Probabilistic Sensorimotor Processing:
Overcoming Uncertainty and Ambiguity to Improve Behavior

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Peter William Battaglia

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser: Paul R. Schrater
Co-adviser: Daniel J. Kersten

June 2008
Acknowledgements

This work was supported a UMN Graduate School Fellowship, NSF Graduate Research Fellowship, a UMN Doctoral Dissertation Fellowship, NIH R01 EY015261, Office of Naval Research Grant N00014-05-1-0124, EU grant 27141 "ImmerSence", SFB 550-A11, and the Max Planck Society.

First I would like to thank my thesis committee members, Al Yonas and Sheng He, for helpful discussions, thoughts, and criticism.

I thank Marc Ernst, Tonja Machulla, and Max Di Luca, at the Max Planck Institute in Tuebingen, who funded my visit to their lab and gave me a great opportunity to conduct the “size-change perception” experiment in this thesis.

Erik Schlicht allowed me to try out ideas on him and refined my thinking in many ways. He is a true friend and excellent scientist.

Robert Jacobs first introduced computational perception to me, deeply motivated my interest in the field, is largely responsible for inspiring me to pursue a doctoral degree, and believed in me more than I believed in myself.

Daniel Kersten’s wisdom and insight have touched every idea about perception and action I have considered in this thesis. He is a scientist’s scientist, whose logical rigor and balance of experimental and theoretical intuition are unparalleled. He was extremely generous with his time and energy, and provided innumerable contributions to my work.

Paul Schrater taught me more than I thought learnable, his creativity and brilliance have driven me to embrace ideas I would never have considered. His passion shaped my interests and I will forever value the deep commitment he made to training me as a scientist and the indelible mark he left on my consciousness in general. I owe him a tremendous debt of gratitude.

My parents have been the ultimate source of my motivation and drive. I respect every decision they made for me, despite often disapproving at the time. By invariably doing what is right, instead of what is easy, they taught me the most important lesson I have ever learned: the deepest source of reward is through hard work.
Dedication

This dissertation is dedicated to my grandparents, my parents, and Stefani.
Abstract

Perception and action are critical functions for most animals’ survivals, but to achieve high performance each must overcome uncertainty and ambiguity, both in perceiving the world and planning and executing actions. The following work examines how human behavior minimizes the negative consequences of uncertainty and ambiguity to improve performance in perceptually-guided tasks.

1) An object’s size and distance each influence the angular size of its visual image, rendering distance estimation based on angular size alone an ambiguous task. But, an “auxiliary” sensory measurement of the object’s size can disambiguate the distance. We ran a psychophysical experiment to test whether humans use sensed size when making distance judgments, and found significant improvements in distance perception due to the auxiliary size sensations.

2) Just as size and distance are each impossible to unambiguously estimate given angular size alone, judging the rate that an object’s size changes given only angular size-change rate is also ambiguous. But an auxiliary distance-change rate sensation can be used to disambiguate the size-change rate. We conducted an experiment to evaluate whether humans use the sensed distance-change rate when making size-change judgments, and found significant improvements in size-change perception due to the auxiliary distance-change sensations.

3) Many behavioral tasks have inherent time constraints that limit the time available for perception and action. We tested whether humans are able to adjust the time they devote to each to minimize the negative consequences of uncertainty that originates from perceptual and motor imprecision. We found humans’ timing choices are consistent with the theoretically ideal actor’s choices, and concluded that humans’ implicit knowledge of their perceptual and motor variability is used to make these choices.
Table of Contents

1. Introduction ............................................................................................................................................... 1
  1.1 Preface ............................................................................................................................................ 1
  1.2 Components of sensorimotor behavior ............................................................................................ 2
    1.3 Computational perception ............................................................................................................... 5
      1.3.1 Background .......................................................................................................................... 6
      1.3.2 Psychophysics and ideal observers ....................................................................................... 7
      1.3.3 Perception as Bayesian Inference ........................................................................................ . 9
      1.3.4 Cue combination ................................................................................................................. 12
      1.3.5 Prior knowledge .................................................................................................................. 15
      1.3.6 Internal generative models .................................................................................................. 17
      1.3.7 Evaluating Bayesian perceptual models ............................................................................. 19
    1.4 Computational sensorimotor control ............................................................................................ 20
      1.4.1 Background ......................................................................................................................... 21
      1.4.2 Internal models ............................................................................................................... .... 22
      1.4.3 Sensory feedback integration .............................................................................................. 23
      1.4.4 Separate roles of perception and action .............................................................................. 26
      1.4.5 Task-relevant behavior .................................................................................................. ..... 27
      1.4.6 Statistical decision theory ................................................................................................... 28
      1.4.7 Human sensorimotor decision-making ............................................................................... 30
      1.4.8 Sub-conclusion ................................................................................................................... 32
  1.5 Derivation of explaining-away rules ............................................................................................. 33

2. Humans Use Size cues to Improve Distance Perception ................................................................. 39
  2.1 Abstract ......................................................................................................................................... 39
  2.2 Introduction ................................................................................................................................... 39
    2.2.1 Background .......................................................................................................................... 39
    2.2.2 Experiment overview ............................................................................................................... 43
  2.3 Methods ......................................................................................................................................... 44
    2.3.1 Stimuli ................................................................................................................................. 44
    2.3.2 Trials ....................................................................................................................................... 47
    2.3.3 Exploration phase .................................................................................................................. 47
    2.3.4 Interception phase .................................................................................................................. 47
    2.3.5 Participants .......................................................................................................................... 49
3. Humans Use Distance-Change Cues to Improve Object Size-Change Perception ...............56

3.1 Abstract.............................................................................................................................56
3.2 Introduction........................................................................................................................57
3.3 Methods ................................................................................................................................60
  3.3.1 Participants ...................................................................................................................61
  3.3.2 Apparatus ....................................................................................................................61
  3.3.3 General procedure .......................................................................................................61
    3.3.3.1 Main experiment ..................................................................................................63
    3.3.3.2 Distance-Change Reliability Experiment .........................................................63
  3.3.4 Design ..........................................................................................................................64
    3.3.4.1 Main Experiment .................................................................................................64
    3.3.4.2 Distance-Change Reliability Experiment .........................................................66
  3.3.5 Data analysis ................................................................................................................66
    3.3.5.1 Main Experiment .................................................................................................66
    3.3.5.2 Distance-Change Reliability Experiment .........................................................70
3.4 Results ..................................................................................................................................70
  3.4.1 Main experiment ..........................................................................................................70
  3.4.2 Distance-Cue Reliability Experiment ..........................................................................78
3.5 Discussion ..........................................................................................................................80

4. Humans Balance Perceptual and Motor Uncertainty to Maximize Reach Accuracy ..........84

4.1 Abstract .............................................................................................................................84
4.2 Introduction ........................................................................................................................84
4.3 Methods .............................................................................................................................86
  4.3.1 Apparatus ....................................................................................................................86
  4.3.2 Target ...........................................................................................................................87
  4.3.3 Procedure: Experiment 1 ............................................................................................88
    4.3.3.1 Practice sessions .................................................................................................88
4.3.3.2 Combined test condition (CC) ................................................................................. 88
4.3.3.3 Visual baseline condition (VB) ................................................................................ 90
4.3.3.4 Motor baseline condition (MB) ............................................................................... 91
4.3.4 Procedure: Experiment 2 ............................................................................................. 92
4.3.5 Participants and compensation .................................................................................... 92
4.3.6 Model.............................................................................................................................. 92
   4.3.6.1 Visual and motor accuracy tradeoff ......................................................................... 93
   4.3.6.2 Ideal reacher derivation ........................................................................................... 94
4.4 Results ........................................................................................................................................... 97
   4.4.1 Experiment 1 .................................................................................................................. 97
      4.4.1.1 Baseline conditions .................................................................................................. 97
      4.4.1.2 CC condition: timing choices ................................................................................ 99
      4.4.1.3 CC condition: task performance .......................................................................... 105
   4.4.2 Experiment 2 ................................................................................................................ 106
4.5 Discussion ................................................................................................................................. 109
5. APPENDIX A .............................................................................................................................. 114
   5.1 Visual variability model ................................................................................................. 114
   5.2 Motor variability model ................................................................................................. 115
   5.3 Maximum likelihood estimation of model parameters .................................................. 116
   5.4 Bootstrapped confidence intervals/SEMs ................................................................. 116
6. Bibliography ............................................................................................................................... 117
List of Tables

