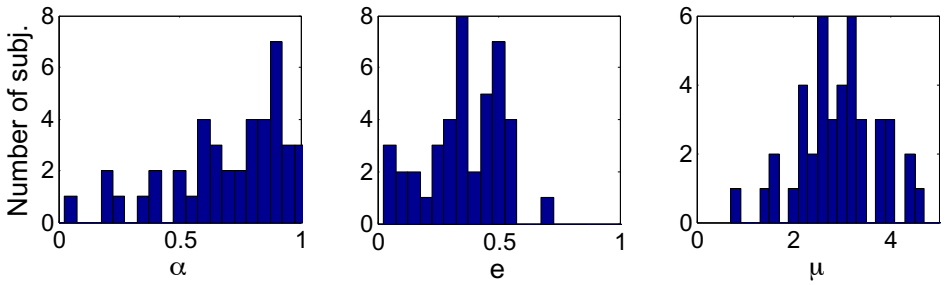

Figure 3.5: Distributions of the MAP estimates of α, e and μ

The top row of Figure 3.5 shows distributions (over all subjects) of the maximum a posterior (MAP) estimates of the DBM parameter, α, the prior belief of the mean probability of repetition, e, and the psychological discriminability of the target direction, μ. The distribution of α has large variation indicating individual differences. The mean of the distribution of e is smaller than .5, implying a bias toward alternation in general. However, we can clearly see that some subjects hold a greater bias toward either repetition or alternation, as indicated by the more extreme estimates of e. There is only one subject who had poor choice accuracy (.74), captured by a low μ value in SDT representation. We did not find any significant correlations between any pair of parameters.

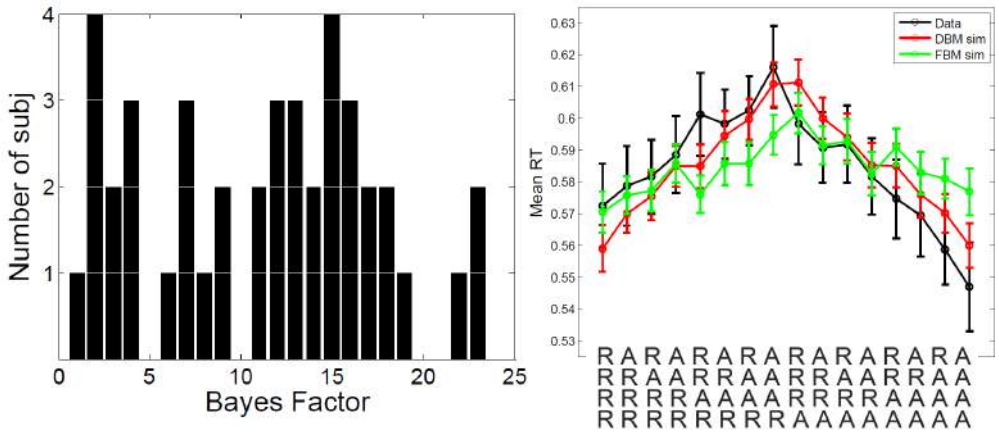

Figure 3.6: Left: distribution of Bayes factors of DBM against FBM, with each brick showing one subject; Right: model predicted RT's compared to data.

We examine DBM and FBM in their abilities of capturing empirical data, using the Bayes factor as a model comparison measure. We calculate the Bayes factor of DBM against FBM, both combined with DDM, for each individual subject. The left panel of Figure 3.6 shows the distribution of Bayes factors over all subjects. Evidence for DBM against FBM is positive for 35 (out of 42) subjects, with a Bayes factor greater than 3 (according to the interpretation scale proposed by Kass and Raftery (1995)). We also compared the RT conditioned on 5 previous trials predicted by DBM and FBM at their best parameterization. The right panel of Figure 3.6 shows that FBM predicts a much smaller sequential effect as com- pared to real-world behavior, whereas DBM can capture a stronger sequential effect as seen in the data.

3.6 Discussion

Previous computational approach to sequential effects focused on Bayes-optimal

learning mechanisms, while giving a simplified treatment to the decision process, such as assuming approximate linear effect of prior on choice RT (Yu & Cohen, 2009; Jones et al., 2013). In this study, we assume an explicit model for within-trial processing, and develop a method for the joint inference of cross-trial learning and within-trial decision making, by augmenting the computational learning model with a principled, sequential hypothesis testing paradigm that is proven to be optimal in both the frequentist and the Bayesian sense (Wald & Wolfowitz, 1948).

Our joint inference and model comparison results sup- port DBM as a better account of human sequential learning than FBM. On the other hand, our results also pro- vide strong evidence for individual differences in their belief of the rate of changing of the environment. The distribution of the inferred α values across all subjects has a large variation that implies potential individual differences.

Our model builds prior knowledge in the starting point of SPRT (and its continuous-time limit, DDM). One of the main theoretical points of proponents of alternative race/accumulator type of models is that the starting point confounds prior knowledge with decision utilities (values). We do not manipulate utilities in the current study, yet future work involves an analysis of how the model would handle decision utilities.

Another future direction is to consider a joint inference of computational and neural models, by extending a newly developed, statistical approach of combining neural and behavioral measures to study cognition (Turner, in press). By jointly fit the computational and neural models, it would become feasible to make simultaneous inference about the correlation between parameters at these different levels.

3.7 Acknowledgments

Funding was in part provided by NSF CRCNS BCS- 1309346 to A. J. Yu.

Chapter 3, in full, is a reprint of the material as it appears in Proceedings of the Cognitive science Society Conference 2014. Zhang S, Huang H and Yu A. Dissertation author was involved in coding the experiment in matlab, collecting and analyzing behavioral data in this project.

References

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 13, 90–102.

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

Cho, R., Nystrom, L., Brown, E., Jones, A., Braver, T., Holmes, P., & Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective and Behavioral Neuroscience*, 2, 283–299.

Gold, J. I. (2002). Good vibrations. Neuron, 33, 842–4.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1), 1-55.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H.(2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological Review*, 120, 628–666.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 377–395.

Kornblum, S. (1973). Sequential effects in choice reaction time. a tutorial review. (S. Kornblum, Ed.). New York: Academic Press.

Laming, D. R. J. (1968). Information theory of choice reaction time. London: Academic Press.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, 111(2), 333.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation, 20(4), 873-922.*

Roitman, J. D., & Shadlen, M. N. (2014). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, 22, 9475–9489.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 598–616.

Turner, B. M. (in press). Constraining cognitive abstractions through bayesian modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), An introduction to model-based cognitive neuroscience. New York: Springer.

Wald, A., & Wolfowitz, J. (1948). Optimal character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19, 326–339.

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? In *Advances in neural information processing systems* (Vol. 21, pp. 1873–1880). Cambridge, MA: MIT Press.

Chapter 4

Infomax models of oculomotor control

Abstract- From a Bayesian point of view, learning is simply the process of making inferences about the world based on incoming data. The efficiency of this learning is determined by the quality of the information provided by the sensors. Thus, a critical part of learning is the existence of a sensory-motor system designed to maximize the information required to achieve goals. Here we show that a wide range of primate eye movement phenomena can be elegantly explained from the point of view of infomax control. The proposed approach describes the velocity profiles of saccadic eye movements as well as previously existing models. In addition, the infomax approach explains phenomena that are beyond the scope of previous models: non-saccadic eye movements, and the difference in end point and velocity profiles observed in saccade- to-target and reach-to-target tasks. The results suggest that the occulomotor control system evolved to be a very efficient real time learning machine.

4.1 Introduction

From a Bayesian point of view, learning is simply the process of making inferences about the world based on incoming data. The efficiency of this learning is determined by the ability of the sensory-motor control system to maximize the

information needed to achieve goals (infomax control).

Humans make over 150,000 saccades per day, spending about 2 hours in saccadic flight, during which useful vision is very poor. It is well known that the velocity profiles of primate saccadic eye movements are quite stereotyped and adhere to consistent relationships between amplitude, duration, and peak velocity. These relationships have been called the "main sequence" (Harris & Wolpert 2006).

Recent models of oculomotor control have been successful at describing saccade velocity profiles using optimal control principles. Typically, these models focus on the relationship between motor commands and forces applied to the eyes, and postulate that the goal of the oculomotor system is to drive the eye to target locations as quickly and accurately as possible. Some models postulate that the eyes minimize the expected deviation from a target end point (Harris & Wolpert 1998; Van Beers 2008). Other models postulate that eye movements minimize the time required to reach the target point (Tanaka et al. 2006), which turns out to be mathematically equivalent. These models ignore the sensory properties of the eyes and assume that the goal of oculomotor control is to reach target points. How these targets are selected is beyond the scope of the models. A recent class of models has focused on explaining how the target points are selected using information maximization principles (Najemnik & Geisler 2005; Butko & Movellan 2010). Up to now these models have focused on the sensory properties of the eyes (e. g. , the fall-off of sensitivity as a function of eccentricity) and have ignored their mechanical properties. Here, we show that by jointly examining the sensory and mechanical properties of the eyes it is possible to explain a range of new phenomena that were beyond the scope of the previous models. The approach shows that a wide range of

primate eye movement phenomena reveal that the primate oculomotor system evolved to be a very efficient real-time learning machine.

4.2 Infomax Model

To model oculomotor control, we first need to have a description of the system of the eye. We follow Harris & Wolpert 1998 in using a state-space model with signal dependent noise to describe the eye. We call the state of the eye at time t as X_t, and describe its changes through time with

$$dX_t = \underbrace{AX_t dt}_{\textbf{drift}} + \underbrace{BU_t dt}_{\textbf{control}} + \underbrace{(C + U_t)dB_t}_{\textbf{noise}} \quad (1)$$

$$X_t = \begin{bmatrix} X_{e,t} \\ \dot{X}_{e,t} \end{bmatrix} \quad (2)$$

Where the matrix A represents the passive dynamics of the system, B describes the effects of the control inputs U_t on the state, and C describes the effect of the Brownian motion B_t on the state. X_t is the state matrix, which contains the eye position $X_{e,t}$ and velocity $\dot{X}_{e,t}$. Notice that the noise scales with the size of the control input, which gives rise to a tradeoff between controlling the system and being certain of the system's state. The values in the A and B matrices were found from human saccades in Robinson 1973, and like Harris & Wolpert 1998 we use these values. These values were retrieved from horizontal saccades, and we model saccades similarly in only one dimension.

Although previous models have assumed that the endpoint of a saccade, which we

will call Z, is known exactly, here we introduce target uncertainty by treating Z as a random variable. Especially if this target is presented in the periphery, subjects will be unsure of the target's location due to sensory uncertainty. Here we model the belief of the target's location as a Gaussian distribution. Additionally, we assume target has its own dynamics described by

$$dZ_t = \eta_t dt + n dV_t \tag{3}$$

where η_t is the model of the target's velocity, and dVt is Brownian motion, with magnitude determined by n. We assume the model of the target dynamics is known. The model could potentially be learned or estimated from the observations of the target, but we do not address this issue here. Even though the dynamics are known, Z is not, so we need a model for how the eye learns about the location of the target.

Similar to Erez et al. 2011, who use a POMDP framework to examine hand-eye coordination, we model the observations Y that the eye collects about the target. These observations change through time as

$$dY_t = (X_{e,t} - Z_t) g(X_{e,t}, Z_t, \dot{Z}_t) dt + dW_t \tag{4}$$

If the observations were noiseless and accurate, they would give the eccentricity of the target with respect to the location of the eye. However, the observations are contaminated by noise, dWt, and the signal-to-noise ratio (SNR) is defined by the visual acuity function g. We choose the following form of the visual acuity function:

$$g(X_t, Z_t, \dot{Z}_t) = e^{-\left[\underbrace{\frac{1}{\rho}((X_{e,t} - Z_t)\beta)^{\rho}}_{\textbf{eccentricity}} + \underbrace{\frac{1}{2}(\dot{X}_{e,t} - \dot{Z}_t)^2\gamma}_{\textbf{velocity}}\right]} \quad (5)$$

where ρ and β define the shape and width of the falloff in SNR due to the target's eccentricity, and r defines the width of the falloff in SNR due to the relative velocity of the eye. For computational simplicity, we restrict ρ to be an even number so we can avoid using an absolute value on $(X_{e,t} - Z_t)$. For example, if we assume the velocity term is zero, and ρ = 4, Figure 4.1 shows a schematic of how the SNR would decrease as the eccentricity (on the x-axis) diverges from zero in either direction.

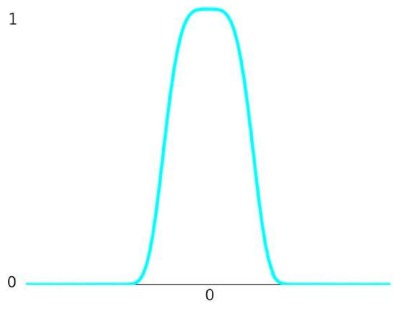

Figure 4.1 Schematic figure of how the SNR decreases as the eccentricity (x-axis) differs increasingly from zero.

A similar shape would hold for the velocity term of the visual acuity function as well. The velocity term models the cost of moving the eyes rapidly. In humans, saccadic suppression masks high-frequency visual information during fast eye movements. Similarly, in the model, a high velocity of the eye with respect to the target reduces the SNR of the observations. This sets up another tradeoff. If the eye is far from a stationary

target, the controller must decide whether it is better to make slow movements that generate more reliable observations, or fast, unreliable movements to decrease the eccentricity of the target. The optimal action will depend on the values of the parameters ρ, β, and r, and the relative cost of action and uncertainty. Taken together, (1), (3), (4), and (5) describe the system to be controlled. What remains is to find a control policy that can generate the actions Ut from times o to final time horizon T such that the system is driven to a desired state.

Learning the control policy

To find the optimal policy, we first need to define an objective function. Here we use a quadratic objective function, and use iterative Linear Quadratic Gaussian (iLQG) control (Todorov & Li, 2005), which will require dealing with both the partial observability and the non-linearity from Najemnik & Geisler 2005. In this paper, we will model eye movements in three tasks (target-directed saccades, smooth pursuit, and eye-hand coordination). Objective function in Erez et al. 2011 will be modified based on task goals.

For the moment, we will focus on target-directed saccades. In this task, we use data collected from horizontal saccades in humans (Collewijn et al. 1988). The participants were asked to saccade from a central fixation point to a flashed target at different amplitudes. Because the task involves positioning the eyes as close as possible to the target, we start with a term for minimizing the squared error between eye and target. Additionally, we include a term to model the cost of action. Let the cost function take the form

$$(X - Z)^2 + U^2 Q \tag{6}$$

where Q is a scalar that captures the tradeoff in cost related to being far from the target point and making actions.

The expression in (6) will be the cost function if the target location Z is known. However, because the location Z is unknown, we cannot use this cost directly. Instead, we need to take the expected value of (6), which leads to

$$(X - \hat{Z})^2 + \sigma_Z + U^2 Q \tag{7}$$

where E[Zl = Z and (Jz is the variance of the estimate of the target location. With (7), we have a quadratic cost function. Notice that even in a situation where the task is to move to a specified location, there is still pressure to find a solution that maximizes the information about the target location.

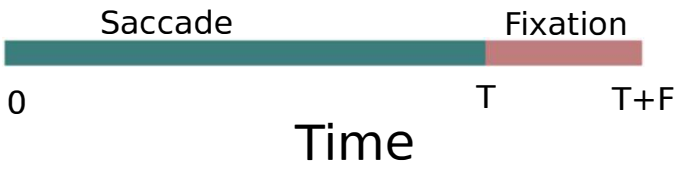

Figure 4.2 The two segments of the finite horizon used for control. First, from time 0 to time T, there is no penalty for the distance from the eye to the target. Second, during fixation (from time T to time T + F), the penalty on the eye position is enforced. The penalty on the magnitude of the action U is enforced for the entire horizon.