Table 4.1 ................................................................................................................................. 102
List of Figures

Figure 1.1 : Object generates image. ................................................................. 10
Figure 1.2 : Cue combination. ........................................................................ 13
Figure 1.3 : Auxiliary information “explains away” ambiguity. ....................... 18
Figure 1.4 : Explaining away Bayes’ net. ..................................................... 34
Figure 2.1 : Size and distance ambiguity. ...................................................... 40
Figure 2.2 : Sensory generative process and scene property inference. ........ 41
Figure 2.3 : Virtual workbench. ................................................................... 45
Figure 2.4 : Stimulus screenshot. ................................................................. 46
Figure 2.5 : Stimulus scene diagram. ............................................................ 48
Figure 2.6 : Size dependence of judged distances. ........................................ 51
Figure 2.7 : Size dependence reduction. ...................................................... 52
Figure 2.8 : Judged distance accuracy. .......................................................... 53
Figure 3.1 : Sensation generation and scene inference. .................................. 59
Figure 3.2 : Size- and distance-change rates schematic. ............................... 62
Figure 3.3 : All trials’ size- and distance-change rates. ............................... 65
Figure 3.4 : Psychometric functions for all distance-cue conditions. .......... 72
Figure 3.5 : 3D psychometric functions for all distance-cue conditions (participant 3). 73
Figure 3.6 : Discrimination boundaries for all distance-cue conditions (all participants). ...... 75
Figure 3.7 : Main experiment PCSS bias, all conditions (all participants). .... 76
Figure 3.8 : Distance-cue bias reduction (all participants). ........................... 77
Figure 3.9 : Distance-Change Reliability Experiment distance cue noise std. dev. 79
Figure 3.10 : Noise magnitudes differences versus PCSS bias differences (all participants). 80
Figure 4.1 : Experimental diagram. ............................................................. 87
Figure 4.2 : Visual, motor, and predicted CC performance vs. elapsed time (participant 4). 93
Figure 4.3 : Dot centroid estimation noise (all participants). ........................ 98
Figure 4.4 : \( \sigma_c \) and measured \[ t_v, t_M \] choices (participant 4). ............. 100
Figure 4.5 : Timing choices and their corresponding \( t_C \) values (participant 1). 101
Figure 4.6 : Regression slopes of \( t_C \) values across dot scatter levels (all participants). 102
Figure 4.7 : Measured vs. predicted \( t_C \) means (all participants). ................. 103
Figure 4.8 : Measured vs. predicted offset standard deviations (all participants). 104
Figure 4.9 : Measured reach endpoint accuracy with respect to predictions (all participants). 106
Figure 4.10 : CC task performance comparison between Exp. 1 and 2 (all Exp. 2 participants). .... 107
Figure 4.11 : Timing distributional area comparison between Exp. 1 and 2 (all Exp. 2 participants) .... 108
1. Introduction

1.1 Preface

Perceptually-guided actions are critically important for animals to complete the tasks required for survival. In primates, perceptual and motor processing regions compose more than half of the brain’s cortex. Neural perceptual processing systems collect sensory data and interpret the states of the body and environment. Motor systems issue commands that instruct the muscles how to execute actions. These processes are highly interdependent: actions may change the content and quality of sensory input, and perceptual interpretations determine what actions are required for particular tasks. Each faces significant challenges, such as ambiguity and uncertainty, yet still exhibit high spatial and temporal precisions and allow intricate manipulations of the body and environment. The importance and high performance of perceptually-guided actions suggest they are facilitated by specialized brain mechanisms that carry out sophisticated computations.

The cooperative functionality of sensation, perception, and action that serve an animal’s goals are termed sensorimotor behavior. Sensorimotor behavior involves decisions about how to deploy the body’s physical capabilities effectively to accomplish different tasks. This requires the brain to consider the task goals, the current states of the body and environment, and the dynamics of the body, including musculoskeletal and neurophysiological properties. Sensory input and prior knowledge provide necessary information about task goals and body and environmental states. This information is interpreted, stored, and relayed to brain systems responsible for motor planning, which select motor commands likely to accomplish the task goals given the body’s dynamics. Sensorimotor behavior requires internal representations of sensory and motor information and efficient computations to exploit the information.

Consider the example of lifting a glass of water to appreciate the complexity of seemingly trivial sensorimotor behaviors. Although a child can perform this task with relative ease, no computer-controlled robot can rival this feat. Reasons for this disparity
are neither mechanical imprecision in the robotic device, nor insufficient number-crunching power, both for which machines match, if not vastly exceed, animals. The disparity exists because machines lack the perceptual and motor-planning capabilities of biological brains.

One goal of sensorimotor research is to uncover the perception and motor-planning principles that serve sensorimotor behavior. The computational modeling approach treats the animal as a black box that takes sensory data and tasks as input, and produces goal-directed behavior as output. Measuring human and animal sensorimotor behavior in experimental settings offers an opportunity to quantify those inputs and outputs, and test hypotheses about the brain functions responsible for the measurements. An important aspect of computational modeling is determining how sensory information, beliefs, and expected action outcomes are represented in the brain’s neural architecture and transmitted among the specialized processing subsystems.

As mentioned, a difficult challenge faced by perception and action systems is to overcome ambiguity and uncertainty, both in sensory input and motor output. The experiments conducted as part of this dissertation examine how human behavior minimizes the negative consequences of ambiguity and uncertainty to improve performance in perceptually-guided tasks. At the beginning of each of the following experimental chapters, a focused literature survey specific to that respective experiment will be presented. The remainder of this chapter will preface those experiments by discussing a broader set of literature related to how the brain represents and computes information about the body and environment states, the task, and the body’s dynamics to facilitate the production of sensorimotor behavior. Because sensorimotor behavior has a fundamentally non-deterministic nature, this chapter will also review the probabilistic structure of information representations and how such representations serve the sensorimotor system. The final section of this chapter will present a theoretical discussion of how an observer can optimally infer multiple unobservable world properties from the observable sensations they generate.

1.2 Components of sensorimotor behavior
Sensorimotor behavior is composed of task specification, information-gathering, and action components, and thus fundamentally constitutes a decision-making process. The task is a critical component of intentional sensorimotor behavior. Some examples include lifting a glass of water, moving from one point to another, or avoiding contact with a moving vehicle. Tasks usually have many potential outcomes, and outcomes have rewards or penalties associated with them. *Utility* is a measure used by economists to quantify the satisfaction or subjective value attached to the rewards and penalties (Berger, 1985). Normative decision theory prescribes an actor to survey the possible outcomes and their associated utilities, then select the outcome with the highest expected utility as the task goal.

A second component of sensorimotor behavior is gathering information about the current states of the body and environment. In order to perform an action, the brain must have knowledge of the spatial layout and physical properties of the surroundings, and also the body’s position, velocity, posture, etc. For instance, when reaching to pick up a glass of water, the environmental state includes the table surface’s angle and position, and the glass’s location, size, and weight. The body state includes the current hand position and velocity, and the current aperture between the fingertips and thumb that will be used to grip the glass. In order to use information about body and environmental states, the brain must store state variables in a form that supports integration with task goals and known body dynamics.

A third component of sensorimotor behavior involves actions, which are selected based on the task goal, knowledge of the body’s dynamics, and possible actions’ associated costs. The muscles receive motor commands and apply forces to manipulate the body and the environment, and because the skeletal structure, muscles, skin, etc. determine the outcome of the motor commands, the brain must select those commands that are expected to achieve the desired outcome given these elements. Often the intended movement differs from the actual movement; such variability is termed *motor noise*. It is important for motor plans to take motor noise into account so that this potential source of error can be managed. Also, different actions have different costs: you do not knock down a brick wall to find out what is behind it, you simply walk around
it. Though both actions result in the same outcome, their different costs make one action more attractive than the other. The effect of a motor command, and the cost associated with its execution may influence the appropriate action to take to achieve a desired outcome. There have been various theories regarding what the brain knows about the body’s dynamics and how costs are accounted for and incorporated into sensorimotor decision-making.

Another less-cited component of sensorimotor behavior involves the dynamics of sensory information. In some situations the quality or general nature of sensory information varies with time, body posture, position, etc. For example, imagine walking into a house on a bright, sunny day when your eyes are still adapted to the broad contrast range of the light outside. You may pause before doing anything because you know your general perceptual quality will improve over the next few moments as your eyes adapt to a finer contrast range. Sometimes one’s own actions cause variation in the quality of sensory information, such as when making eye movements.