In this target-directed task, the entire eye movement in one trial includes first a saccade and then a short fixation period at the target. To apply the cost function (7) to the entire movement, we also separate the movement control into these two segments, as shown in Figure 4.2. The first segment, which we will call the saccade, only contains

penalties on the actions. The second segment, from time T to time T + F, which we call the fixation, also includes the penalty on the state of the eye. With this, the complete minimization objective becomes

$$\int_{T}^{T+F} ((X_t - \hat{Z})^2 + \sigma_Z)dt + \int_{0}^{T+F} U_t^2 Q dt \qquad (8)$$

Following [9], we include $\hat{Z}$ and σ_Z in the state X, and plan according to the belief state. Using an extended Kalman-Bucy filter, we can find the dynamics of $\hat{Z}$ and σ_Z. The extended Kalman-Bucy filter equations are as follows:

$$d\hat{Z}_t = \eta dt + K_t dI_t \qquad (9)$$

$$dI_t = (dY_t - f(X_t, \hat{Z}_t, \dot{\hat{Z}}_t))dt \qquad (10)$$

$$K_t = \sigma_{Z,t} \frac{\partial}{\partial \hat{Z}_t} f(X_t, \hat{Z}_t, \dot{\hat{Z}}_t) \qquad (11)$$

$$d\sigma_{Z,t} = -K_t^2 dt + n_t^2 dt \qquad (12)$$

where we've defined

$$f(X_t, \hat{Z}_t, \dot{\hat{Z}}_t) = (X_{e,t} - \hat{Z}_t) g(X_t, \hat{Z}_t, \dot{\hat{Z}}_t) \qquad (13)$$

and g is as in (5). Using the product rule, we can find

$$\frac{\partial}{\partial \hat{Z}_t} f(X_t, \hat{Z}_t, \dot{\hat{Z}}_t) = g(X_t, \hat{Z}_t, \dot{\hat{Z}}_t)(1 - \beta(X_{e,t} - \hat{Z}_t)^\rho) \qquad (14)$$

Using the above equations, we can incorporate the observation process Y with the estimate of the target location and give the dynamics of $\hat{Z}$ and σ_Z in relation to time.

The final step is to augment the A and B matrices from (1) to include the

dynamics of $\hat{Z}$ and σ_Z in relation to changes in the other state variables. We can find the necessary terms in the augmented matrices by taking the partial derivatives of (12) with respect to $X_{e,t}$ and velocity $\dot{X}_{e,t}$, $\hat{Z}$ and σ_Z. This will allow us to linearize the dynamics of the system around a given state or sequence of states.

With the linearized dynamics, the belief state Z, and the quadratic cost function, we can now solve for a control policy using iLQG. The policy learned by iLQG is a closed-loop policy. This means the optimal action at a given time can depend on the state rather than just the time; in other words, the optimal policy can react to changes in the environment. This feature of the policy is interesting, and differs from previous models, and its implications are discussed in more detail in Section V. Once the control policy has been learned, it can be applied to a noise-free simulation to give the expected trajectory of the eyes for a set of parameters.

4.3 Evaluation Methods

To evaluate the suitability of the infomax model for describing oculomotor control, we looked at three different eye movement paradigms. The first is in describing the velocity profiles of horizontal saccades, as was described earlier. Second, we examined the qualitative suitability of our model for predicting smooth pursuit in amenable situations in simulation. We also examined a task where the eye played a supportive role to the hand, which had to reach a target.

Saccades

To learn the parameters of the system that describes the eye's movement and observations, we perform a pattern search to minimize the root squared error between the velocity profiles of 5, 10, 20, 30, 40, and 50 degree saccades generated by the optimal controller under the fixed set of parameters and the behavioral data from [10]. We allow the saccade duration T to change with each amplitude, but all other parameters are held constant across amplitudes.

To compare the infomax model to other models, we use a cross-validation paradigm, where each amplitude saccade is held out in turn. The parameters for each model are set from the remaining amplitudes, and an optimal movement for the held-out amplitude is generated with the learned parameters. Then, the error is calculated, and averaged across the amplitudes.

Smooth pursuit

To model eye movements other than saccade, we need only make minimal changes to the objective function (8). Rather than considering the task where the eyes are required to move to a particular location (as was the case in our model of the saccade task from Collewijn et al. 1988), here we only consider the goal of minimizing the variance of the estimate of the target location. As such, the objective comes closer to pure information maximization, and is defined as

$$\int_{T}^{T+F} \sigma_Z dt + \int_{0}^{T+F} U_t^2 Q dt. \qquad (15)$$

Although we have tried the following experiments with an objective function

closer to (8) with similar results, it is more compelling to show that even without a strict penalty on the location of the eyes, we can generate qualitatively similar behavior to smooth pursuit, so we will focus on this case.

Hand-eye coordination

We model eye movements in a rapid reaching task (data collected and described by Huang et al. 2012) in two conditions: Eye+hand, in which the reward is given based on the endpoint of hand movement; Eyes only, in which the reward is given based on the endpoint of saccadic eye movement. In the experiment, subjects were instructed to reach the target at a distance of 20 cm (25 degree visual angle) from the starting location either with hand (Eyes+hand) or eye (Eyes only) movement at a time window of 600 ms. For the former, subjects can freely move their eyes and thus eye movements only serve to guide hand movements. For the latter, subjects' reward will be based on the endpoints of eye movements and thus eye movements contribute directly to the task goal. We used Eyelinkl000 system to track eye movement and Phasespace motion capture system to record hand movement in the experiment.

We model the hand as a point mass, and used the dynamics for the hand as described in Tanaka et al. 2006.

For Eyes+hand condition, without constraint on eye movement to the target, the objective function for eye movement is

$$\int_{T}^{T+F} ((X_{h,t} - \hat{Z})^2 + \sigma_Z)dt + \int_{0}^{T+F} U_t^2 Q dt \qquad (16)$$

where $X_{h,t}$ is the position of the hand at time t. For Eyes only condition, with the task goal of moving eyes to the target, similar as the target-directed saccade task, the objective function for eye movement is (8).

4.4 Results

Predictions of optimal saccades for static targets

Figure 4.3 compares the infomax model prediction of saccade velocity profiles over a range of amplitudes. Optimal velocity profiles (Fig 4.3b) captured the important shape features shown in behavioral data (Fig 4.3a) (i. e. symmetric for low amplitudes and asymmetric/left-skewed peak for high amplitudes). The optimal positions (Fig. 4.3c) also show similar trajectories as in the observed behavior.

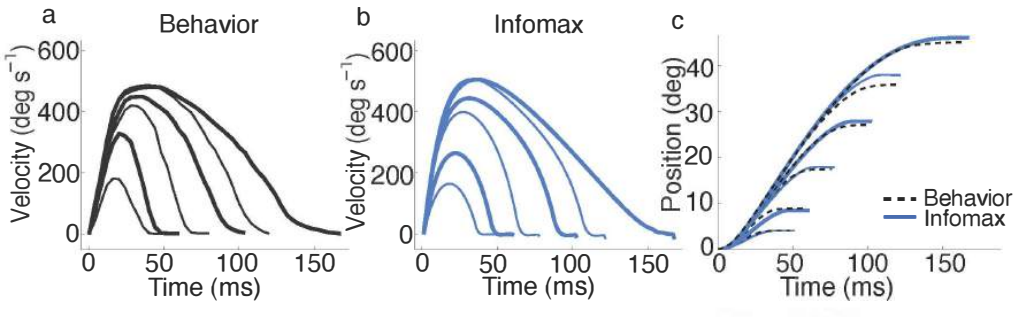

Figure 4.3: Comparison of behavioral result and infomax predictions. a. Observed velocity profiles of horizontal saccades when the target is at 5°, 10°, 20°, 30°, 40° and 50° ([10]). b. Optimal saccadic velocity profiles for corresponding amplitudes shown in a. c. Optimal eye positions (solid blue line) and observed eye positions (dashed black line) for the amplitudes shown in a.

In Figure 4.4, we compare infomax with previous models (Internal Model from Chen-Harris et al. 2008; Minimum Variance Model from Van Beers 2008). RSE

comparison (Fig 4.4c) suggests there is no significant difference (summed over all velocity profiles between behavior and model predictions) between those 3 models ($p>0.1$).

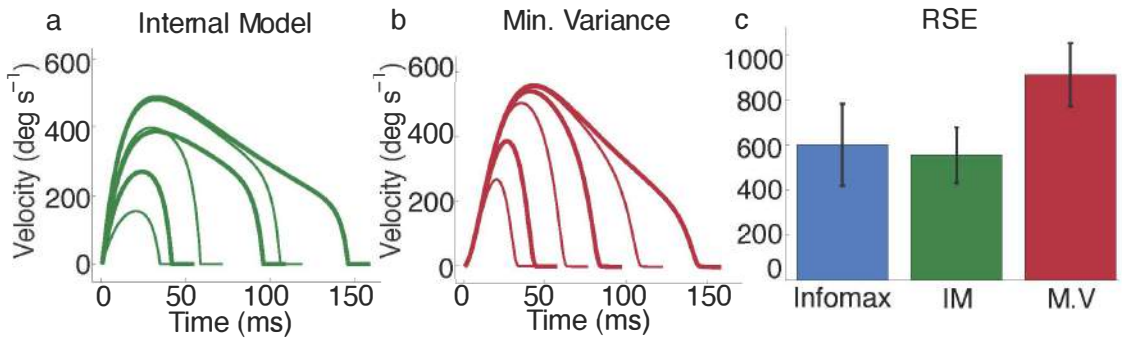

Figure 4.4: Comparison of infomax and previous models. a. Internal Model from Chen-Harris et al. 2008. b. Minimum Variance Model from Van Beers 2008. c. Mean RSE over all the amplitudes, error bars show the standard error of the mean.

Prediction of saccadic and smooth pursuit eye movement for moving targets

Figure 4.5a and 4.5b show infomax prediction for eye movements when the target moves at 20 deg/s with no location difference between initial fixation and the onset of the moving target. Eye velocity trace in Fig 4.5a suggests the eye will increase velocity rapidly and continuously until reaches the target velocity (~ 40 ms after target onset), and then track the target using pursuit eye movements. Eye position trace in Fig 4.5b shows eye positions closely match target locations.

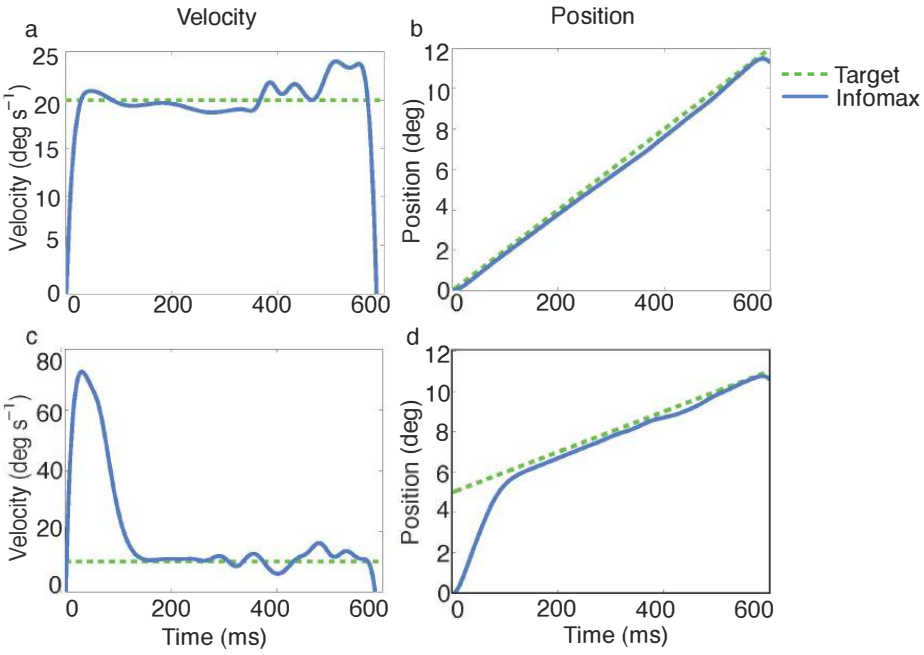

Figure 4.5: Representative eye movement responses to moving targets. Top row: a. Eye velocity trace (solid blue line) and b. corresponding eye position and target position in response to a zero offset target moving rightward at 20 deg/s (dashed green line). Bottom row: c. Eye velocity trace (solid blue line) and corresponding eye position and target position in response to a moving target initially located at 5 deg in the right visual field and then moves rightward at 10 deg/s (dashed green line).

Figure 4.5c and 4.5d shows model prediction of eye movement when the moving target is initially located 5 deg to the right of fixation, and then moves to the right at 10 deg/s. Eye velocity trace in Fig 4.5c shows the eye will first make a quick catch- up saccade-like movement to the target (~150 ms after target onset) and then track the moving target at a matching speed.

The eye position trace in Fig 4.5d suggests the eye will reach the target at the end of the first quick movement and then track the target position.

Prediction of eye movement in Hand-eye coordination

Figure 4.6a shows model predictions of eye movement in hand-eye combination (solid blue) and eyes only (solid red) conditions. Comparing with behavioral data (dashed blue, dashed red lines) observed in the experiment (Figure 4.6b), eye movement endpoints in the task show that subjects undershoot target with saccadic eye movements when the task goal is to reach the target with the hand (top panel in Figure 4.6b). However, the undershooting disappeared when the task goal was to fixate the target with eye movements (bottom panel in Figure 4.6b). Infomax predictions (Figure 4.6a) of optimal eye positions for the hand-eye condition (solid blue) and eye only (solid red) are consistent with this observation from the behavioral data (dashed blue and red).

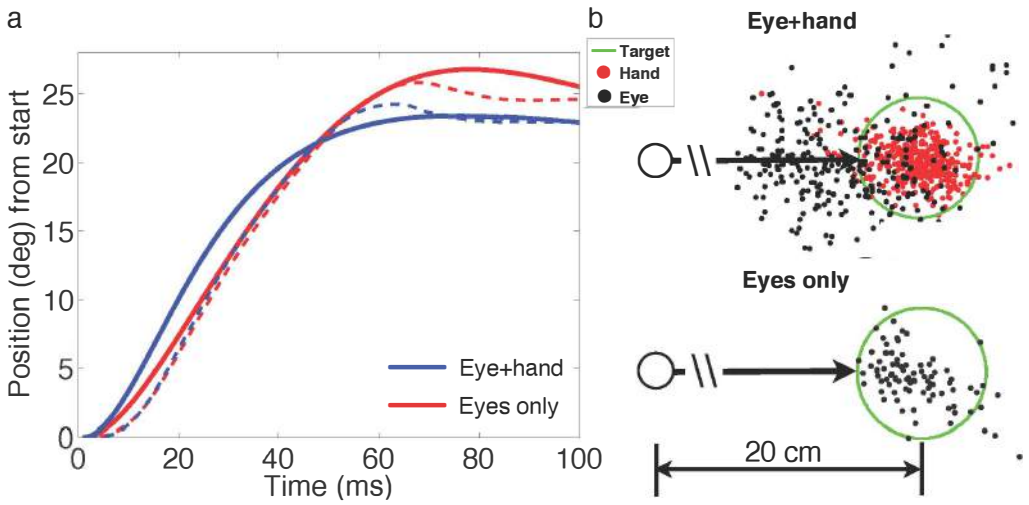

Figure 4.6: Eye movements in the reaching task. A. Comparison between informax prediction of optimal eye positions (solid lines) and behavioral data (dashed lines). b. Eye movement endpoints in Eye+hand condition (top panel) and in Eyes only condition (bottom panel). Green circle is the target; red dots are hand endpoints in Eye+hand condition; black dots are eye endpoints.