To provide an intuitive feel for a sensorimotor behavior, let’s once again use the example of lifting a glass of water, this time using the components from the previous few paragraphs. First, a person perceives thirst. This constitutes a task, where successfully drinking water produces a utility gain and not drinking any water produces either no utility change (remaining thirsty), or even a utility penalty (dehydration). Thus the task goal is to successfully drink a glass of water because this is expected to yield the highest utility. To serve this task goal, perceptual systems infer the environmental and body states; visual information is processed to identify and localize the glass and table, visual and proprioceptive information (including felt joint angles and muscle extensions/contractions) are integrated to infer the body’s current location and posture. The resulting perceptual interpretation might suggest that the glass is one meter away on a table, and the body is sitting in a chair with the hands at the sides. That information is transmitted to motor planning systems which use knowledge of the body’s dynamics to select a sequence of motor commands for the muscles. Those motor commands are predicted to cause the right arm to raise, extend toward the glass, close the fingertips around the glass, lift the glass, return toward the mouth, and tip the glass to deposit the
water into the mouth. Simultaneously the head is positioned to allow the hand to approach the mouth directly, and the mouth itself opens to form a seal with the glass and receive the water. The motor commands are issued, a series of actions is executed, and an outcome occurs. The new state is then analyzed by perceptual systems to confirm that the water was consumed. If indeed it was, the task is no longer valid because the initial motivating condition (thirst) is no longer present. If the action had resulted in the glass being knocked over, the water missing the mouth and landing on the shirt, or some other mistake, the task would be updated and continued (i.e. dry up the water, get a new glass of water, and so on). Although the example is common and simple in comparison to many tasks, it illustrates the multiple concerted activities that the brain orchestrates to produce a basic sensorimotor behavior. Subjectively the behavior seems effortless, but the sequence of perceptual and motor processes is non-trivial and requires sophisticated, efficient information representations and computations to accomplish.

The following two sections of this introductory chapter compose a literature review with two foci: research that investigates computational perception exclusively and research that studies computational sensorimotor control. The studies are not an exhaustive compilation of perceptual and sensorimotor research, but are intended to provide necessary information to understand the important computational problems of perceptual inference and motor planning. The fourth and final section will present a theoretical treatment of probabilistic perceptual inference of multiple scene properties based on multiple observations.

1.3 Computational perception

Perception is the process by which sensory data and prior knowledge are used to generate beliefs about environment and body states. For instance, when viewing a banana, the luminance and color information is transduced into a neural signal by the retina, and sent to the visual cortex of the brain. The brain uses the evidence contained in the signal to compute beliefs about the identity and spatial properties of the object in the scene (the banana) that caused the signal. Within the signal that arrives at the visual cortex, there is information about the banana, information about other nearby objects, and
extraneous information generated by unpredictable processes such as optical irregularities in the eye’s lens, or neural noise during transduction and transmission from retina to visual cortex. Green and Swets (1966) were among the first to quantify the process of extracting signals and discarding noise from stimuli. Even when a pure, noiseless signal can be obtained, perceptual processing must overcome obstacles such as signal redundancy (Attneave, 1954), incomplete information about the scene attribute(s) in question, and ambiguity about the cause of the stimulus. Obtaining high-quality beliefs about a scene can be aided by combining multiple sensations with prior beliefs, and generating interpretations based on knowledge of how sensations are related to their causes in the world.

1.3.1 Background

Attneave (1954) reported that Shannon’s (1951; Shannon and Weaver, 1949) ideas about signal coding had applications to visual perception. Specifically, visual signals are spatially and temporally redundant; a simple example is that measuring the response of a retinal photoreceptor at a certain position and time allows greater-than-chance prediction of proximal photoreceptor responses at, just before, or just after the measurement time. Further, given the vast number of photoreceptors and their possible states (firing/not firing), it is computationally intractable for the brain to examine and represent all information in non-redundant signals, thus redundancy is a critical feature that allows visual perception to avoid unnecessary computation. Barlow (1961) extended this idea by proposing that an organizational principle of biological visual systems is to reduce redundancy in the internal representation of visual scenes. The logical end of redundancy reduction is a meaningful representation, perhaps composed of identified “objects” with spatial properties. More recently Barlow (2001) concedes that the role of sensory processing is not exclusively to reduce redundancy, but this does not diminish the value of statistical models to describe perceptual processing.

Helmholtz’s (1867) characterized perception as a process of “unconscious inference”, in which sensations are inverted using knowledge of how they are generated to recover the properties of the scene. Statistical inference that incorporates sensory
generative knowledge and prior information falls under the broad category of probabilistic Bayesian models of perception (Knill and Richards, 1996). Several lines of research have provided substantial support for the idea that sensory processing reflects Bayesian inference. But generally, many proposed perception models are consistent with Bayesian perception. A Bayesian model provides a cohesive, principled framework in which to describe the process of drawing beliefs from sensory data and prior knowledge, adding explanatory power to empirical behavioral observations and mechanistic features (receptive fields, neural firing patterns, etc.).

The details of Bayesian models of perception and research supporting them will be discussed after the following methodological overview.

1.3.2 Psychophysics and ideal observers

A common experimental method for studying biological perception is called psychophysics. A psychophysics experiment presents an observer with a sensory stimulus and asks the observer to make a decision about the stimulus. Psychophysics experiments measure the upper limit of perceptual performance. An example of a psychophysics experiment (Hecht et al., 1942) is: I present a very dim flash of light to you, and ask you to press a button if you see it. I can vary the luminance of the flash of light, and repeat this trial many times. Based on your responses, I can estimate what minimal amount of light is required for you to detect it.

A powerful tool for perceptual research is ideal observer analysis (Geisler, 2003). Generally an observer takes sensory data and prior knowledge as input, and outputs task behavior; an ideal observer is a computational model for which the output behavior is optimal given task constraints. Thus the ideal observer is, by definition, the upper bound on task performance given the inputs. For example, when confronted with two objects, successful task performance in an object identification task is associated with correctly identifying each object while success in a discrimination task is associated with correctly classifying the objects as “same” or “different”.

An ideal observer may make mistakes, but this is due to noise or ambiguity in sensory input, imperfect prior knowledge, or uncertainty imposed on the model’s
computations that reflect constraints on task performance (Geisler, 2003). When constructing an ideal observer model, assumptions about noise or inherent limitations are formally specified, and performance is predicted on the basis of those constraints. Ideal observer performance and psychophysiologically-measured human or animal performance can be compared: all real observers, human or animal, must perform within the bounds of ideal performance and chance performance (defined as randomly selecting behavioral output while ignoring sensory data and prior knowledge completely). Because of these bounds, a real observer’s performance may be described by an efficiency score, which quantifies “how optimal” the observer is. An efficiency score lies between 0% and 100%, corresponding to chance and ideal performances, respectively. This may reflect what proportion of the total available information and prior knowledge the observer actually based the behavioral decision on, or how well the observer uses the information. The efficiency score provides a dimension on which to compare different potential observer models to the real observer. Any potential model whose efficiency score is less than the real observer’s may be ruled out as a description of real behavior because it fails to operate at the level of the real observer. If a potential model outperforms the real observer it must be sensitive to more of the sensory information, or use the information more efficiently, than the real observer. Barlow (1981) discusses physical limitations that govern visual perception, and suggests that the brain’s functionality is proportional to the extent of those limitations. Ideal observer analysis is essentially the inverse: using functional bounds to constrain hypothesized physical, mechanistic neural models. In addition to supporting and rejecting models, ideal observer analysis is helpful for formally quantifying sensory data, prior knowledge, and task constraints, all of which must be internally represented in real observers for performance to approach optimality.

Formally, an ideal observer makes a decision based on knowledge of model parameters and a loss function, and this decision causes an outcome (either deterministically or probabilistically). In perception, model parameters are variables that represent the state of the body and environment; knowledge of parameters may be accumulated from sensory data and prior knowledge, and take the form of probability distributions over possible parameter values. The loss function defines the rewards and
penalties associated with different outcomes, as determined by the task. As mentioned, the decision may probabilistically determine the outcome, meaning there may be outcome uncertainty for different decisions. But for perception independent of action, it is appropriate to consider decisions and outcomes to be equivalent because perceptual states are believed to be equally easy to decide upon (Maloney, 2002). Cases in which decisions and outcomes are not be equivalent frequently occur in motor control; for instance when hitting a baseball with a bat, the decision may be to strike the ball with the bat but the inherent variability among swing outcomes may cause some swings to miss the ball.