4.5 Discussion

We showed that saccadic eye movements emerge as the solution to an information maximization problem with sensors that have a limited field of view and limited temporal resolution. The information maximization principle explains the velocity profiles observed in saccadic eye movements as well as previous principles, including minimum end-point variance (Harris & Wolpert 1998) and minimum time (Tanaka et al. 2006). More importantly, information maximization explains eye movement phenomena that were beyond the scope of previous models. We showed that, for moving targets, infomax generates both saccades and smooth pursuit eye movements. When target onset location is close to the initial eye fixation (foveal), infomax predicts smooth pursuit eye movement which closely tracks target positions. When the target appears in a peripheral location, infomax predicts a catch-up saccade followed up by smooth pursuit eye movement. Qualitatively, this behavior was observed in empirical studies (Robinson et al. 1986; Erkelens 2006). While previous models (Grossberg et al. 2012) explained smooth pursuit from the point of view of minimizing tracking errors, here we explain both saccadic movements and smooth pursuit from the point of view of maximizing information about the location of a target.

We designed an experiment in which target tracking and information maximization make different predictions. Subjects were instructed to reach a target with their hands (Hand condition) or with their eyes (Eye condition). Subjects were rewarded based on the endpoints of hand movements or eye movements, respectively. For the Eye condition, subjects made eye movements as predicted both by the infomax approach and

by the target tracking approach. However, for the Hand condition, eye movements undershot the target by -2.5 degrees. This result was predicted by the information maximization approach but contradicted the target tracking models. According to the infomax model, the reason why people undershoot in the Hand condition but not in the Eye condition is that moving the eyes close to the target does not improve the accuracy of the hand motion.

It should be noted that the infomax model generates closed- loop control policy for the eyes. At first glance, this might seem an undesirable feature since saccades are widely believed to be open-loop, ballistic movements. However, due to the limited temporal bandwidth of our eye model, when the eyes move quickly they provide very little visual information, and thus virtually operate in open loop mode. The decision to move slowly in closed loop mode, or quickly in open loop mode, can be seen as the result of optimizing a common information maximization principle.

4.6 Conclusions

From a Bayesian point of view, learning is equivalent to making inferences based on the information gathered by the sensors. Efficient learners are thus those that control their sensors so as to maximize the expected value of information. Here, we showed that a wide range of properties of the oculomotor system, including the velocity profiles of saccades, the transition between smooth pursuit and saccadic movements, and eye hand coordination in reaching tasks can be explained from the point of view of information maximization. In summary we showed that, by considering that the occulomotor system

has evolved to be a very efficient real-time learning machine, one can make sense of a wide range of phenomena that were previously addressed using different principles or that were beyond the scope of previous models.

It should be noted that our work is agnostic with respect to brain implementation issues. For example, while we show that saccadic movements and smooth pursuit movements serve a common goal (information maximization) it is perfectly plausible for the two forms of movements be controlled by different brain systems. It is also possible that, as recently suggested (Erkelens 2006; de Xivry & Lefevre 2007) saccades and pursuit are two outcomes of a single sensorimotor system. Regardless, our work suggests that the brain systems involved in oculomotor control have evolved to serve a common computational principle: efficient, real-time learning.

4.7 Acknowledgements

This research was supported by NSF lIS 0968573 and NSF lIS 0808767.

Chapter 4, in full, is a reprint of the material as it appears in Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on. Talbot W, Huang H and Movellan J.

References

Butko, N. J., & Movellan, J. R. (2010). Infomax control of eye movements. *Autonomous Mental Development, IEEE Transactions on*, *2*(2), 91-107.

Chen-Harris, H., Joiner, W. M., Ethier, V., Zee, D. S., & Shadmehr, R. (2008). Adaptive control of saccades via internal feedback. *The Journal of Neuroscience*, *28*(11), 2804-2813.

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *The Journal of Physiology*,*404*(1), 157-182.

de Xivry, J. J. O., & Lefevre, P. (2007). Saccades and pursuit: two outcomes of a single sensorimotor process. *The Journal of Physiology*, *584*(1), 11-23.

Erez, T., Tramper, J. J., Smart, W. D., & Gielen, S. C. (2011, April). A POMDP Model of Eye-Hand Coordination. In *AAAI*.

Erkelens, C. J. (2006). Coordination of smooth pursuit and saccades. *Vision research*, *46*(1), 163-170.

Grossberg, S., Srihasam, K., & Bullock, D. (2012). Neural dynamics of saccadic and smooth pursuit eye movement coordination during visual tracking of unpredictably moving targets. *Neural Networks*, *27*, 1-20.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780-784.

Harris, C. M., & Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biological cybernetics*, *95*(1), 21-29.

Huang, H., Plank, M., Gepshtein, S., & Poizner, H. (2012). Target predictability and eye-hand coordination in a rapid reaching task. *Journal of Vision*, *12*(9), 411-411.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387-391.

Robinson, D. A. (1973). Models of the saccadic eye movement control system.*Kybernetik*, *14*(2), 71-83.

Robinson, D. A., Gordon, J. L., & Gordon, S. E. (1986). A model of the smooth pursuit eye movement system. *Biological cybernetics*, *55*(1), 43-57.

Tanaka, H., Krakauer, J. W., & Qian, N. (2006). An optimization principle for determining movement duration. *Journal of neurophysiology*, *95*(6), 3875-3886.

Todorov, E., & Li, W. (2005, June). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. In*American Control Conference, 2005. Proceedings of the 2005* (pp. 300-306). IEEE.

Van Beers, R. J. (2008). Saccadic eye movements minimize the consequences of motor noise. *PLoS One*, *3*(4), e2070.

Chapter 5

The Influence of Depression on Cognitive Control: Disambiguating Approach and Avoidance Tendencies

Abstract - It has been difficult to disambiguate the degree to which approach and avoidance motivation dysfunctions contribute to performance in goal-directed tasks in depression. Here, we propose to use a novel experimental paradigm, i.e. a computer simulated driving-task, to study the impact of depressed mood on cognitive control by measuring approach and avoidance actions in continuous time. In this task, 39 subjects with minimal to severe depression symptoms used a joystick to move a virtual car as quickly as possible to a target point without crossing a stop-sign or crashing into a wall. We recorded their continuous actions on a joystick and found that depressed mood 1) affects stopping position; and 2) biases the magnitude of late deceleration (avoidance) but not early acceleration (approach). Taken together, these results suggest that depressed mood promotes stronger active avoidance near stopping target, while minimally impacting approach motivation.

5.1 Introduction

Depression has been linked to both motivational deficits and impaired motor control, which is evidenced in different complex goal-directed tasks (Caligiuri & Ellwanger 2000) and increased odds of car accidents (Selzer, Rogers & Kern, 1968; Bulmash et al. 2006). Despite this compelling evidence of poor cognitive control in depressed individuals, the effect of depressed mood on the implementation of driving actions remains largely unknown. Moreover, due to the complex nature of depressive symptoms and related neural systems (Drevets, 2001; Liotti & Mayberg, 2001), evidence of cognitive control deficits in depression based on standard cued response tasks is at best mixed suggesting a need for more sensitive behavioral paradigms to assess cognitive control in depression.

Motivational Deficits in Approach-Avoidance Task

Anhedonia (the decreased capability to experience pleasure and seek reward) is an essential symptom of Major Depressive Disorder (MDD; DSM-V; American Psychiatric Association, 2013) and is a core feature of reward-processing deficits. Depressed individuals report lower positive affect in response to positive stimuli (Dunn, Dalgleish, Lawrence, Cusack, & Ogilvie, 2004) and are slower to approach positive social cues (Vrijsen, Oostrom, Speckens, Becker & Rinck, 2013). In addition, they are less discriminative of and sensitive to monetary rewards to perform more accurately on cognitive tasks (Henriques & Davidson, 2000). These reward-seeking deficits appear in the context of decreased recruitment of the nucleus accumbens and prefrontal cortex

(PFC) upon exposure to rewards, which is believed to reflect impairment in attentional switching, regulatory processes (Heller et al., 2009; Epstein et al 2006) and in flexibility to alter reward-seeking behavior (Disner, Beevers, Haigh & Beck, 2011).

While motivational research often focuses on approach deficits in depression, a disturbance in avoidance processes may be equally important and relevant to motor control performance among depressed individuals (Trew, 2011). For example, depressive individuals learn faster to avoid risky gambles (Smoski et al., 2008), and demonstrate faster motor response to withdraw from negative stimuli such as negative faces (Seidel et al., 2010). Depression is associated with a greater tendency to move away from undesired states (avoidance goals; Dickson & MacLeod, 2004) and more reported avoidant schemas and emotions (Aldao, Nolen-Hoeksema & Schweizer, 2010). Overall, these findings suggest depressed individuals may have attenuated motivation to approach reward as well as greater sensitivity to and higher motivation to avoid punishment. However, few studies (Layne, Merry, Christian & Ginn, 1982) have attempted to quantify these motivational biases in depressed individuals, particularly in terms of their influence on motor-control in complex goal-directed tasks.

The Present Study

To gain a more precise understanding of how motivational dysfunction in depression may impact motor control, we designed an experiment allowing us to disambiguate approach and avoidance motivational tendencies underlying motor-control behavior in continuous time. Specifically, we designed a computer simulated driving task including both approach (e.g., speeding to approach the target) and avoidance (e.g.,

deceleration to avoid passing target) motivational components. The two primary goals of this study were: 1) to examine and quantify how these distinct motivational biases may impact driving behavior in individuals with depressed mood; 2) to develop a task that will allow modeling the relative contribution of approach and avoidance motivations in depression and other clinical disorders. We hypothesized that, in this task, depressed mood would alter both subjects' targeted stopping position and motor action. That is, based on evidence of increased sensitivity to punishment and avoidance tendency in depression, we expected that individuals with depressed mood would stop further away from instructed stopping position to avoid inaccuracy (i.e., passing target/crash).

5.2 Method

Participants

Thirty-nine college students (age range: 18-22 years old, 23 females /16 males) participated this study (approved by the Human Research Protections Program at San Diego State University). They were recruited from San Diego State University through an online system as part of Psychology 101 class during the spring of 2013. They were contacted and scheduled for an experiment session during winter quarter 2013. All participants signed informed consent, and were compensated $25 and 2.5 course credits for completing the study.

Before subjects performed the experimental task, they were administered the Beck Depression Inventory (BDI-II; Beck AT et al. 1996) to use as a current index of depression severity. Participants' BDI scores ranged from 0 to 40 (mean = 13.5, std =

10.77; median = 10). For the following analyses, we grouped subjects by the severity of their depression symptoms based on their BDI score (cutoff based on Beck et al., 1996). The final groups included 13 non-depressed subjects with BDI<=8 (Non Dep), 14 mildly depressed subjects with 9<=BDI<=15 (Mild Dep), and 12 moderately to severely depressed subjects with BDI>=15 (Mod-Sev Dep).

Driving Task

The experiment was comprised of two conditions (Task1: stop sign; Task2: wall) and programmed using Matlab on a 15.4-inch Mac-book. Participants completed two blocks of 15 trials for each condition in the following order: Task1, Task2, Task1, Task 2. They were instructed to drive a virtual car on a computer screen from an initial position to either a stop sign or a wall (equal distance: 10.62 cm/465 pixels) by pushing or pulling a joystick (Logitech Extreme 3D Pro) to control the acceleration and the deceleration of the car (Figure 5.1a). They were instructed to drive as quickly as possible and stop as close to the stop-sign or the wall as possible without crossing the stop-sign or crashing into the wall, respectively. Each trial had a fixed time-window of 6 seconds.

The car dynamics adhered to the standard second order Newtonian differential equations of motions with viscous friction. The force applied to the car at every point in time was proportional to the deviation of the Joystick from the resting point. If forward the force pushed the car forwards. If backwards it push the car backwards. (Please see appendix for detail).

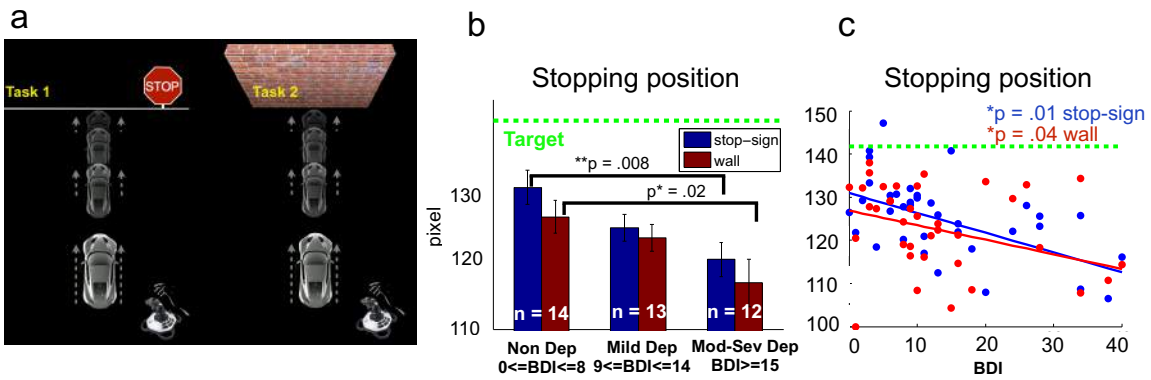

Figure 5.1: a. Experiment paradigm. In Task1 (stop-sign), subjects were instructed to drive the car to the stop-sign as quickly as possible and stop as close as possible to the sign without crossing the white line. In Task 2 (wall), subjects received the same instructions as in task 1, with the addition that they could not crash into the wall (instead of not crossing the stop-sign). b. The influence of depressed mood on stopping position in stop-sign and wall condition of stopping position: Group comparison; c. The influence of depressed mood on stopping position in stop-sign and wall condition of stopping position: stopping distance as a function of BDI.

Data Analytic Approach

The two main dependent variables were participants' stopping position, i.e. distance from stop line/wall in pixels and participants' continuous acceleration/deceleration in joystick control (measured as the joystick position). The first variable corresponds to car position when T= 6s, i.e., trial time window, and the second one to recorded continuous joystick action for acceleration/deceleration. Based on the observed trajectories, further analyses were performed on individuals' maximum acceleration during the initial trial phase (i.e., max acceleration during the first 200 pixels) and their individual maximum deceleration in the late trial/close to target phase (i.e., max deceleration during the last 200 pixels].

To analyze these data, a linear mixed model (LMM) was fitted for each main

dependent variable, with subject model as a random effect and depression severity group as well as condition as fixed effects. To ensure participants' performance was stable and avoid confounding factors such as subjects' learning curve in the task, only data collected in the second block for each condition were included in the analyses. We first assessed for any group differences in stopping distance and stopping velocity. If differences were observed, we followed up by investigating when this difference happened within a trial epoch, and what actions caused this difference. For significance tests of main effects and interactions, we report change in log likelihood ratio (approximated by a chi-square distribution). We also provide the model coefficients of interest with corresponding p values.

5.3 Results

Depressed mood associated with more remote stopping position

We first examined the effect of depressed mood group (i.e., Non Dep, Mild Dep and Mod-Sev Dep) and experimental condition (i.e., Stop sign, Wall) on stopping distance. There was a significant group main effect (chi-square = 12.60, df = 2, p = .002) and significant interaction between group and condition (chi-square = 12.98, df = 4, p = .011), while no main effect of condition was observed (chi-square = 2.86, df = 1, p>.05).

Specifically, in the stop-sign condition, relative to Non Dep individuals, the Mod-Sev Dep group stopped significantly further away from the target (B = -11.08, p =.008; Non Dep mean = 131.62, Mod-Sev Dep mean = 120.54) while the Mild Dep and the Non Dep groups did not differ in stopping distance (p>.05;see Figure 1b). In the wall

condition, relative to Non Dep individuals, the Mod-Sev Dep group had a significantly further stopping distance (B = -9.86, p = .02; Non Dep mean = 127.13, Mod-Sev Dep mean = 117.27). The Mild Dep and the Non Dep groups did not differ in their stopping distance (p>.05) in the wall condition.

Individual's average stopping position was negatively correlated with BDI scores in both stop-sign and wall conditions indicating that the degree of depression was associated with differences in behavior on this task, we found that (Figure 1c). In other words, those individuals with the highest BDI scores stopped furthest away from the target position. Similar as the result from mixed linear model, this negative correlation was slightly higher in stop-sign condition (regression coefficient =-.4593, p = .01) than in the wall condition (regression coefficient = -.3376, p = .04), but this difference is not significant (chi-square = 3.62, df = 2, p = .16).

Depressed mood affects the magnitude of deceleration (avoidance), but not acceleration (approach)

Next, we examined continuous car position between three groups within the 6 seconds trial time window for the stop-sign and wall conditions separately. Figure 2 indicates that car position between Non Dep and Mod-Sev Dep groups started to differ when it was close the target (2a. stop-sign and 2b.wall), in particular during the last 2 seconds of a trial (Figure 2a&2b Zoom in). It also shows there were no significant position differences between Non Dep and Mild Dep groups throughout the 6-sec trial window, which is consistent with the stopping position results above. To further investigate the cause of this position difference between the Non Dep and Mod-Sev Dep

groups, we looked at joystick actions between those the two groups. Action (accelerating if > 0, and decelerating if <0) is plotted as a function of distance to target in Figure 2c&2d.

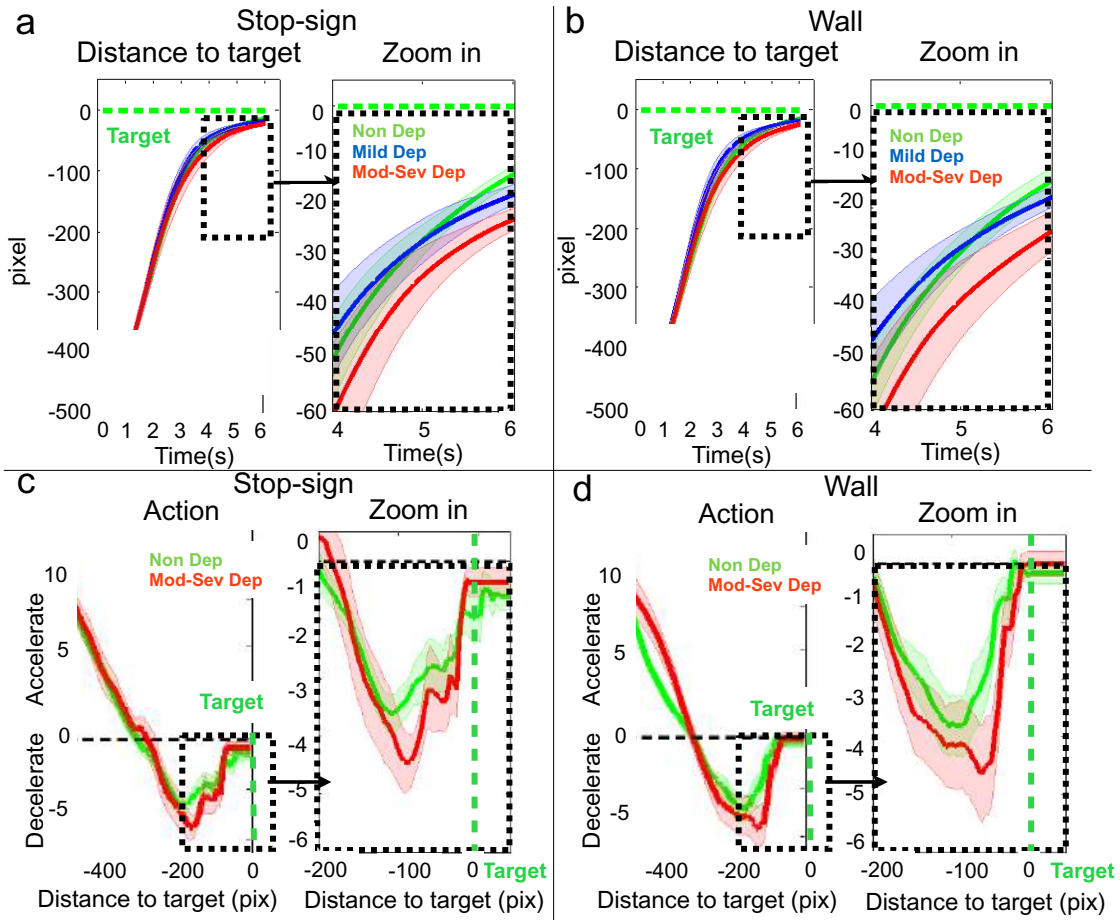

Figure 5.2 a & b: Car position as a function of time within a trial (6-seconds): a (stop-sign), b (wall); c & d: Action as a function of distance to target: c (stop-sign), d (wall).

Maximum Acceleration: there was no significant group main effect (chi-square = 0.10, df = 1, p = .75), condition main effect (chi-square = .58, df = 1, p = .45), or interaction effect (chi-square = .89, df = 2, p = .64) for maximum acceleration (i.e.,

individuals' peak acceleration within first 3-seconds trial window) between Non Dep and Mod-Sev Dep groups in the stop-sign and the wall condition.

Maximum Deceleration: We found a significant group effect (chi-square = 17.75, df = 1, p <.0000) and group by condition interaction (chi-square = 22.58, df = 2, p<.0000) for maximum deceleration (i.e., individuals' peak deceleration within last 2-sec trial window). No significant condition main effect was observed (chi-square = .001, df = 1, p = .97). More specifically, the Mod-Sev Dep group had overall greater maximum deceleration (i.e., larger dip) relative to the Non Dep group (B=-.97,p<.001). In addition, this group difference was statistically significant in the stop-sign condition (B=-1.47; Mod-Sev Dep Mean= -5.92 pix/s^2; Non Dep Mean=-4.09 pix/s^2) but not in the wall condition (p>.05).

5.4 Discussion

In this study, we aimed to disambiguate approach from avoidance motivational dysfunction in depression and its effects on motor control. We designed a computer-simulated driving-task to examine how the severity of depression (measured with BDI scores) would impact the motivational underpinnings of individuals' driving behavior through their accelerating and decelerating control. As we outline below, our results are consistent with the hypothesis that individuals with depressed mood engage in stronger avoidance actions while minimally impacting approach-based actions when getting closer to a target. As a consequence, it may negatively deceleration and stopping distance. We summarize our results and potential implications below.

Relative to healthy controls, moderately to severely depressed individuals stopped further away from the stop sign or wall, suggesting they may have a different 'target' stopping position. In addition, more severely depressed individuals demonstrated a significantly stronger deceleration when approaching target. These results suggest that greater depression severity may promote stronger proactive avoidance of the target (pulling away) in terms of planning stopping position and adjusting movement closer to end goal. This is consistent with empirical evidence of both heightened sensitivity to and stronger avoidance of negative stimuli in depressed individuals, both behaviorally (Kellough, Beevers, Ellis & Wells, 2008; Seidel et al 2010) and neurally (Holmes & Pizzagalli 2008).

Alternatively, given evidence of decreased reward-seeking and approach motivation in depression (Beck, 2011), it could be argued that more severely depressed individuals have less motivation to perform the task due to anhedonia (Der-Avakian & Markou, 2012). That is they may be less motivated to be precise about their stopping position and as a result stop further away. However, our results show that the high depression group did not differ from healthy controls in their positive control action (approach) values, but rather applied higher deceleration when the car was getting closer to the stop sign. Overall, these results are broadly consistent with evidence of impaired pro-active cognitive control in clinically depressed individuals, both behaviorally (i.e., longer response latency to interference) and neurally (i.e, longer event-related potentials in dorsal ACC; Vanderhasselt et al., 2014). Thus, depressed individuals may be less efficient in incorporating environmental cues that indicate increased need for cognitive control (e.g., seeing a stop sign in the distance).

Another plausible explanation is that more depressed individuals are lacking fine and/or more precise motor control caused by impaired sensorimotor skills, which would lead to more inaccuracy and variability in stopping positions. That is, while there is somewhat mixed evidence of psychomotor retardation in depression (Sabbe, Hulstijn, Hoof, Tuynman-Qua & Zitman, 1999), higher depression severity could result in slower sensory processing of the target and slower motor initiation. In future work, we plan to incorporate individual differences in sensorimotor skills in the modeling of how goal setting and reward-processing influence motor control.

5.5 Summary and Future Directions

In conclusion, these results provide empirical evidence of altered motivational influences in motor control performance among individuals with depressed mood. We found that those individuals with the most severe depression stopped furthest away from the target and showed the greatest deceleration when approaching the instructed stopping target. Together, these results show that the influence of depressed mood in modulating approach and avoidance motivations and cognitive control may be more fluid and dynamic than expected, with the potential to influence different stages (action planning and execution) of a motor-control task. They further highlight the usefulness of a continuous time analysis as done in the present study. Finally, our task provides a useful platform to study the precise cognitive processes underlying emotion/cognitive control interaction in depression. Specifically, we plan to use computational models (e.g. control theory) to examine if/how the severity of sensorimotor impairment interacts with

motivational deficits in more severely depressed individuals. This research will in turn help to more precisely quantify cognitive control deficits in depression and their associated neural circuitry.

5.6 Acknowledgements

Chapter 5, in full, is submitted to Emotion (brief report). Huang H, Movellan J, Paulus M, Harlé K.

5.7 Appendix

Car dynamics in the experiment:

In our experiment (both stop-sign and wall conditions), the car has a linear dynamic system, which is modeled by the following equation:

$$dXt = AXtdt + BUtdt,$$

in which Xt = state [car position, car velocity], Ut = control action (acceleration or deceleration, based on joystick position), A = [0 1; 0 -.35] is the dynamic matrix, and B = [0; 0.5] the input matrix. Those valued are chosen such that the car's velocity is the rate of change of the position:

$$dXt = Vtdt ,$$

and velocity is controlled by joystick action with the influence of viscosity $dVt = -.35Vtdt+.5Utdt$, in which Ut is measured through joystick position (ranges from -10 to

10, with 0 being the resting position/no action, positive number indicating accelerating action and negative number indicating decelerating action).

In the reported analysis, we used *Xt* at t = T (T = 6s, trial time window) as the stopping position, maximum *Ut* during the first 200 pixels as participant's max accelerating action in the during the initial trial phase, and minimal *Ut* during the last 200 pixels as participant's max decelerating action in the later trial phase.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review*, *30*(2), 217-237.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory. The psychological corporation. *San Antonio, TX*.

Beck, J. S. (2011). *Cognitive behavior therapy: Basics and beyond*. Guilford Press.

Bulmash, E. L., Moller, H. J., Kayumov, L., Shen, J., Wang, X., & Shapiro, C. M. (2006). Psychomotor disturbance in depression: assessment using a driving simulator paradigm. *Journal of affective disorders*, *93*(1), 213-218.

Caligiuri, M. P., & Ellwanger, J. (2000). Motor and cognitive aspects of motor retardation in depression. *Journal of affective disorders*, *57*(1), 83-93.

Der-Avakian, A., & Markou, A. (2012). The neurobiology of anhedonia and other reward-related deficits. *Trends in neurosciences*, *35*(1), 68-77.

Dickson, J. M., & MacLeod, A. K. (2004). Approach and avoidance goals and plans: Their relationship to anxiety and depression. *Cognitive Therapy and Research*, *28*(3), 415-432.

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467-477.

Drevets, W. C. (2001). Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders.*Current opinion in neurobiology*, *11*(2), 240-249.

Dunn, B. D., Dalgleish, T., Lawrence, A. D., Cusack, R., & Ogilvie, A. D. (2004). Categorical and dimensional reports of experienced affect to emotion-inducing pictures in depression. *Journal of abnormal psychology*, *113*(4), 654.

Epstein, J., Pan, H., Kocsis, J., Yang, Y., Butler, T., Chusid, J., ... & Silbersweig, D. (2006). Lack of ventral striatal response to positive stimuli in depressed versus normal subjects. *American Journal of Psychiatry*, *163*(10), 1784-1790.

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., Reiss A. L., & Schatzberg, A. F. (2007). Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biological psychiatry*, *62*(5), 429-437.

Heller, A. S., Johnstone, T., Shackman, A. J., Light, S. N., Peterson, M. J., Kolden, G. G., Kalin N.H., & Davidson, R. J. (2009). Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation. *Proceedings of the National Academy of Sciences*,*106*(52), 22445-22450.

Henriques, J. B., & Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition & Emotion*, *14*(5), 711-724.

Holmes, A. J., & Pizzagalli, D. A. (2008). Response conflict and frontocingulate dysfunction in unmedicated participants with major depression. *Neuropsychologia*, *46*(12), 2904-2913.

Kellough, J. L., Beevers, C. G., Ellis, A. J., & Wells, T. T. (2008). Time course of selective attention in clinically depressed young adults: An eye tracking study. *Behaviour research and therapy*, *46*(11), 1238-1243.

Layne, C., Merry, J., Christian, J., & Ginn, P. (1982). Motivational deficit in depression. *Cognitive Therapy and Research*, *6*(3), 259-273.

Liotti, M., & Mayberg, H. S. (2001). The role of functional neuroimaging in the neuropsychology of depression. *Journal of clinical and experimental neuropsychology, 23*(1), 121-136.

Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., ... & Fox, P. T. (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness.*American Journal of Psychiatry*, *156*(5), 675-682.

Mitterschiffthaler, M. T., Williams, S. C., Walsh, N. D., Cleare, A. J., Donaldson, C., Scott, J., & Fu, C. H. (2008). Neural basis of the emotional Stroop interference effect in major depression. [Research Support, Non-U.S. Gov't]. *Psychol Med, 38*(2), 247-256. doi: 10.1017/S0033291707001523

Sabbe, B., Hulstijn, W., Hoof, J. van, Tuynman-Qua, H., & Zitman, F. (1999, September). *Retardation in depres- sion: assessment by means of simple motor tasks* (Vol. 55) (No. 1). Elsevier/North-Holland Biomedical Press.

Seidel, E. M., Habel, U., Finkelmeyer, A., Schneider, F., Gur, R. C., & Derntl, B. (2010). Implicit and explicit behavioral tendencies in male and female depression. *Psychiatry research*, *177*(1), 124-130.

Selzer, M.L., Rogers, J.E., Kern, S., 1968. Fatal accidents: the role of psychopathology, social stress, and acute disturbance. Am. J. Psychiatry 124, 1028–1036

Siegle, G. J., Steinhauer, S. R., Thase, M. E., Stenger, V. A., & Carter, C. S. (2002). Can't shake that feeling: event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biological psychiatry*, *51*(9), 693-707.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of behavior therapy and experimental psychiatry*,*39*(4), 567-576.

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical psychology review*, *31*(7), 1156-1168.

Vanderhasselt, M.-A., De Raedt, R., De Paepe, A., Aarts, K., Otte, G., Van Dorpe, J., & Pourtois, G. (2014). Abnormal proactive and reactive cognitive control during conflict processing in major depression. *Journal of Abnormal Psychology, 123*(1), 68.