When quantifying maximum task performance using ideal observer analysis, it is often useful to incorporate known constraints into the ideal observer model (Geisler, 2003). This idea can be illustrated using an example from visual perception. An overly simple ideal observer may consider the image of a scene projected on a 2D surface, much like a camera’s photograph, to be the input sensory data. This ideal observer is unrealistic because it does not precisely correspond to biological optics. The eye is stimulated by light that passes through the eye’s optics, which contain irregularities that absorb photons, distort the light, and generally add random noise to the image that falls on the retina. Naturally a real observer’s performance could never match the overly simple ideal observer’s, because even if every other aspect of the real observer’s behavior matched ideal functionality, the variability introduced by the optical noise would prevent the real observer from matching the ideal. A more realistic ideal observer would take this inherent random noise into account, and make decisions based on this more biologically-plausible model.

1.3.3 Perception as Bayesian Inference

Bayesian inference is a probabilistic framework in which measurements are used to compute beliefs about what caused them. Bayesian inference is well-suited to the problem perceptual systems face because the state of the world is hidden, but it generates sensations that carry information about it. Bayesian inference provides an optimal, prescriptive solution to perception that is valid if world states are assumed to be random
variables, probability is assumed to represent degree of belief about their values, and the observer has knowledge of the true probability distribution over world states and the probabilistic conditional dependencies between world states and sensory measurements. For a particular perceptual context, Bayesian inference provides a functional relationship between sensory input and belief about the world. A growing consensus of empirical results confirms human perception yields beliefs consistent with Bayesian predictions (Kersten et al., 2004).

Bayesian inference is used to compute the probability of the value of a parameter that is not directly measurable, but which generates directly measurable sensory data. Since ideal observers base their decisions on probability distributions over parameter values, Bayesian inference prescribes optimal decision-making strategies when there is uncertainty about any parameter value that is relevant for task performance. For example, consider the simple task in which we are trying to estimate an object’s identity from an image, with minimal chance of error. In Figure 1.1, the circle (called a node) labeled “O” is a parameter that represents the object’s identity, which is not directly measurable. The node labeled “I” is a parameter that represents the information contained in the image of the object, which is directly measurable. The arrow connecting the nodes represents probabilistic dependence: the image of the object (“I”) looks the way it does as a random function of the identity of the object (“O”).

Figure 1.1: Object generates image.
This is a simple graphical model that characterizes an object in the world generating an image. The node labeled “O” represents the object’s identity, the node labeled “I” represents the image information, and the arrow represents the generative process of image formation that causes the image to appear the way it does depending on the object’s identity.
For instance, if the object is a baseball, most likely the image will contain much white color information, although it is possible that the baseball is dirty so the image instead contains brown color information. Using probability notation our problem can be formalized as such: we want to find the posterior probability of “O” given “I”, \( p(O \mid I) \), so that we can select the object identity with the greatest probability as our estimate. Bayes’ rule is a relatively straightforward consequence of probability theory regarding joint and conditional probability distributions:

\[
\begin{align*}
    p(O, I) &= p(I, O) \\
    p(O \mid I)p(I) &= p(I \mid O)p(O) \quad \text{Equation 1.1} \\
    p(O \mid I) &= \frac{p(I \mid O)p(O)}{p(I)}
\end{align*}
\]

Here \( p(I \mid O) \) is termed the conditional likelihood, or generative model, and represents the relationship between the object’s identity and the particular image information it generates, i.e. a clean baseball generates much white color information, a dirty baseball generates much brown. \( p(O) \) is called the prior probability and represents the probability of different objects being present in a scene before observing it, i.e. the chances of stumbling upon a white baseball versus a dirty baseball. \( p(I) \) is called the total likelihood and represents the probability of the measured image occurring across all possible scenes, i.e. in nature it is common for an image to contain brown color information but uncommon for an image to contain fluorescent pink color information.

Bayesian inference is the optimal model for computing the probability of parameter values using measurements. Often applied in ideal observer analysis, a Bayesian model of a perceptual process can be used as a methodological tool for suboptimal model rejection. If a real observer’s performance matches optimal Bayesian predictions, this does not necessarily imply perception stores probability distributions and computes posterior beliefs, only that the computations the brain is performing are consistent with those a Bayesian observer would use. Moreover, Bayesian models do not specify physiological mechanisms for perceptual inference; instead they define
theoretical bounds on task performance and imply computational mechanisms. Because recent evidence supports near-optimal performance for a variety of perceptual tasks, researchers have begun searching for the brain’s computational mechanisms responsible for this behavior. Specific hypothesized components include representations of the statistics of sensory information, prior knowledge of scene attribute statistics, and generative models for sensory data. The following sub-sections present work that has investigated how such components play a role in perception.

1.3.4 Cue combination

Strong confirmation that the brain represents the statistics of perceptual estimates comes from research that studies how the brain combines multiples pieces of information about single scene attributes (see Ernst and Bülthoff, 2004 for additional review). Pieces of sensory information generated by a scene are termed cues; integrating the information available in multiple cues is termed cue combination, or cue integration. Clark and Yuille (1990) defined “weak fusion” and “strong fusion” as distinct cue combination methods. A weak fusion observer computes a separate estimate of the scene attribute from each cue, and then combines those estimates as a linearly-weighted average with weights proportional to the estimates’ respective reliabilities. A strong fusion observer computes one estimate of the scene attribute from all cues taken together, which allows nonlinear cue combination. Generally, weak fusion observers correspond to approximately optimal Bayesian estimators in the case that the cues are conditionally independent given the scene attribute (Yuille and Bülthoff, 1996), which occurs when variability in one cue is unrelated to variability in another. Other cases require a strong fusion observer for optimal Bayesian estimation of the scene attribute. The generation of shading and texture cues to surface slant cannot usually be considered independent processes, because there are auxiliary factors like global luminance that may contribute to the formation of both (Yuille and Bülthoff, 1996). In these cases, the optimal observer must use cues in a nonlinear way to infer scene attributes.
A common method of studying how multiple cues are combined to estimate a scene attribute is called *perturbation analysis* (Young et al., 1993; Landy et al., 1995; Knill, 1998b). In perturbation analysis, two cues dependent on some scene attribute are presented simultaneously but each cue indicates a different value of the scene attribute. By examining an observer’s judgment of the value of the scene attribute, one can infer the relative contribution of each cue’s information to the final estimate. Consider the following simple example of perturbation analysis to study linear cue combination. For a task in which an observer is asked to estimate a surface’s slant angle, an experimenter might present stereoscopic and texture compression cues that suggest surface angles of 40 and 48 degrees, respectively. If the observer judges the surface angle to be 44 degrees, the simplest explanation for the judgment is that the individual cue estimates are being averaged with equal weights. If the observer judges the angle to be 42 degrees, the individual cue estimates are being averaged with weights of 0.75 and 0.25 for stereo and texture cues, respectively.

A Bayesian model of perception specifies that sensory information is applied toward perceptual judgments in proportion to its reliability. It is important here to highlight a distinction: there are models of perception that posit the use of sensory information’s reliability to compute perceptual estimates and there are models that go a
step further and propose that many perceptual computations are optimal estimates, which implicitly necessitate the use of the sensory information’s reliability. In order to prove that a reliability-contingent perceptual computation is optimal, it must be shown that the final estimate and its respective uncertainty are comparable to the ideal observer’s. Among the following experiments, those experiments that explicitly tested for optimal estimates will be noted. See Jacobs (2002) for a discussion of what factors may determine visual cue reliability and what role reliability plays in cue combination. See Schrater and Kersten (2000) for a detailed examination of perceptual cue combination in the context of Bayesian decision-making.

Several experiments tested perception of visual object geometry and found reliability-weighted cue combination. Young et al. (1993) reported that observers who are asked to discriminate between the degree of surface curvature of elliptical cross-section cylinders using motion (kinetic depth effect) and texture (compression due to perspective scaling) cues compute separate curvature estimates and combine them linearly according to their respective reliabilities, supporting a weak fusion observer model. Jacobs (1999) followed this result to report that such motion and texture cue combination was optimal. Johnston et al. (1994) reported evidence for a weak fusion model of curvature perception based on stereo (stereopsis information indicating by binocular disparities) and motion cues. Several studies reported weighted combination for slant (Backus et al., 1999) and texture edge perception (Landy and Kojima, 2001). Knill (1998a; 1998b; 1998c) found that human surface slant estimates are weighted combinations of several textural cues. In particular, humans rely heavily on textural foreshortening cues and assume isotropic “texel” distribution when stimulus information is consistent with such an assumption, which may reflect perceptual exploitation of the most robust components of available textural information. Knill and Saunders (2003) and Hillis et al. (2004) each reported that stereo and texture cues to surface slant are combined optimally.