Vrijsen, J. N., van Oostrom, I., Speckens, A., Becker, E. S., & Rinck, M. (2013). Approach and avoidance of emotional faces in happy and sad mood.*Cognitive therapy and research*, *37*(1), 1-6.

Wagner, G., Sinsel, E., Sobanski, T., Kohler, S., Marinou, V., Mentzel, H. J., . . . Schlosser, R. G. (2006). Cortical inefficiency in patients with unipolar depression: an event-related FMRI study with the Stroop task. [Research Support, Non-U.S. Gov't]. *Biol Psychiatry, 59*(10), 958-965. doi: 10.1016/j.biopsych.2005.10.025

Chapter 6

Inverse Optimal model of Depressive Behavior in a Simulated Driving Task

Abstract-Depressed individuals have been shown to have impaired sensorimotor system, skewed goal setting towards reward or punishment, and lack of motivation. However, it has been difficult to distinguish the their effects in goal directed motor tasks, as the observed behavior is confounded from their mixed influences. Here, we propose to use inverse optimal control approach and a computer-simulated driving task to analyze and factorize performance deficits into three components: sensorimotor speed, goal setting and motivation. 66 subjects with minimal to severe depression symptoms (0<=BDI<=39) participated in our experiment. In Task 1, subjects were instructed to push a joystick as quickly as possible once they observe motion onset of a virtual car. In Task 2, subjects were instructed to use the joystick to move a virtual car as quickly as possible and stop it as close as possible to a stop sign. We estimated sensorimotor speed from Task 1 based on recorded continuous joystick position and subject's reaction time to car motion-set and movement time to push the joystick to instructed position. Taking into account of the sensorimotor speed from Task 1, we recovered the underlying reward function that best explain subject's behavior in Task 2, which includes the intended stop distance to the stop sign (goal state), and the amount of effort one is willing to spend to

achieve the goal state (motivation). Results suggest, that relative to healthy controls, depressed individuals: 1) have slower sensorimotor speed; 2) have different goal states; 3) only severely depressed individuals (BDI>=29) have significant lower motivation. These results indicate that inverse optimal control framework can quantify in individual with depression the combination of sensorimotor deficits and impaired reward-evaluation that includes different goal setting and motivational factors.

6. 1 Introduction

Sensorimotor speed, *goal setting*, and the amount of effort one is willing to spend to achieve the goal (i.e. *motivation)* are three factors that influence performances in goal-directed motor tasks. It is important to distinguish their effects to understand the behavior differences observed in depressive individuals. At present, however, most of the experimental paradigms are restricted to observing the behavior of discrete actions and using reaction time as the measure for motor performance. Indeed, it is difficult to distinguish those factors from discrete actions or reaction time, because those factors jointly influence both the cognitive control (movement planning) and movement execution (Sobin & Sackeim 1997). For example, in a goal-directed motor task, a slower action may be caused by slower movement execution due to impaired sensorimotor system; or by different goals, for example, minimizing the control noise with slower velocity; or lack of motivation, for example, having less subject value for the external reward. Thus it is difficult to investigate their individual influence from the confounded result.

It has been shown that depressed individuals have different performance from healthy controls in following aspects: increased reaction time (Sobin & Sackeim 1997) and lower velocity (Caligiuri & Ellwanger 2000), which indicates impaired *sensorimotor system*; more sensitive to punishment than to reward (Trew 2011), which indicates skewed *goal setting*; and less willing to spend effort to gain reward (Treadway et al. 2009; Der-Avakian et al. 2012), which indicates the lack of *motivation*. However, it is important to realize that those results are from mixed effects caused by joint influences of *sensorimotor speed*, *goal setting, and motivational factors*. To be able to interpret the different behavior in depressed individuals, first we need to have a better understanding of the underlying causes of those differences. Thus the goal of this paper is to provide an experimental paradigm to assess sensorimotor speed, goal setting and motivational factors, and apply a computational framework that can take into consideration of all factors, to quantity their individual and joint effects to explain observed behavior.

Motor Control in a sensorimotor feedback loop

Optimal control theory has been shown to be an effective computational framework to explain human movements in continuous time (Todorov & Jordan 2002). Motor control in a goal-directed task is a dynamic process of sensorimotor integration, in which the brain takes sensory information and uses it to make continuous motor actions. Optimal control theory frames this dynamic process in a feedback control loop, in which the optimal controller estimates the current state at time t, produces a motor command based on the goal and keeps an efference copy (the expected outcome of the motor command) at the state estimator, and sends the motor command to muscles to generate

the movement. The state estimator will update the efference copy with the delayed sensory observation to predict state at next time point t+1 and the optimal controller will generate new motor commands until the goal is reached.

In optimal control theory, actions are chosen to optimize a performance criterion. The performance criterion is defined as a reward-function that includes task-related performance measure and action cost. For example, in a task that instructs subjects to drive to a location A as quickly as possible, the performance measure can be the stopping distance to A, and the action cost can be the accumulated effort of accelerating and decelerating controls. Different individuals may have different target stopping distance to A, and different weights to assess the ratio of the closeness to the target location over the action cost (the amount of effort one is willing to spend to achieve the intended stopping distance), thereby forming different reward-functions.

With different reward-functions in mind, there will be different action-planning strategies, which are defined as control-policies. A control-policy comprises a series of dynamic decisions modulating actions at given states in continuous time (Shadmehr 2008). In *a forward model,* with experimentally defined reward function (e.g., points), we can derive the optimal control-policy to optimize the reward function. In *an inverse model* (Ng & Russell, 2000), with observed continuous actions, we can infer the control-policy, and recover the reward-function used in developing this control-policy. Thus the objective of inverse optimal control is to infer individuals' reward-function based on observed behaviors.

Inverse optimal control approach will provide a quantitative comparison of how different reward evaluation between depressed and healthy controls lead to observed

behavioral differences. Within this framework, we can firstly measure sensorimotor speed that has minimal influences from reward evaluation, and take into this baseline sensorimotor speed in the dynamic of the control-process, and use observed behavior to infer the reward function in a goal-directed motor task.

In summary, we will apply inverse reinforcement learning approach to investigate how sensorimotor impairments and reward-evaluation in depressed individuals influence their motor control in a simulated driving task.

6. 2 Methods

Participants

66 college students (20 male and 46 female subjects, age 18-27, mean age 20.6, std =2.04) in UCSD participated this study in Fall quarter 2013 and Winter & Spring quarter 2014. They signed up through UCSD SONA system, and then completed phone-screening and on-line BDI (Beck Depression Inventory, BDI-II, Beck et al. 1996) measure. Qualified subjects completed the experiment (with a second BDI measured prior to the task) in the lab, and were compensated by 2 course credits. Their onsite BDI range from 0 to 39 with mean BDI=12.59 (std=10.55), median BDI=10. Subjects were divided into four groups based on their onsite BDI (Beck AT et al. 1996) as follows: Non-dep ($0 \leq BDI \leq 5$, n = 17), Min-mild dep ($6 \leq BDI \leq 19$, n = 33), Mod-dep ($20 \leq BDI \leq 28$, n = 9), and Sev-dep ($29 \leq BDI \leq 63$, n = 7).

Experiment

Subjects were instructed to complete two tasks in this experiment. Both tasks were computer experiments (on a 15 inch MacBook Pro) programmed in Matlab. We recorded their continuous actions using a gaming joystick (Thrust- master HOTAS Warthog Flight Stick). The goal of Task 1 (Move-and-Go) is to measure individual's perceptual and motor speed, and the goal of Task 2 (Speed-and-Stop) is to apply inverse optimal control model to recover reward-function in a goal-directed task.

Task 1: Move-and-Go

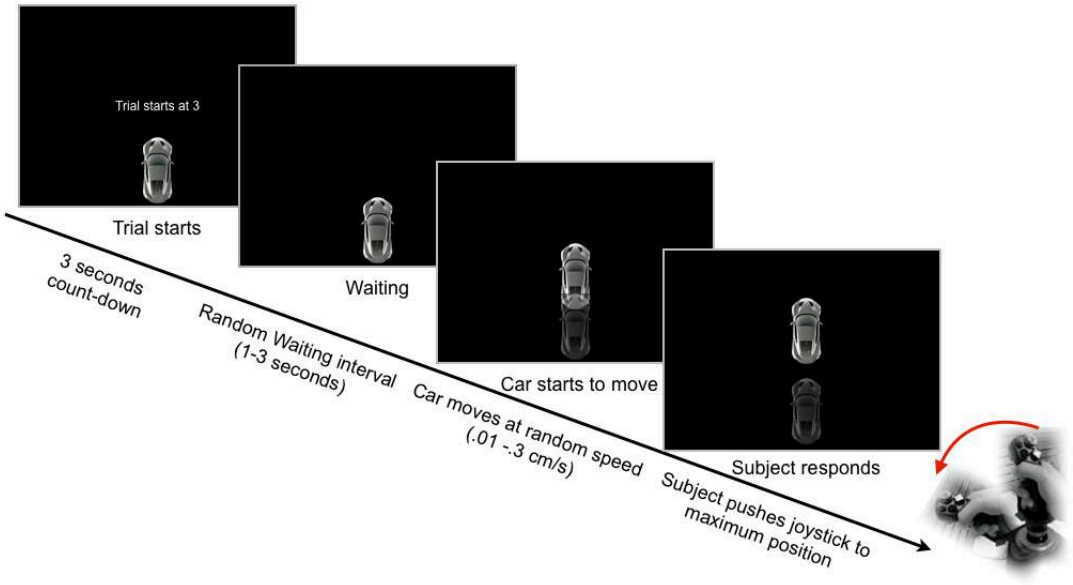

Figure 6.1: Task 1

Subjects performed Task 1 twice (120 trials, before and after Task 2). In each trial (Figure 6.1), a car would appear on the bottom of the screen, and subjects were instructed to push the joystick from resting position forward to the maximum position as quickly as possible once they observe the car move. Each trial started with a 3-second countdown

and a random waiting interval (1-3 seconds), then the car would start to move at a randomly selected speed (.01-.3 cm/second). Trials ended once subjects pushed the joystick at its maximum forward position.

Task 2: Speed-and-Stop

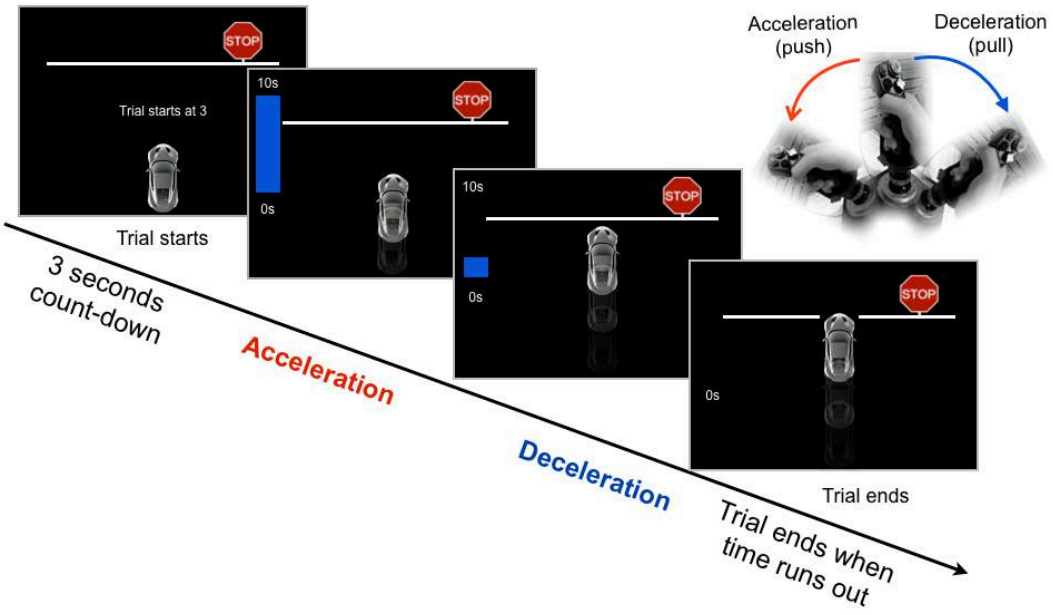

Figure 6.2: Task 2

There were 3 blocks, with 20 trials/block in Task 2. In each trial (Figure 6.2), subjects were instructed to drive a virtual car as quickly as possible and stop at a stop sign (distance: 10.62 cm) without crossing the stop-line, with a fixed time window of 10-second. Each trial started with a 3-second countdown and ended when time ran out, with no performance feedback (e.g., points) in the end. The car has a linear dynamic system (see Appendix), in which the car position is controlled by continuous joystick position.

6.3 Models

We propose to use inverse optimal control model to distinguish sensorimotor speed, goal setting and motivational effects in observed behavior (Fig 6.3). To achieve that, we first assessed individual's sensorimotor system by estimating their sensory speed (delay in perceiving sensory observation at time *t*) and motor speed (delay in executing motor command at time *t*) (Task 1: move-and-go). Then we estimated their goal state (intended stopping distance) and motivation (the amount effort one is willing to spend to achieve the goal state) in the reward function in Task 2 (speed-and-stop), with the sensory and motor delay parameters from Task 1.

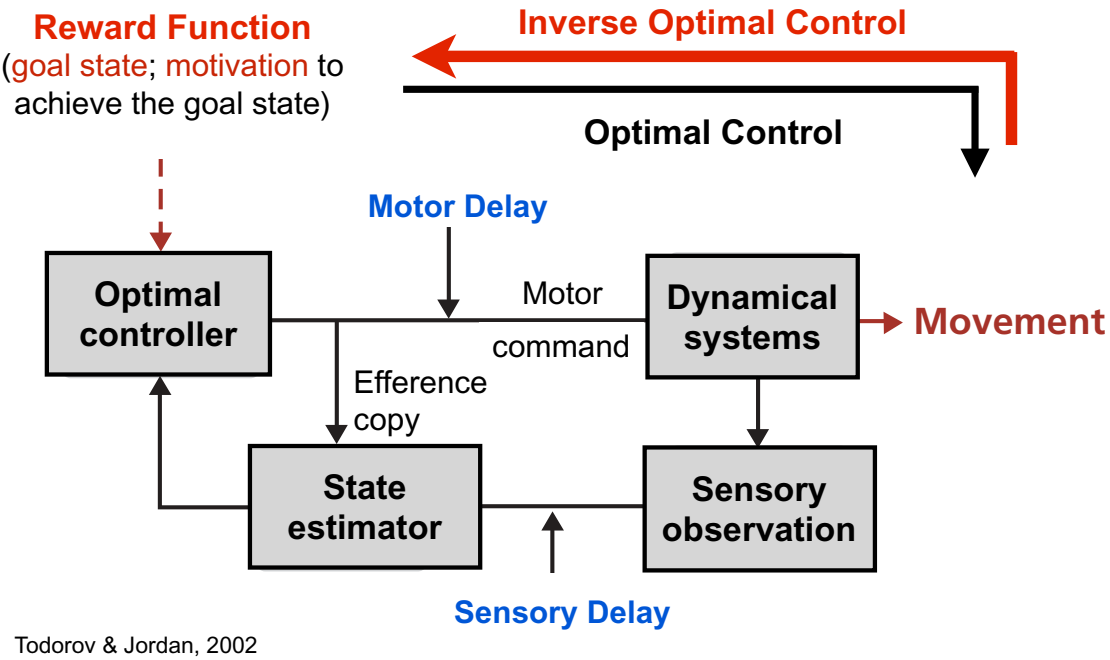

Figure 6.3: Computational Framework of the Inverse Optimal Model

Task 1 (move-and-go) was designed to estimate sensory speed γ and motor speed β. Subjects' perceived car position was modeled as a delayed true car position due to the limit of sensory processing speed γ (Appendix: Eq.1). The higher the γ, the closer the perceived car position is to the true car position. Joystick position is modeled as a delayed execution from target joystick position, due to the limit of motor execution speed β (Appendix: Eq. 3). The higher the β, the closer joystick action is to the desired target position. Thus the minimal time for joystick to reach target joystick position is movement time (Appendix: Eq. 4). Reaction time to car motion-set and true car position were used to recover γ, and recorded joystick action and movement time were used to recover β, by applying Maximum Likelihood Estimation (i.e. optimizing over γ, β between predicted reaction time, movement time and observed data.)

Task 2 (speed-and-stop) was designed to estimate individual's reward-function. It is a function of goal stopping distance (*goal*) and the ratio between internal reward for achieving the goal and the energy expenditure (*motivation*). *Goal* measures individual's intended stopping distance from the stop sign. *Motivation* measures individual's willingness to reach the goal stopping distance. With higher motivation, one will be willing to spend more effort to gain internal reward. In a quadratic reward function, *goal* represents the optimal point of the reward function, and *motivation* represents the hessian of the reward function. We formulate the driving task as a Linear Quadratic Gaussian (LQG) problem with a linear dynamic system taking into account of the sensorimotor speed estimated from Task 1, and a quadratic reward function of *goal* and *motivation*.

Linear dynamic system Assuming the driving task as a linear dynamic system (Equation 1) with a partial hidden state X_t and observable feedback Z_t, in which X_t is a

vector including the (hidden) true car distance to target stopping position at time *t*, joystick action at time *t*, and perceived car distance to target stopping position at time *t*.

Linear dynamic system: $dX_t = AX_t\,dt + BU_t\,dt$ (1)

Observation: $Z_t = CX_t + V_t$ (2)

In which, A is a dynamics matrix (SI: Eq. 7) with motor and perceptual speed estimated from Task 1 and parameters of car dynamics (assuming known), B is input matrix (SI: Eq. 8) which takes into consideration of subject's motor speed, C = [0 0 1]' (i.e. perceived car distance), and V_t is Gaussian noise.

Quadratic reward function We assume the reward function r (X_t, U_t) is a quadratic function of current state X_t and action U_t (Equation 3). It evaluates current state X_t based on its distance from the goal state (G) and the ratio of the weight on this distance over the energy expenditure, which is defined as motivation M in our framework. (Please see SI for detail.)

Reward function: $r(X_t,U_t) = g(X_t, G, M) - U_t^2$ (3)

In forward LQG problems, the optimal controller generates an optimal control policy that maximizes a given reward function. Figure 6.4 shows in a forward model of this driving task, the different effects of model parameters (motor speed β, goal stopping distance G and motivation M) on car position and joystick control. In inverse LQG problems, observed movements are used to infer the underlying reward function that best explains the observed behavior. We estimated G (goal stopping position) for each individual subject using MLE, and recovered M (Appendix: Eq. 16) from analytical result using recorded continuous car position and joystick actions (See Appendix for detail).

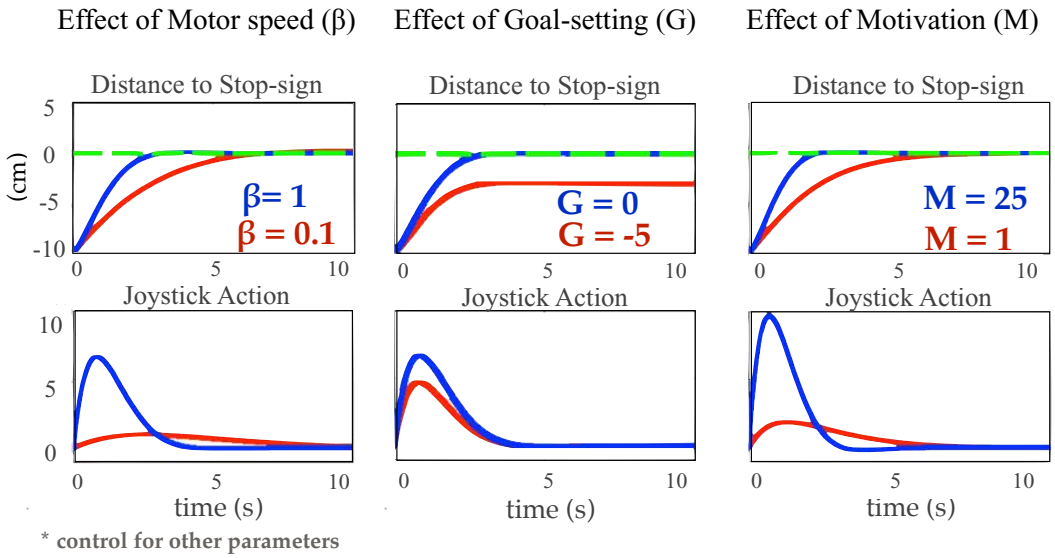

Figure 6.4: Influences of model parameters. Left (β): higher motor speed lead to faster arrival time to target; Middle (G): different goal distances lead to different stopping position; Right (M): higher motivational level lead to faster arrival time and closer distance to target;

6.4 Results

Sensorimotor speed

Model results from Task 1 showed depressed individuals had significantly slower sensory speed (Fig 6.5 Left) (** < .01), as well as significantly slower motor speed (Fig 6.5 Right) than Non-dep group (* < .05, **<= .01). No significant differences of sensory speed or motor speed were observed among the three depressive groups (p>.05). Those results are consistent with their longer reaction time to car motion-onset (mean reaction time across different car speeds: Non-dep: 1.19 s; Mid-dep: 1.33 s; Mod-dep: 1.45 s; Sev-dep: 1.42 s) and longer movement time to push the joystick from resting to maximum

forward position (mean movement time: Non-dep: 0.20 s; Mid-dep: 0.27 s; Mod-dep: 0.29 s; Sev-dep: 0.32 s).

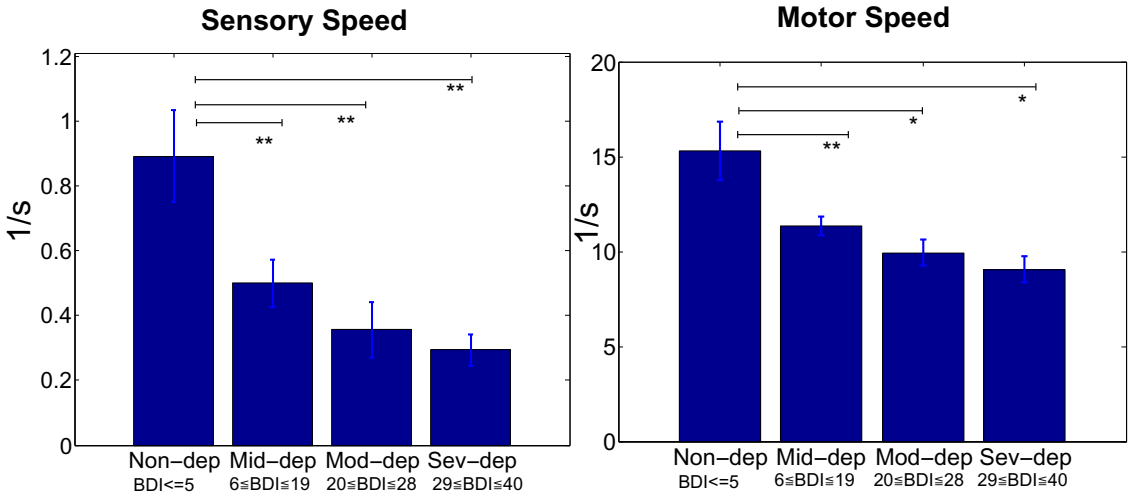

Figure 6.5: Sensory and motor speed estimated from Task 1.

Goal stopping distance

A linear mixed effect model (LMM) was fitted for mean goal stopping distance (G) across three blocks (Fig 6.6), with subject model as a random effect and depressive groups as fixed effect. We found a significant main effect of depressive groups (chi-square = 27.586, df = 3, $p < .001$) on mean goal stopping distance G. More specifically, relative to Non-dep individuals, all depressive groups had significant further stopping distance (B (Mid-dep) = - 0.11, p = .01; B (Mod-dep) = -.17, p = .006; B (Sev-dep) = -.38, $p < .001$). Consistent with grouped results, we also found a significant negative effect of BDI on goal stopping distance (B = -.009, $p < .001$).

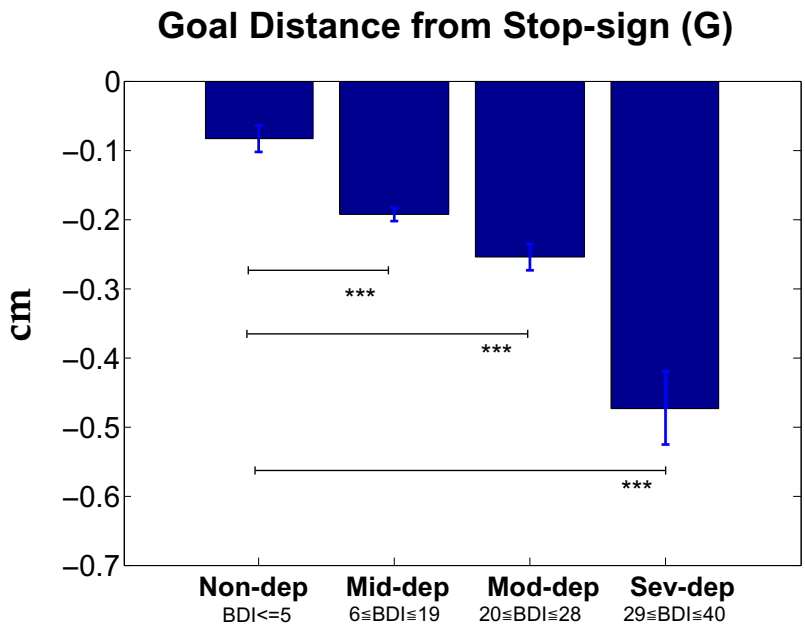

Figure 6.6: Goal distance estimated from Task 2. Averaged across blocks.

We then examined the effect of experimental blocks (Fig 6.7) on goal stopping distance and found a significant block main effect (chi-square = 18.744, df = 3, $p < .001$), and significant interaction between group and block (chi-square = 104.14, df = 5, $p < .001$). More specifically, comparing to each group's goal stopping distance in Block 1, Non-dep individuals had significantly decreasing goal distance in the 2nd block (B = .02, p =.003), while Mod-dep individuals had significant increasing goal distance in the 2nd block (B = -.07, $p < .001$), and Sev-dep individuals had significant increasing goal distance both in the 2nd block (B = -.18, p =.002) and the 3rd block (B = -.36, $p < .001$). No significant change in Mid-dep group over blocks was observed ($p>.1$).

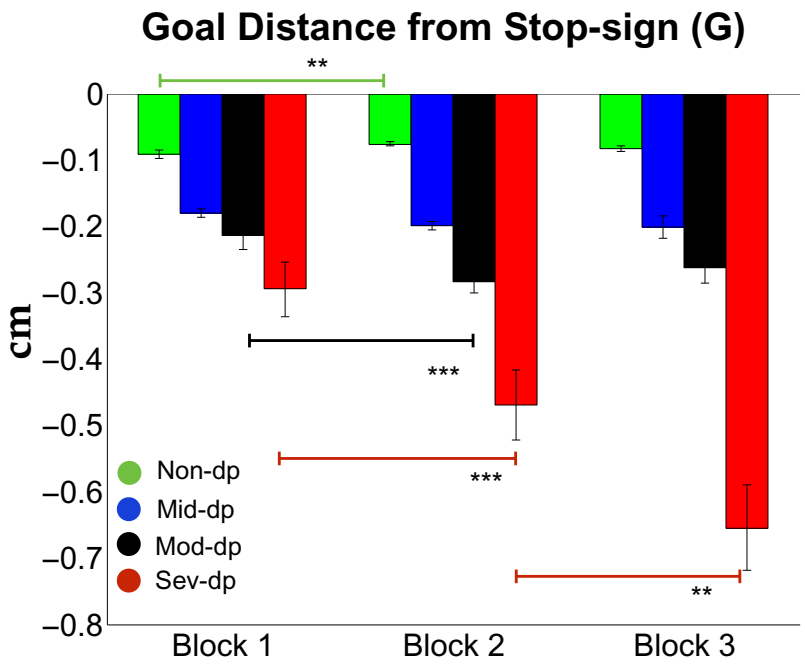

Figure 6.7: Goal distance estimated from Task 2. Separated by blocks and depressive groups.

Motivation to achieve the goal stopping distance

For motivation factors (Fig 6.8), only severely depressed individuals (BDI>=29) were reported to have significantly low motivation to achieve their goals (Fig 8), while no significant difference among Non-dep, Mid-dep and Mod-dep groups was observed.

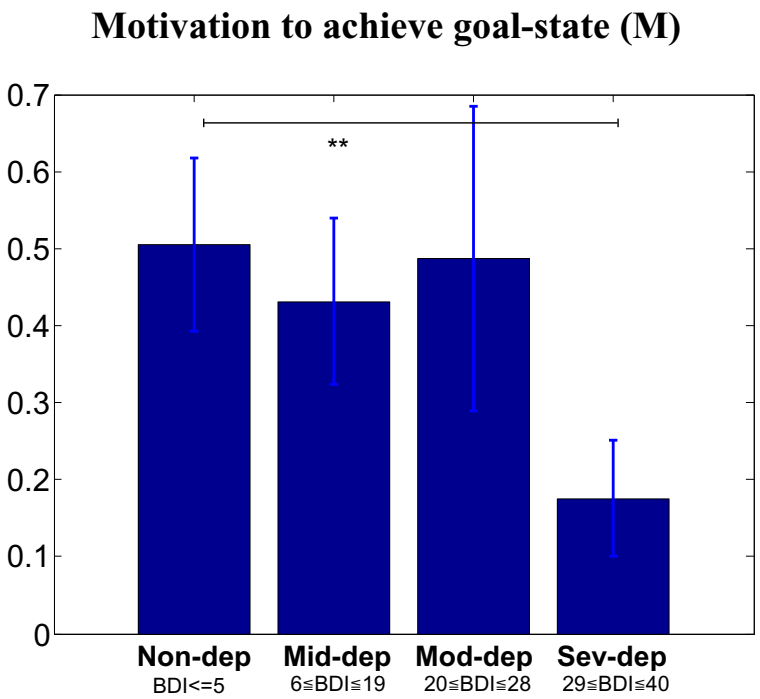

Figure 6.8: Motivation level estimated from Task 2.

Joint Distribution of Goal and Motivation

We then used mean G (-.2108) and mean M (.4309) to group subjects into four regions (Short G-High M, Long G-High M, Short G-Low M, and long G-low M) and looked at the distribution of depressed individuals in those regions. As shown in Figure 6.9, none of the healthy controls were in either of the Long Goal distance regions, but their percentage in Low Motivation or High Motivation is not significantly different (53% vs. 47%). None of the Mod & Sev-dep individuals was in Short Goal distance - High Motivation region. Most of the Sev-dep individuals are in the Long G-Low Motivation region (71%).