There have been several studies that investigated perceptual estimates of scene attributes based on cues from different modalities. Ghahramani et al. (1997) reported reliability-weighted linear combination of visual and auditory information for perceptual
estimates of target location, while Battaglia et al. (2003) and Alais and Burr (2004) each reported optimal cue combination of visual and auditory cues to target location. Körding et al. (2007) showed visual and auditory cues to an object’s location are only combined when they are likely to have been caused by the same source, in accordance with a causal inference Bayesian model of perception. Shams et al. (2005) reported that the visual illusion of multiple flashes of light induced by auditory beeps (Shams et al., 2000; Shams et al., 2002) is an optimal percept based on the reliability of the visual and auditory cues.

Van Beers et al. (1999) found evidence for reliability-weighted linear combination of proprioceptive and visual cues to fingertip location. Violentyev et al. (2005) reported a haptic-induced visual illusion analogous to the auditory-induced visual illusion discovered by Shams et al. (2000). Ernst and Banks (2002) reported that humans combine visual and haptic cues to an object’s size optimally to compute a final perceptual estimate. Helbig and Ernst (2007) reported optimal visual and haptic shape cue integration. Hillis et al. (2002) reported that when cue combination occurs among only visual cues, the individual information from each cue is discarded so that only the final visual estimate remains. But conversely, when a visual cue and a haptic cue are combined, the individual cue information is retained along with the final estimate which allows discrimination among stimuli whose final perceptual estimates are the same, but whose underlying composition of cues is different.

1.3.5 Prior knowledge

Another feature of Bayesian perception is the incorporation of prior knowledge to improve judgments. Specifically, sensory data is combined with prior knowledge in order to compute a posterior probability distribution over the value of a scene attribute. In some cases, prior knowledge may behave like another independent sensory cue, providing information about the desired scene attribute. In other cases prior knowledge is used to disambiguate an otherwise ambiguous cue. The following discussion considers prior knowledge as pre-sensation beliefs about scene properties; an alternative use is to treat all knowledge not contained within sensory input as prior knowledge. Only a few experiments have documented human use of prior knowledge because of some inherent
difficulties. First, many scene attributes vary greatly, like ambient illumination and object size, thus an appropriate prior distribution over possible attributes would be nearly uniform thus rendering its effects on perceptual estimates comparable to using no prior knowledge at all, and thus experimentally indistinguishable from prior non-use. Second, context is thought to play a role in prior knowledge’s application, thus measuring the prior is very sensitive to experimental conditions and real-world priors may not readily transfer into experimental settings. Third, the effect of prior knowledge may be scientifically categorized as “bias” in some cases. In order to prove that prior knowledge is being used in a Bayesian manner it needs to be shown that the use of the prior knowledge increases the quality of resulting perceptual estimates, as opposed to bias which does not necessarily improve estimation. Despite these challenges, the following experiments documented human use of prior knowledge for perception.

Mamassian and Goutcher (2001) measured human observers’ estimates of an object’s shape from shading cues which require an assumption about what direction the light arrives from to be useful. This assumption about the light direction is a kind of prior that serves as auxiliary knowledge for use in disambiguating the otherwise ambiguous shading cue. Mamassian and Landy (2001) investigated how multiple priors’ weights are decided by the brain, specifically lighting direction and surface slant priors, and concluded that the weights reflect their relative reliabilities. Adams et al. (2004) modified observers’ light direction priors by providing haptic feedback that suggested a different lighting direction than the default. Weiss et al. (2002) modeled a variety of visual motion illusions as optimal percepts based on observers’ uses of object motion prior knowledge. Körding et al. (2004) showed that when observers estimate a force pulse’s magnitude, they form their estimates by integrating current sensory information and previous experience with force pulses, which constituted empirical prior knowledge. Schwartz et al. (2005) use a smooth contour prior to explain various line tilt judgment “misperceptions” as optimal inferences. Tassinari et al. (2006) reported that humans use empirical prior knowledge provided by training experience in combination with sensory information to estimate reach target locations.
1.3.6 Internal generative models

A core component of Bayes’ rule is inverting knowledge of how scene properties generate sensory data to develop beliefs about those scene properties. Treating perception as Bayesian inference requires the brain to incorporate a model of how scene attributes generate sensory data, called the generative model or conditional likelihood function (Knill and Richards, 1996; Kersten et al., 2004). Freeman (1994) framed the “generic viewpoint assumption” (Malik, 1987; Nakayama and Shimojo, 1992) as a consequence of Bayesian inference using sensory generative knowledge for shape perception. Körding and Wolpert (2004) reported that humans who were asked to reach to targets under impoverished visual feedback of their fingertip positions learned a relationship between visual information and true fingertip position, and used it to improve reach accuracy beyond the level of the sensory information alone. Stocker and Simoncelli (2006) show observers’ motion judgments are consequences of internal knowledge of sensory noise and prior assumptions of slow movement.

A distinct feature of Bayesian inference is termed explaining-away (Kim and Pearl, 1983; Pearl, 1988), in which an ambiguous sensory measurement is disambiguated by an additional, auxiliary cue that does not depend on the scene property being estimated. The following details a qualitative account of perceptual explaining-away.

Knill and Kersten (1991) reported that an object’s shape affects observers’ judgments of surface albedo, consistent with a generative model that explains away the effect of the shape on the image luminance. Figure 1.3A and Figure 1.3B illustrate Knill and Kersten’s (1991) instance in which an internal generative model allows shape to disambiguate an otherwise ambiguous luminance cue to the albedo of a surface (adapted from Knill and Kersten, 1991, with permission). In Figure 1.3A, the grayscale image shows two objects with different shapes. The observer’s task is to decide what the albedo is in a horizontal cross-section across each object (dashed white boxes). Under the image, the rows represent the actual luminance profile of the pixels across each object (labeled “L”), and the perceived albedo profile across each object for a typical observer (labeled “A”). The different perceived albedo profiles are due to the different perceived shapes of the objects, indicated by their different respective edge cues. In the left object,
because the shape looks flat on the front, the luminance difference across the object’s center is attributed to different albedo values. In the right object, the left side of each cylinder is perceived to face the light source more directly; different surface albedos are not required to explain the luminance differences across the center of the object. Figure 1.3B shows a graphical model for this perceptual inference problem. “A” represents the objects’ albedos, “S” represents the objects’ shapes, “L” represents the objects’ image luminance, and “E” represents the objects’ image edges. Image luminance is a function of both shape and surface albedo (and light source direction but we can assume that is constant here). The observer’s perceptual task is to estimate the surface albedo, but since different shapes result in different image luminance values, the effect the shape has on the image luminance must be explained-away in order to disambiguate the albedo. Because there is auxiliary edge information that provides an independent estimate of the shape, the shape’s contribution to image luminance can be disregarded and the albedo unambiguously estimated. The observer requires knowledge of how shape and albedo generate the image data (luminance and edge) to use it for perceptual estimation.

Figure 1.3: Auxiliary information “explains away” ambiguity. (A) The image depicts two objects with different shapes. The luminance profiles within the
dotted boxes are identical for both objects, as plotted below the objects in the row labeled “L”. The perceived albedo, as plotted in the row labeled “A”, is different for the different objects.

(B) The graphical model describes the process of how luminance measurements are a function of albedo “A” and shape “S”. The contributions of “S” and “A” to “L” are ambiguous and indistinguishable. But the availability of edge information, “E”, allows independent inference of “S”, and thus offers the possibility of “explaining away” the effect of “S” on “L” to unambiguously infer the value of “A”.

Bloj et al. (1999) reported that color perception was influenced by the geometry of the surface on which the color appeared, which implies the brain computes a color percept using a generative model of how color depends on light reflection between surfaces. Chapters 2 and 3 present experiments that test whether a generative model of the visual image size of the object is used to incorporate that auxiliary size and distance information.