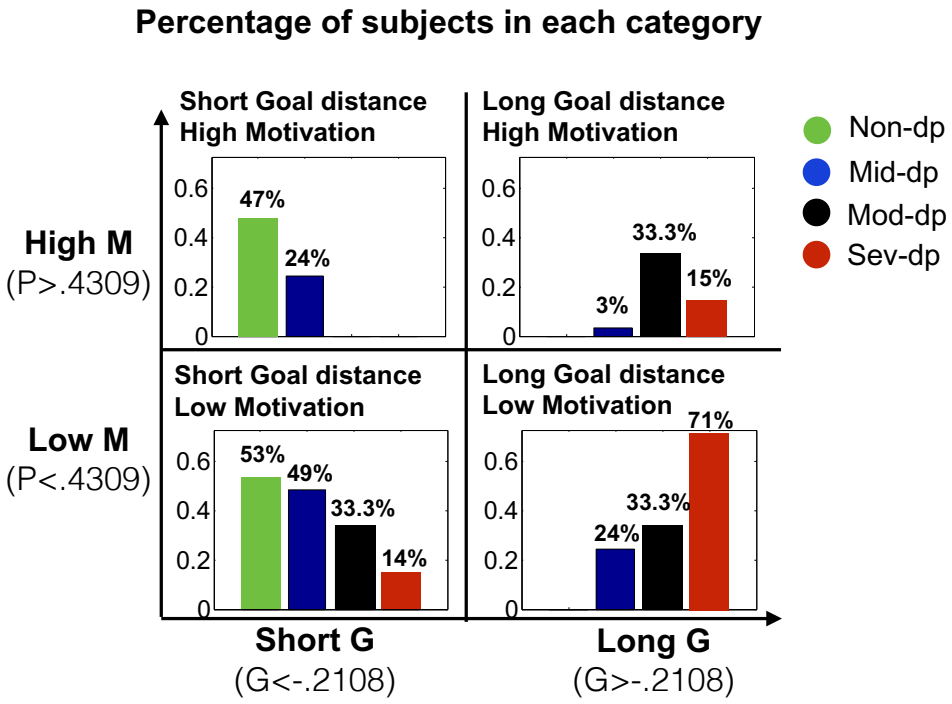

Figure 6.9: Joint distribution of Goal distance and Motivation level among depressive groups.

Reward function and action cost

Next, we explored the optimal accumulative action cost that generated from different reward function using a forward model. Model simulation (Fig 6.10 Left) suggests action cost increases as the goal stopping distance decreases (closer to stop-sign), and as motivation increases (willing to spend more effort to achieve goal stopping distance). Taking into account of the increased goal distance in depressed groups, and no significant motivation difference except for Sev-dep individuals, we can map their action cost based on individual reward function estimated from the model (Fig 6.10 Right). It indicates depressed individuals used reward function that leads to lower action cost.

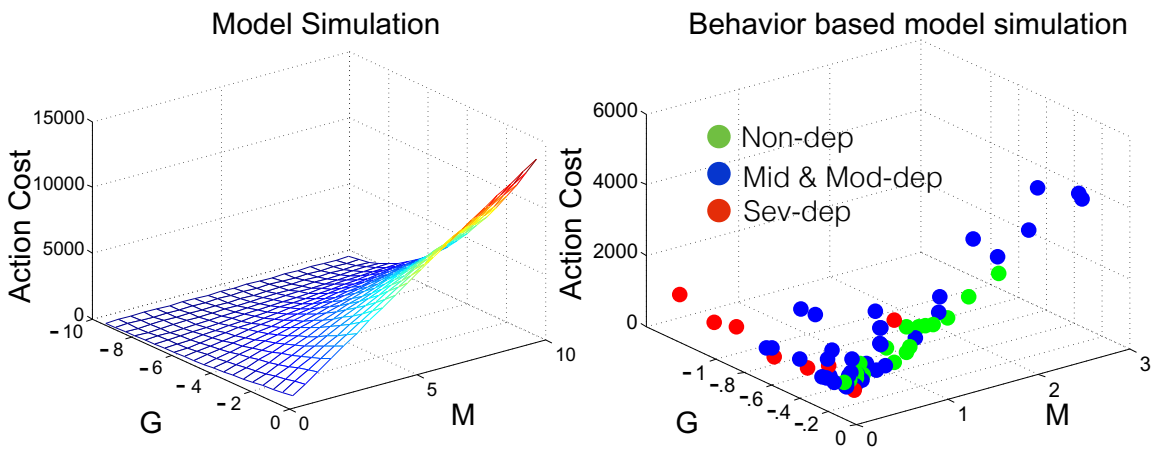

Figure 6.10: Action cost: model simulation and data. Left: Model Simulation of action cost as a function of Goal distance (G) and Motivation (M). Right: Model prediction of action cost based on estimated Goal distance and Motivation from behavioral data.

6.5 Discussion

In this paper, we proposed to use inverse optimal control approach to disentangle the effects of sensorimotor speed, goal setting, and motivational factors in goal directed motor tasks. We tested this framework in a simulated driving task and found that depressed individuals have slower sensorimotor speed, different goal setting (e.g. further away from stop-sign), and no significant motivation difference except in severely depressed individuals (BDI>=29).

Firstly, the proposed experimental paradigm expands from how we previously measure psychomotor disturbance symptoms using discrete actions to continuous actions. Thus in addition to reaction time, it can record the entire movement trajectory and provides more insights of how action is chosen at each time step at continuous time.

Secondly, this framework addressed the issue that observed motor behavior is a confounded result from sensorimotor speed, goal setting and motivational factors. It has been shown that depressed individuals have impairments of brain reward pathways which includes amygdala (Davis et al., 1998), anterior cingulate gyrus (Mayberg et al., 1994), etc., as well as impaired sensorimotor pathways which includes basal ganglia and frontal cortex (Nestler et al. 2002). Distinguishing those factors can provide insights to locate the corresponding neural circuits. Being able to measure those different factors is also of very importance in providing more effective measures for psychomotor disturbance in depressed individuals. Consistent with past studies that demonstrated the correlation between depression severity and psychomotor retardation (Blewett, 1992; Lemke et al., 1999), we found slower sensorimotor speed in depressed individuals. In addition, taking into account of the slower sensorimotor speed in depressed individuals, we showed that minimal to moderately depressed individuals do not have significantly less motivation, but only differ in where their subjective goal is. With more accurate measure of those three factors, more efficient treatment plans can be advised to individuals that need specific improvements.

Thirdly, this approach distinguishes goal setting from motivation in reward function, which may provide deeper understanding of anhedonia. As a core symptom in depression, anhedonia is considered as reward-motivation related impairment commonly observed in depressed individuals. While there have been studies showed that depression is associated with reduced hedonic activity, there are also mixed evidence as for if anhedonia will lead to reduced pleasure experience (Heller et al. 2009). Clinical diagnosis of anhedonia does not discriminate between the lack of interest or pleasure. It is likely for

some one to give best effort even though he/she does not enjoy doing it, which is an example of low pleasure but still high motivation. For the case that one does not care to try if not enjoy it, it is lower pleasure and low motivation. With behavior observation showing reward impairment in depressed individuals, it is critical to examine how much of the difference in depressive behavior is caused by different goal setting that may be due to lack of pleasure, or by less willingness to spend effort to gain pleasure due to lack of motivation. If not clearly distinguishing the effects of goal setting and motivation, then mixing those two can lead to faulty interpretation of depressed individuals' reward evaluation. Inverse optimal control model provides an analytical way to address this problem, by defining goal as the 'intended state', while motivation as the 'amount of effort to achieve the goal state', thus can help to measure those two factors independently. We found as depression worsens, subjects had further goal distance from stop sign. It might indicate the tendency of risk-averse in depressed individuals, which is consistent with their higher sensitivity to avoid punishment (Dickson & MacLeod, 2004; Aldao et al. 2010). This finding may provide evidence of the decreased pleasure caused by increased depression severity. However, we only found significant lower motivation difference in severely depressed individuals (BDI>=29). This suggests that, for minimal-moderately depressed individuals, the behavioral difference may be mainly from goal setting, but not the amount of effort to achieve the goal. But for severely depressed individuals, they are different in both goal setting that may be caused by lack of pleasure, and in the amount of effort to achieve the goal that is caused by lack of motivation.

6. 6 Future work

Our current result suggests depressed individuals have further goal distance from the stop sign. It has several possible explanations. First, due to higher risk-aversion in depressed individuals (Smoski et al. 2008), it is likely that they prefer to stay further away to minimize the risk of crossing the stop sign. Secondly, as increasing goal distance will lead to higher action cost, it is also likely that they prefer a further distance to reduce action cost. In the future, we will use facial expressions recorded in the task to investigate this issue.

6.7 Acknowledgements:

Chapter 7, in full, is the material submitted to Cognition (brief article), Huang H, Movellan J, Paulus M & Harlé K.

6.8 Appendix:

Sensory speed γ and motor speed β

Task 1 (move-and-go) was designed to estimate sensory speed γ and motor speed β. We model subjects' perceived car position Y_t as a delayed true car position X_t due to the limit of sensory processing speed γ (appd: Eq.1). The higher the γ, the closer the perceived car position Y_t is to the true car position X_t . We assume subjects will decide the car starts moving once the perceived car position Y_t reaches a position threshold X_{thd}.

Thus the minimal time for the perceived car position Y_t to reach the threshold X_{thd} is reaction time RT (Equation 2):

Perceived car position Y_t : $dY_t = \gamma(X_t - Y_t)dt$ (1)

Reaction Time: $RT = argmin_t \{Y_t \geq X_{thd}\}$ (2)

We model joystick position C_t as a delayed execution from target joystick position *Utarget*, due to the limit of motor execution speed β (Equation 3). The higher the β, the closer joystick action is to the desired target position. Thus the minimal time for C_t to reach U_{target} is movement time (Equation 4).

Joystick position C_t : $dC_t = \beta(U_{target} - C_t)dt$ (3)

Movement Time: $MT = argmin_t \{C_t \geq U_{target}\}$ (4)

In above equations, X_t (true car position), RT (reaction time to car motion-onset), C_t (recorded joystick position), U_{target} (target position) and MT (movement time) are known. We use RT and X_t to recover X_{thd}, γ and Y_t, and use C_t and MT to recover β, by optimizing over γ, X_{thd}, and β to give the minimal errors between predicted RT , MT and observed data.

Inverse Linear Quadratic Gaussian Model (LQG)

We formulate the driving task as a LQG problem with a linear dynamic system and a quadratic reward function. In forward LQG problems, the optimal controller generates an optimal control policy that maximizes a given reward function. In inverse LQG problems, we use observed movements to infer the underlying reward function that best explains the observed behavior.

Linear dynamic system Assuming the driving task as a linear dynamic system (Equation

5) with a partial hidden state X_t and observable feedback Z_t, in which X_t is a 3x1 vector including the (hidden) true car distance to target stopping position at time t, joystick action at time t, and perceived car distance to target stopping position at time t.

Partial observable linear system: $dX_t = AX_t\,dt + BU_t\,dt$ (5)

Observation: $Z_t = CX_t + V_t$ (6)

With:

$$A = \begin{bmatrix} a & b & 0 \\ 0 & -\beta & 0 \\ \gamma & 0 & -\gamma \end{bmatrix} \quad (7)$$

$$B = \begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix} \quad (8)$$

$$C = [0, 0, 1] \quad (9)$$

In which, a, b are car dynamics parameters (assuming known), V_t is Gaussian noise, β and γ are motor and perceptual speed that are estimated from Task 1. Note that in the state X_t, the hidden true car position and perceived car position are measured as a distance to goal stopping position (parameterized as the goal state in the reward function), which we will estimate through MLE using observed behavior.

Quadratic reward function We assume the reward function r (X_t, U_t) is a function that evaluates the state X_t based on its distance from the goal state G (through $g(X_t, G)$), and the action U_t (through $U_t^2 q$).

reward function: $r(X_t, U_t) = g(X_t, G) - U_t^2 q$ (10)

Without loss of generality, let q = 1 (i.e. optimal action will not change if scaling

the reward function), thus $r(X_t,U_t)$ is a function of goal state G and motivation M, in which M is defined as the ratio of the distance between current state and the goal state over the energy expenditure. We assume subjects were using a stationary (infinite horizon) policy and the reward function has a diagonal form (i.e. no joint influence between state elements in the reward function).

In LQG setting, subjects first estimate true state from observation using a Kalman filter to convert the problem to a fully observable system, and then solve it as a LQR (Linear- Quadratic-Regulator) problem:

$$d\hat{X}_t = A\hat{X}_t dt + BU_t dt + L_t(Z_t - C\hat{X}_t)dt \qquad (11)$$

$$U_t = -K\hat{X}_t \qquad (12)$$

In which L_t is Kalman gain. U_t is a linear combination of the states and K can be estimated from U_t and recorded behavior data through linear regression. This suggests a quadratic value function:

$$v(\hat{x},t) = -\frac{1}{2}\hat{x}_t' w \hat{x}_t \qquad (13)$$

Then the HJB equation (Bellman, 1957) for this linear system will give as a quadratic reward function in the following form:

$$g(\hat{x}) = -\frac{1}{2}\hat{x}'(-2A'w + k'k)\hat{x} \qquad (14)$$

In which we define M as motivation:

$$g(\hat{x}) = -\frac{1}{2}\hat{x}' M \hat{x} \qquad (15)$$

$$M = -2A'w + k'k \qquad (16)$$

In which A and k are known from equation (7) and (12), and w can be solved by using optimal LQR solution.

References:

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review*, *30*(2), 217-237.

Blewett AE. Abnormal subjective time experience in depression. Br J Psychiatry. 1992; 161:195–200.

Caligiuri, M. P., & Ellwanger, J. (2000). Motor and cogni- tive aspects of motor retardation in depression. *Journal of Affective Disorders*, *57*(1-3), 83 - 93.

Davis, M. (1998). Are different parts of the extended amygdala involved in fear versus anxiety?. *Biological psychiatry*, *44*(12), 1239-1247.

Der-Avakian, A., & Markou, A. (2012, January). *The neu- robiology of anhedonia and other reward-related deficits* (Vol. 35) (No. 1). Elsevier Applied Science Publishing.

Dickson, J. M., & MacLeod, A. K. (2004). Approach and avoidance goals and plans: Their relationship to anxiety and depression. *Cognitive Therapy and Research*, *28*(3), 415-432.

Heller, A. S., Johnstone, T., Shackman, A. J., Light, S. N., Peterson, M. J., Kolden, G. G., ... & Davidson, R. J. (2009). Reduced capacity to sustain positive emotion in major depression reflects diminished maintenance of fronto-striatal brain activation. *Proceedings of the National Academy of Sciences*, *106*(52), 22445-22450.

Lemke MR, Puhl P, Koethe N, Winkler T. Psychomotor retardation and anhedonia in depression. Acta Psychiatr Scand. 1999; 99(4):252–6.

Mayberg HS, Lewis PJ, Regenold W, Wagner HN Jr. Paralimbic hypoperfusion in unipolar depression. J Nucl Med. 1994; 35(6):929–34.

Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., & Monteggia, L. M. (2002). Neurobiology of depression. *Neuron*, *34*(1), 13-25.

Ng, A. Y., & Russell, S. (2000). Algorithms for inverse rein-forcement learning. In *in proc. 17th international conf. on machine learning* (pp. 663–670). Morgan Kaufmann.

Shadmehr, R., & Krakauer, J. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research*, *185*(3), 359-381.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of behavior therapy and experimental psychiatry*,*39*(4), 567-576.

Sobin, C., & Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, *154*(1), 4-17.

Todorov, E., & Jordan, M. I. (2002, November). Optimalfeedback control as a theory of motor coordination. *NatNeurosci*, *5*(11), 1226–1235.

Treadway, M. T., Buckholtz, J. W., Schwartzman, N., Ashley, Lambert, W. E., & Zald, D. H. (2009, 08). Worth the eefrt? the effort expenditure for rewards task as an objective measure of motivation and anhedonia. *PLoS ONE*, *4*(8), e6598.

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, *31*(7), 1156 - 1168.

Chapter 7

Facial Expression of Depressed Individuals in a Simulated Driving Task

Abstract- Following last chapter, here we presented some of the preliminary analysis of facial expressions recorded in the simulated driving task. The goal of this project is to explore the possibility of using facial expression to measure subject's emotional state in the task, with the aim to explain the observed behavior and provide insights of the model parameters. Preliminary result suggests, comparing to healthy controls (BDI<=5), depressed individuals (BDI>=20) had significantly less joy expressions, both in mean evidence reported by FacetSDK (Emotient.com), and histogram distribution. They also have significantly more fear expressions in the histogram distribution. All together, these results indicate the validity of the hypothesis that further goal distance is caused by stronger risk-aversion in depressed individuals. More data and further analysis will be needed to test this hypothesis.

7.1 Introduction

As a mood disorder, emotional dysfunction is a hallmark of Major Depressive Disorder (MDD) (Rottenberg 2005; Morris et al. 2009) and it suggests that emotional

expression such as facial expression may be used as an effective measure of depression. Cohn et al. 2009 proposed to use facial actions in the diagnosis of depression (with the use of manual FACS, Ekman et al. 2002). Other studies using facial coding techniques found that MDD patients are less likely to react to positive stimuli with facial expressions associated with positive emotion (Gehricke & Shapiro 2000; Sloan et al. 2002). Reed et al. 2007 recorded MDD patients' facial expression in response to comedy and found that they showed reduced facial response to positive stimuli. Reed's finding suggests that MDD individuals may actively process information to create mood congruent responses. However, while past studies have been mainly focused on the detection of facial expression on positive or negative stimuli, we have little knowledge of their emotional state while performing a goal-directed motor task. For example, it has been shown that depressed individuals are more sensitive to punishment than to reward (Trew 2011). But so far no studies have provided evidence of the facial expressions to confirm this hypothesis. There are also studies shown depressed individuals are more risk-aversion (Smoski et al. 2008), which suggests fear expressions may be detected in those tasks, as it may drive risk-averse actions (Lerner & Keltner 2001). Thus it is important to assess, compared to healthy controls, if depressed individuals will have different emotional response based on their own risk assessment of the task. For example, if they will show less joy that is caused by depressed mood, and more fear that caused by risk-aversion. With advanced facial expression recognition technology (Emotient.com), we can now have automated face video analysis that gives reliable results at continuous time. Here, we propose to use a simulated driving task, and use the software developed by Emtient.com, to analyze the recorded face video.

7. 2 Method

We recorded subjects' facial expression in a simulated driving task (as described in previous chapter) using Logitech HD webcam C615 (30 fps) when subjects attempted to control the virtual car with the joystick. There are 19 healthy controls (BDI<=5; mean BDI = 2.26, std = 1.73) and 11 depressed subjects (BDI>=20; mean BDI = 27, std = 6.86) included in this analysis. Facial expressions were then analyzed from the video (~20 min) using FACET SDK (Littlewort et al. 2011; Malmir et al. 2013), developed by Emotient (Figure 1). FACET measures the emotional responses of users in continuous time, recognizing seven basic emotions: anger, contempt, disgust, fear, joy, sadness and surprise, and generating frame-by-frame outputs of those emotions.

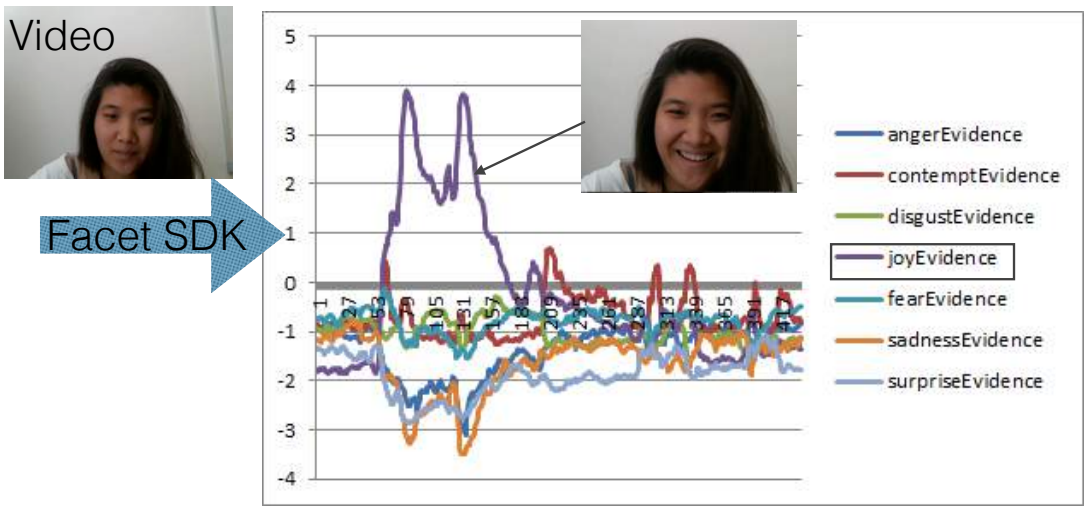

Figure 7.1: Example of processing video using Facet SDK

7. 3 Results

We first looked at the mean facial expression between Non-dep and Dep individuals (Figure 7.2). Result suggests depressed individuals had significantly less joy expressions (p= .006). No significant differences among other expressions were observed. Examples of Facet SDK output can be seen in Figure 7.3.

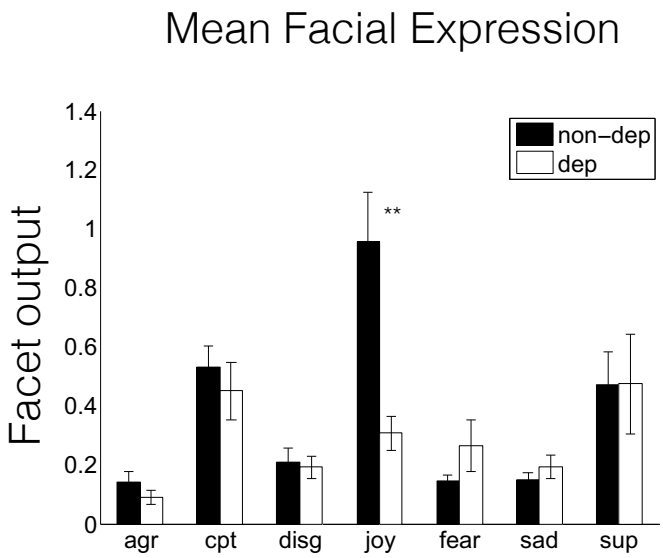

Figure 7.2: Mean facial expression of 7 basic emotions in Non-dep and Dep group.

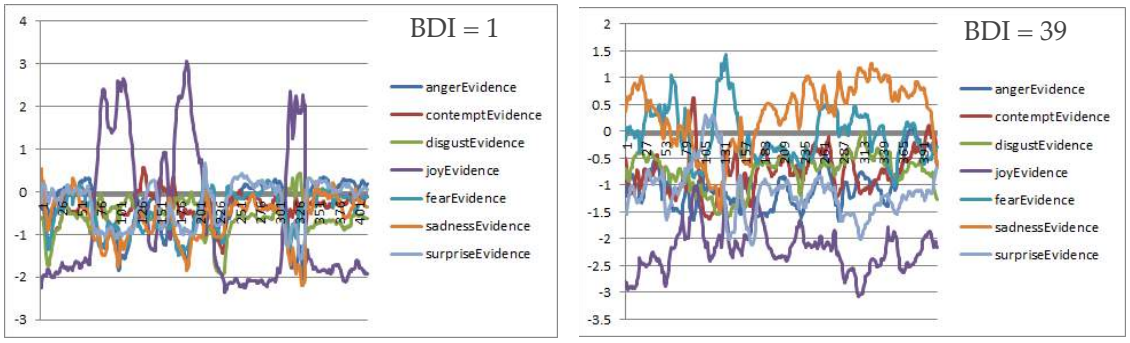

Figure 7.3 Examples of Facet SDK results for a non-dep individual (BDI =1) and a dep individual (BDI = 39).

We then looked at this histogram distribution of joy and fear expressions in those two groups. As shown in Figure 4, depressed group had significantly less joy expressions ($p < .001$) and more fear expressions ($p < .001$) than healthy controls in the task.

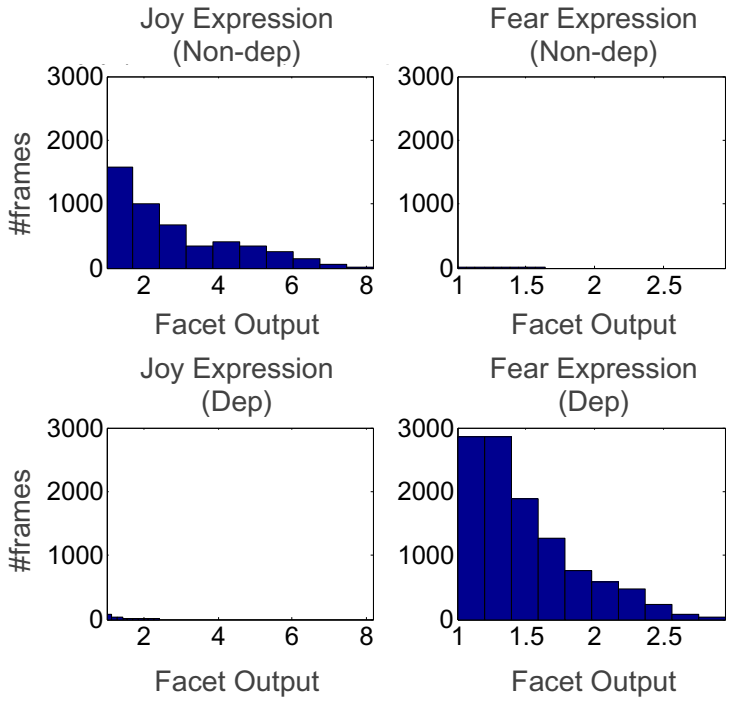

Figure 7.4: Histogram of Joy (Left) and Fear (Right) among Non-dep and Dep group

7.4 Discussion and Future work

In previous chapter, we found that depressed individuals have further goal distance from the stop sign, which can be explained by risk-aversion that caused by the fear of crossing stop sign, or the tendency of choosing a reward function that is associated with a lower action cost. From current facial expression result, in particular fear expressions, we believe there is a strong indication that the difference in goal setting is

caused by risk-aversion. However, more analysis needs to be performed to test this hypothesis. For example, if we can show a correlation between fear intensity and goal stopping distance. In addition, we will explore the dynamic of facial expressions at continuous time. For example, if depressed subjects show more fear when the car is approaching the stop sign.

References

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... & De la Torre, F. (2009, September). Detecting depression from facial actions and vocal prosody. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on* (pp. 1-7). IEEE.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facs manual. *A Human Face*.

Gehricke, J. G., & Shapiro, D. (2000). Reduced facial expression and social context in major depression: discrepancies between facial muscle activity and self-reported emotion. *Psychiatry Research*, *95*(2), 157-167.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology*, *81*(1), 146.

Morris, B. H., Bylsma, L. M., & Rottenberg, J. (2009). Does emotion predict the course of major depressive disorder? A review of prospective studies. *British Journal of Clinical Psychology*, *48*(3), 255-273.

Reed, L. I., Sayette, M. A., & Cohn, J. F. (2007). Impact of depression on response to comedy: a dynamic facial coding analysis. *Journal of abnormal psychology*, *116*(4), 804.

Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, *14*(3), 167-170.

Sloan, D. M., Bradley, M. M., Dimoulas, E., & Lang, P. J. (2002). Looking at facial expressions: Dysphoria and facial EMG. Biological Psychology, 60, 79 –90.

Smoski, M. J., Lynch, T. R., Rosenthal, M. Z., Cheavens, J. S., Chapman, A. L., & Krishnan, R. R. (2008). Decision-making and risk aversion among depressive adults. *Journal of behavior therapy and experimental psychiatry*,*39*(4), 567-576.

Trew, J. L. (2011). Exploring the roles of approach and avoidance in depression: An integrative model. *Clinical Psychology Review*, *31*(7), 1156 - 1168.

Conclusions

Part I: Inverse learning the underlying decision-process from observed choices in perceptual decision tasks.

In Chapter 2, we investigated if human subjects can learn spatial statistics and use this information in optimizing the search strategy in a visual-search task. Experimental results showed that subjects did internalize target spatial distribution and utilize them in deciding the search sequence. But they showed 'matching-like' behavior in their choices of 1st fixation, instead of 'maximizing', the optimal decision strategy. However, the debate between 'matching' vs. 'maximization' failed to take into account of the subjective belief of fluctuations in the underlying stimuli statistics. Thus we proposed to use a dynamic belief model (DBM) to capture this subjective belief into the learning process. We presented model prediction using DBM and FBM (fixed belief model) combined with two decision strategies (maximizing vs. a matching strategy) and a melioration strategy based on a limited trial history. Behavioral data with model comparisons (DBM + max, DBM + match, FBM + max, FBM + match, melioration) indicate that, even though subjects' choices of their first fixation appear to be matching-like behavior under a fixed belief model, it is best explained using a dynamic belief model with a maximizing decision strategy (DBM + max). First, DBM + maximization can explain the choices of 1st fixations cross trials within a block. Second, DBM + maximization can also explain how current fixation choices were influenced by recent trial history. Taken together, we have showed that, instead of interpreting 'matching-like'

choice behavior using the sub-optimal decision strategy (matching) with a fixed belief model, subjects' decision-process in fact can be better explained using the optimal decision strategy (maximization), with a dynamic belief model, which can best describe the learning process that is influenced from both long-term statistics and short-term statistics. In Chapter 3, we proposed to use DBM, combined with a DDM (Drift Diffusion Model) to explain sequential effects in a 2AFC task. We estimated subject-specific parameters of cross-trial learning from DBM and within-trial decision-making from DDM. We showed that comparing to FBM (Fixed Belief Model), DBM can produce a stronger sequential effect as observed in the data, and can capture individual differences in the belief of non-stationarity of the environment. All together, Chapter 2 and Chapter 3 showed that Dynamic Belief Model is a viable computational framework to explain human decision-making processes under uncertainty.

Part II: Inverse learning the objective function from movement trajectory in goal-directed motor tasks.

In Chapter 4, using optimal control framework, we proposed an Infomax model of human oculomotor control, under the assumption that eyes are not only the movement executor but also the information collector to achieve the goals. In addition, we should take into consideration of the sensory constraints of the eye, in which the observation is influenced by both target eccentricity and eye velocity. We showed that this model can explain saccadic eye velocity profiles, as well as smooth pursuit and eye movement in a rapid reaching task. In Chapter 5-6, we presented an inverse optimal control model to study depressive behavior in a simulated task, and showed that the observed behavior

difference in depressed individuals were caused by the joint effects from sensorimotor speed, goal-setting and motivation. In particular, we showed that for mild-moderately depressed individuals, they differ from healthy individuals in sensorimotor speed and goal setting, but for severely depressed individuals, in addition to slower sensorimotor speed and different goal setting, they also had significantly lower motivation from healthy individuals. Taken together, we showed that optimal control theory can be used to examine the underlying factors of different movement trajectories, and it can be applied to investigate the causes of different behavior observations.

In conclusion, observations from human behavior, either in the form of discrete decisions, or in the form of continuous movement trajectory, both describe how we use sensory information to achieve the goals under sensorimotor constraints. Thus to infer the underlying principles of observed behavior, we can examine what the goals that one is trying to achieve, how is the sensory information being used in decision-making/motor-control process, and what are the sensorimotor constraints in achieving the goal. We proposed to use two computational approaches (Bayesian inference and Optimal control theory) in this dissertation to address those issues, and showed that they can be applied in a variety of tasks to give insights of human sensory-motor processing.