1.3.7 Evaluating Bayesian perceptual models

Full Bayesian models of perception make quantitative predictions given stimuli and parameters that encode the particular model hypothesis. Although qualitative models are important, the ultimate goal for computational perception is to test quantitative hypotheses about internal computations. Hypotheses may include the content of observers’ sensory generative and/or prior knowledge, and can be represented as parameters in computational models. Ernst and Bülthoff (2004) review a number of experiments that measured observers’ internal beliefs about sensory noise by fitting weights to quantify the relative contributions of multiple sensory cues to perceptual judgments. Körding and Wolpert (2004) and Schwartz et al. (2005) quantified observers’ sensory generative knowledge by assuming an optimal Bayesian perception model and solving for the sensory generative knowledge required to cause the measured observers’ responses. Stock and Simoncelli (2006) reverse-engineer observers’ internal conditional likelihood functions and prior knowledge by separating the effect of trial-to-trial variability on measured responses from sensory noise and prior uncertainty.

There are a variety of features that distinguish a fully Bayesian observer from competing model observer, and most experiments mentioned are only sensitive to a subset of these features. For instance, even when an observer’s perceptual estimates and
response variability are consistent with the ideal observer, this does not imply that observer has knowledge of his or her own estimate reliability. A few studies (Landy et al., 2007; Whiteley and Sahani, 2008) suggest observers know their own estimate reliability to some degree, but this topic is ripe for further scrutiny.

1.4 Computational sensorimotor control

Sensorimotor behavior relies on perceptual interpretations for determining the task constraints and the body and environment states. Studies have investigated this cooperative functionality and revealed some of the general information representations and computational mechanisms that underlie sensorimotor behavior. As mentioned, percepts are often derived from noisy, incomplete, and redundant sensory data. Likewise, the outcomes of movements often deviate from the intentions of the brain. Such deviations are termed motor noise, caused by inherent variability in the firing of neurons that carry motor commands to the muscles, variability in muscle responses to motor commands, and variability in online sensory feedback during the movement which causes mis-estimation of the body state and consequent mis-corrections. Another important aspect of motor control is that there a variety of potential motor plans can accomplish a desired task goal. But, these potential plans may differ by the inherent danger or metabolic, or other biological, costs of execution. Incorporating the potential costs when selecting motor plans increases the overall utility of task performance (Harris, 1998). The high level of behavioral performance animals achieve in spite of estimate and outcome uncertainty, and variable action costs, justifies the use of probabilistic decision-making models to account for how the brain plans and executes sensorimotor behavior. Thus within the corpus of sensorimotor research, many studies have evaluated behavior in the context of Bayesian decision theory and other optimal models (Körding and Wolpert, 2006). Moreover, a consensus is growing that motor planning and execution are often near-optimal given task constraints, body dynamics and limitations, and online sensory feedback (Scott, 2004; Todorov, 2004). The over-arching theme of these recent results is that the brain places a premium on very precise, very efficient motor output, and the motor planning system uses all available information to generate moment-by-moment
recalculations and updates of motor behavior. In many instances, the statistics of perceptual estimates can be helpful, if not crucial, for planning near-optimal task-guided behavior, so it is natural to expect principled, robust mechanisms for incorporating such information into sensorimotor planning.

1.4.1 Background

Several important studies laid the groundwork for more detailed investigations of sensorimotor control and the role of sensory feedback. Keele and Posner’s (1968) classic study of how visual feedback improves reach accuracy concluded that during a reach, the information contained in visual feedback of the hand’s position was used to improve reach accuracy as early as 190-260 ms after it was received by the eyes. Carlton (1981) concluded that visual feedback could influence movement as soon as 135 ms after presentation. Fitts (1954) quantified how humans trade off speed and accuracy when planning reaches and concluded that people plan reaches according to a simple logarithmic relationship between movement duration and the ratio of target distance over target size. Schmidt et al. (1979) modified this so-called “Fitts’ Law” by supporting a linear relationship between movement error and the ratio of target distance over movement duration. Meyer et al. (1982) suggested that these seemingly contradictory relationships could each be correct under different force-impulse profiles, and different visual feedback conditions.

Systematic trade-offs between different movement attributes have been observed for duration, accuracy, and velocity, and often lead to stereotyped movements (Fitts, 1954; Schmidt et al., 1979; Lacquaniti et al., 1983). Because these stereotyped patterns emerge despite the fact that multiple adequate movement plans exist for any start and endpoint combination, it has been hypothesized that particular movement profiles may reflect paths that minimize biomechanical (Flash and Hogan, 1985; Uno et al., 1989), or perhaps task-relevant (Harris and Wolpert, 1998), costs. Effort has been put into modeling such trade-offs as optimizations given particular principles, like minimizing joint jerkiness (Flash and Hogan, 1985), torque (Uno et al., 1989), etc. Optimal models
of movement path selection imply that the brain compares different trajectories, and selects one with minimal associated penalties and maximum expected benefits.

1.4.2 Internal models

To generate a motor plan, the brain must select a sequence of motor commands that are predicted to bring the current body state to some desired body state. This is thought to require an internal representation of the body’s dynamics because it must be known what effect motor commands have on muscles, and how that translates to various body states. Kawato et al. (1987) proposed that sensorimotor behavior relies on an internal model, specifically an “inverse model”, which takes a desired trajectory as input and outputs the motor command necessary to cause the body to execute this trajectory. Once the brain selects a desired movement vector, it can send this vector to the inverse model to be converted into motor commands. Wolpert et al. (1995) reported that human motor behavior also incorporates sensory feedback using an optimal predictive estimation model, called a Kalman filter (Kalman and Bucy, 1961), which uses a forward model of sensorimotor dynamics. A forward model inputs motor commands and predicts sensory signals from the consequent motor output (Jordan, 1996; Miall and Wolpert, 1996). The advantage of a forward model is that it provides nearly immediate access to predicted sensory feedback, without the delays due to motor command transmission and sensory afferance and processing (Desmurget and Grafton, 2000). Kawato (1999) provides an excellent review of theories of, and evidence for, the brain’s use of internal models for sensorimotor control. Though beyond the scope of this paper, Shadmehr et al. (2005) reviews how the motor system may learn such internal models. Another type of internal model corresponds to world dynamics. In order to interact with a moving, or otherwise dynamic, target, it is important to know what the future state of the target may be. McIntyre et al. (2001) and Zago et al. (2004) suggest that humans use an internal representation of gravitational forces to predict future positions of a falling object when trying to intercept it. Conversely, when making a perceptual judgment of the future position, the object’s future position is predicted under the assumption of non-
acceleration. Zago et al. (2005) reported that such internal models of gravitational acceleration could be modified using feedback.

A detailed examination of internal models is beyond the scope of this discussion, but the key idea is that motor planning and online control systems use physically-realistic models to generate motor plans, and predict sensory feedback of those plans. In some ways this is analogous to the generative models required for Bayesian inference in perception (note: there are fundamental differences as well, so the comparison should not be over-extended).

1.4.3 Sensory feedback integration

Sensory feedback is also used for probabilistic computations. First, it may be helpful to highlight the distinction between sensory feedback of the body state, and sensory feedback of the environment state. Sensory feedback about the body is required to update the brain’s estimate of the body state, which may be compared to pre-planned desired trajectories, or perhaps the task itself, to decide whether corrective controls must be issued to the muscles to accomplish the task goal. Sensory feedback about the environment is required to update the brain’s representation of the task itself. The use of sensory feedback about body state for sensorimotor control occurs before, and during, a movement. Also, there are different types of sensory feedback that are used: vision and proprioception are considered to be the most important. Proprioceptive feedback means the sensation of the joints’ angles and muscle contraction/extension, which indicate posture and position of the body.

The brain’s use of visual feedback during a reach has been further investigated since Keele and Posner (1968; Carlton, 1981; Hay and Beaubaton, 1986; Jeannerod, 1988; Saunders and Knill, 2003; Saunders and Knill, 2005). Recently, Saunders and Knill (2003; 2005) used perturbation analyses to uncover the role of visual feedback during reaches. They manipulated visual feedback of hand positions and velocities at different points during a reach to assess how reach control reacted to perturbations online. Saunders and Knill (2003) concluded that humans continuously use visual feedback of the hand to control reaches, in contrast to classical views that online sensory feedback is
followed by showing that visual motion and position information were both used
continuously during reaches, and consistent with reliability weighted information
integration models. Both studies report sensory feedback latencies of roughly 160 ms.
These results indicate that sensory feedback about body state is essential for online reach
control, and used quickly and efficiently to improve reach performance.

Reference frames and coordinate transforms pose challenges for sensorimotor
behavior. Soechting and Flanders (1989b) showed that pointing errors were often
attributable not to sensory target mislocalization, but to the coordinate transformation that
converts the target’s location from external, 3D world coordinates into internal, joint-
space coordinates. Soechting and Flanders (1989a) showed this coordinate
transformation error may derive from a linear approximation strategy used for the
transform. McIntyre et al. (1997) reported that pointing errors to remembered targets
revealed a viewer-centered coordinate frame for storage of target location. Carrozzo et
al. (1999) reported that for continuous visuomotor interaction, in which target positions
are not stored in memory, external coordinates are converted to internal coordinates in a
gradual, two-stage process. Schlicht and Schrater (2007) suggest that coordinate-
transform variability may determine which coordinate frame task information is stored in.
See Lacquaniti and Caminiti (1998) for a review of neurophysiological representations of
visuomotor coordinates.

Sober and Sabes (2003) investigated how different sources of sensory feedback
about hand position, visual and proprioceptive, are used for planning a reach. They
identified two main functions that require sensory feedback about hand state: 1.)
selecting a movement vector plan, which includes subtracting the hand position vector
from the target position vector, and 2.) transforming the movement vector plan from 3D
world coordinates into the joint-space coordinates in which muscle commands are
specified. They concluded that selecting a movement vector plan to a target relies more
heavily on visual information, while transforming movement vectors into joint-space
coordinates relies more heavily on proprioceptive information. A potential reason for
this dissociation is that transforming sensory information between coordinate-frames
incurs neural-firing variability, meaning that using a piece of sensory outside of its original coordinate frame automatically makes the information less reliable because of additional coordinate transformation noise. Since visual information is indigenous to 3D world coordinates and proprioceptive information is indigenous to joint-space coordinates, if we assume coordinate transformation noise exists, then it is natural for each cue to have a higher weight during computations in its native coordinate frames. Sober and Sabes (2005) supported this view by further showing that these coordinate-specific information weights were also dependent on the modality by which the task itself was perceived. This means the relative weighting of proprioceptive information compared to visual information is heavier for felt (proprioceptively-detected) targets than for viewed (visually-detected) targets.

Knill (2005) examined how different visual cues to surface slant are used for different slant estimation tasks. They reported that humans assign different relative weights to stereo-disparity and textures cues depending on the task; for a perceptual reporting task, stereo-disparity is assigned a lower relative weight than for a reaching task that involves placing an object flat on the slanted surface. Greenwald et al. (2005) potentially explained Knill’s (2005) result by showing that stereo-disparity cues to 3D surface orientation are processed more rapidly than texture cues and because of this decreased latency, they are assigned a higher relative weighting for online feedback control than for motor planning. Also, they report the weightings are consistent with optimal cue combination that accounts for the temporal dynamics of cue information availability. Bruggeman et al. (2007) found similar results when comparing observers’ perceptual distance judgments to pointing judgments. Specifically, they showed that perceptual, “distance-matching” judgments were more susceptible to illusory perspective information than pointing judgments, which relied more heavily on binocular slant information. It is not clear if there is a relationship between Sober and Sabes’ (2003; 2005) and Knill’s (2005), Greenwald et al.’s (2005), and Bruggeman et al.’s (2007) results, but each suggests the brain has sophisticated understanding of the structure and quality of sensory information, and exploits this knowledge to optimize reach performance.
Several studies have reported that under degraded, or absent, sensory feedback about an object’s size or position, humans increase the hand’s grip aperture when reaching to grasp the object (Wing et al., 1986; Sivak and MacKenzie, 1990; Chieffi and Gentilucci, 1993; Bertheir et al., 1996). This indicates that motor planning is aware of the reliability of sensory feedback, and adjusts behavior accordingly.

So far, the works presented in this section have illustrated some of the brain’s solutions to computational challenges involved with sensorimotor control. Specifically, the brain employs internal models and sensory information’s structure and reliability representations to plan and execute movements. But not only does the brain perform sophisticated, probabilistic computations to produce movements, but also to select the task goals that the movements serve. The following research addresses issues related to how the task constraints affect sensorimotor behavior.

1.4.4 Separate roles of perception and action

Schrater and Kersten (2000) and Maloney (2002) suggest that differences between perceptual and sensorimotor interpretations of the same scene could be due to different loss functions being applied to the posterior probability distribution of world states given sensory information. Essentially this view amounts to: the task of perception is different from various motor tasks. The following studies investigated disparities between perception-only and visuomotor tasks.

There is controversy regarding the functional roles of perceptual and visuomotor processing. Ungerleider and Mishkin (1982) proposed that visual processing is divided into two distinct cortical processing streams, ventral and dorsal, one responsible for recognizing objects and their properties (“what” pathway), and one responsible for interpreting spatial relationships among objects in the scene (“where” pathway), respectively. Goodale and Milner (1992) championed this notion and numerous results in support of such dissociation have emerged (Milner and Goodale, 1995). Goodale (1998) revised the idea that the separation is purely visual by suggesting that in light of strong evidence that sensory and motor processing are tightly coupled, the dissociation is better understood as a functional separation between object recognition-related processing and
visuomotor processing. Support for such a functional dissociation came from studies that reported motor output resisted some notable visual illusions, indicating their interpretations are founded on different information (Aglioti et al., 1995; Carey, 2001; Glover and Dixon, 2001; Glover and Dixon, 2002; Glover, 2002; Danckert et al., 2002; Westwood and Goodale, 2003b; Glover, 2004). These results received substantial criticism on methodological grounds (Franz et al., 2000; Franz, 2001; Bruno, 2001; Franz et al., 2005) as well as simply disagreeing results (Hartung et al., 2005). Although no consensus has emerged regarding whether separate functional pathways support perception and action, there is sufficient evidence (Westwood and Goodale, 2003a; Kroliczak et al., 2006; Knill, 2005; Bruggeman et al., 2007) to suggest sensorimotor control may use different information and computations to produce actions than perception.

1.4.5 Task-relevant behavior

Several studies suggested that task performance is the critical criterion for movement planning, and governs motor control selections and execution. Using the neurophysiologically-supported assumption that the variance of neural motor signals is proportional to the magnitude of those signals, Harris and Wolpert (1998) modeled stereotypical arm and eye movement trajectories as optimal for minimizing arm and eye endpoint variability. This cost function penalizing predicted endpoint variability makes intuitive sense because the goal of movements is usually to achieve a desired body state, which may include position, posture, velocity, acceleration, etc. Harris and Wolpert’s (1998) assumption that neural motor signal variance is proportional to signal magnitude means small signals are preferred, so some potential biomechanical costs, such as not breaking your arm as you execute a movement, may be avoided simply as a consequence of these planning criteria and not explicit, intentional avoidance.

Scholz and Schoner (1999) and Scholz et al. (2000) proposed the idea that movement control is dependent on the task-relevance such that task-irrelevant movement dimensions are not controlled and allowed to vary freely. These studies analyzed standing-up movements as well as pistol-firing behavior to support this hypothesis.
Todorov and Jordan (2002) elaborated on this idea by showing that when the brain plans reaching behavior, it only performs online corrective controls when the hand deviates from the desired trajectory in a way that will disrupt task performance. This is characteristic of a control-theoretic construct called a *stochastic optimal controller*, which effectively only responds to sensory feedback that indicates a threat to task success. The reason such a controller is optimal is because motor commands are thought to introduce isotropic signal-dependent motor noise, so every control that does not substantially improve task performance actually decreases performance by adding motor noise to the overall movement. Resultant behavior follows the “minimum intervention principle”, and matches various observations of biological movement.

### 1.4.6 Statistical decision theory

Sensorimotor behavior can be treated as a decision process (Maloney, 2002). Here it is useful to briefly introduce some key ideas from statistical decision theory to preface the subsequent studies. To formalize the discussion of decision-making in the beginning of this paper, I will define several terms and explain their roles. *Statistical decision theory*, or SDT, is a set of prescriptive rules that specify how to choose actions based on information. *Bayesian decision theory*, or BDT, is a related set of rules that make slightly different assumptions about what information is available regarding the world state; this will be discussed in a moment. Decisions select actions, $\alpha$, whose effect depends on the world state, $\theta$. *Utility* is value associated with the consequences of an action in some world state, and is usually defined in terms of a *loss function*, $L(\theta, \alpha)$ that inputs $\alpha$ and $\theta$, and outputs utility (terminology varies, and often the loss function is replaced with a “gain function”, $G(\theta, \alpha)$, which is simply the additive inverse of $L(\theta, \alpha)$: $G(\theta, \alpha) = -L(\theta, \alpha)$). Positive loss is negative utility. As an example of loss functions, consider standing on a busy street. You “standing in the street”, and a “vehicle moving toward you” compose the world state. “Remaining still”, as well as “jumping out of the way”, are your potential actions. If your loss function inputs “standing in street, vehicle moving toward you” and “remaining still”, to output “very positive” loss (meaning: bad consequences), and also inputs “standing in street, vehicle moving toward
you” and “jumping out of the way”, to output “zero” loss (meaning: neither good, nor bad consequences), you should choose the action “jumping out of the way” to minimize loss, or bad consequences.

In some cases, the state of the world is not precisely known, but the actor may have measurements, $X$, generated by the state of the world. According to SDT, the actor should select each decision after considering likelihood of all possible world states, or the conditional likelihood, $p(X | \theta)$ (just as in Equation 1.1). Since $X$ is all the information the actor has, it makes sense to select action decisions, $\alpha$, as a function of $X$, $\alpha = \delta(X)$. Among the possible decision rules, $\delta$, the actor computes the expected loss as a function of $\delta$, or risk, $R(\delta, \theta)$, given the actual state of the world, $\theta$, for any possible $X$ that the world generates:

$$R(\delta, \theta) = \int p(X | \theta)L(\theta, \delta(X))dX$$

Equation 1.2

From this the actor picks the $\delta$ that has the lowest risk.

BDT assumes that the actor has some information about what world states are likely to occur, known as the prior probability, $p(\theta)$ (again, just as in Equation 1.1). BDT prescribes the actor to make decisions based on the measurements, $X$, as well as knowledge of $\theta$, represented by $p(\theta)$. The consequent Bayes’ risk, $BR(\delta)$, is not a function of $\theta$ because the actor uses $p(\theta)$ to consider all possible values of $\theta$ and compute the decision rule, $\delta$, with the lowest expected loss:

$$BR(\delta) = \int p(\theta)R(\delta, \theta)d\theta = \int \int p(\theta)p(X | \theta)L(\theta, \delta(X))dX d\theta$$

Equation 1.3

From this the actor can once again pick the $\delta$ that has the lowest Bayes’ risk.

Applying Bayes’ rule, from Equation 1.1, to Equation 1.3, allows the following substitution:
\[ BR(\delta) = \int \int p(\theta)p(X | \theta)L(\theta, \delta(X))dXd\theta = \int \int p(X)p(\theta | X)L(\theta, \delta(X))dXd\theta \]

Equation 1.4

Since \( X \) can be thought of as a fixed, sensory measurement, \( p(X) \) is simply some positive number, \( \delta(X) \) can be replaced by \( \alpha \), and the integral over \( X \) is removed:

\[ BR(\alpha) = \int p(\theta | X)L(\theta, \alpha)d\theta \]

Equation 1.5

This means that \( BR(\delta) \) can be rewritten as \( BR(\alpha) \), and by minimizing \( BR(\alpha) \), the actor obtains the action, \( \alpha \), with the lowest expected loss, given the sensory measurements. A digression regarding which is the “right” model, SDT or BDT, is beyond the scope of this discussion (see Maloney, 2002, and Kersten and Schrater, 2002, for more depth). The key point is that decisions are made based on knowledge of the world’s state, and the value of different action/world-state combinations. Additionally, there may be uncertainty about the outcome of an action, as the studies in the following section show.

One final note about decision-theoretic treatments of sensorimotor behavior echoes a previous point about Bayesian models of perception. First, because of the structure and important factors for sensorimotor behavior, decision theory offers an ideal context in which to build formal, quantitative models. These models do not necessarily imply that the brain is fundamentally a decision-theoretic machine that is flexible enough to handle any sort of decision problem. Rather, those behaviors that are consistent with ideal actions must use all available information to the ideal actor to make optimal decisions.

1.4.7 Human sensorimotor decision-making

Trommershäuser et al. (2003a; 2003b) applied SDT to examine how people select reach target locations under motor uncertainty, and various loss functions, as defined by
reach target location payoffs. They displayed a visual image on a touchscreen display of two overlapping circles: one was a “reward circle”, the other was a “penalty circle”. Participants were asked to reach to the touchscreen and contact a location of their choice; if they contacted the reward circle they won money, if they contacted the penalty circle they lost money, if they contacted neither they neither won nor lost money, and if they contacted both (in the overlap region) they won the reward circle money minus the penalty circle money. In their SDT model, Trommershäuser et al. (2003a; 2003b) considered a participant’s reach endpoint, \( \omega \), on a single reach to be a decision, \( \alpha \), plus some error due to motor noise, \( \Delta \), i.e. \( \omega = \alpha + \Delta \). Thus \( \omega \) was a random variable dependent on the participant’s decision, whose distribution could be expressed as:

\[
p(\omega | \alpha) = N(\alpha, \sigma_\alpha) = \alpha + N(0, \sigma_\Delta).
\]

This expression assumes \( \Delta \) is zero-mean, normally distributed with standard deviation, \( \sigma_\Delta \). Participants’ knowledge of the state of the world, \( \theta \), was assumed to be perfectly deterministic and matched the visual display of the reward and penalty circles. The loss function, \( L(\theta, \omega) \), was considered to be the amount of money associated with contacting a specific position with respect to the reward and penalty circles. Because there was no uncertainty as to the state of the world, the SDT and BDT risks were equivalent to each other, and to the loss function:

\[
BR(\omega) = R(\theta, \omega) = R(\omega) = L(\theta, \omega) = L(\omega) = L(\alpha + \Delta).
\]

But because the actual action, \( \omega \), was a random variable dependent on the decision, \( \alpha \), in order to compute the expected loss, or risk, of different decisions, \( \alpha \), one must integrate over possible values of \( \omega \), or alternatively \( \Delta \):

\[
R(\alpha) = \int p(\omega | \alpha)R(\omega)d\omega = \int p(\omega | \alpha)L(\omega)d\omega = \int p(\Delta)L(\alpha + \Delta)d\Delta.
\]

This expression can be minimized to select that decision, \( \alpha \), expected to minimize loss (or maximize reward). Trommershäuser et al. (2003a; 2003b) showed that humans indeed aim their reaches near-optimally with respect to the ideal model. These results imply that humans internally represent their motor noise and apply that information to optimally select movement plans.
A follow-up question Trommershäuser et al. (2005) asked was, can peoples’ internal motor noise knowledge be modified? They reported that when visual feedback of a participant’s fingertip was manipulated to suggest the fingertip position was more variable than it actually was during a reach, participants learned to compensate for this increase in perceived error by shifting their aim points, $\alpha$, consistent with the optimal reach model. This is evidence for a robust internal mechanism that monitors motor noise and updates planning strategies based on sensory feedback.

Trommershäuser et al. (2006b) and Wu et al. (2006) each reported human limits to this near-optimal reach target decision behavior. Trommershäuser et al. (2006b) showed that delaying the onset of target information until mid-reach diminished participants’ reach target decision performance away from optimal, while Wu et al. (2006) showed that when reaching to complex target configurations with more than one reward or penalty circles, humans often diverged from optimal reach target decisions.

Whiteley and Sahani (2007) and Landy et al. (2007) each asked whether action decisions take perceptual uncertainty into account. Whiteley and Sahani (2007) asked participants to make Vernier judgments and penalized errors asymmetrically: incorrectly choosing “left” would be penalized more greatly than incorrectly choosing “right”. Participants’ choices reflected a strategy that minimized negative consequences due to perceptual uncertainty across different reward/penalty and uncertainty conditions. Landy et al. (2007) performed an experiment in which line orientation judgments were penalized asymmetrically but found suboptimal behavioral choices in many cases. These studies represent first attempts to examine this question and a more complete picture will likely materialize in the future.

1.4.8 Sub-conclusion

Section 1.3 presented experimental results that suggest the brain uses reliability metrics, prior knowledge, and statistical information representations to infer body and environmental states more accurately. Section 1.4 presented experimental results that suggest the brain’s sensorimotor control uses internal models, online sensory feedback, and near-optimal action planning for more precise behaviors. In general, the
sensorimotor system can be thought of as an information-processing front-end driving an action-planning and execution back-end.

Green and Swets (1966) characterized many psychophysical tasks as signal processing (detection) problems in which stimuli are transduced in neural codes (sensations), and operations are performed on those codes to yield actions (decision rules). In this framework, Bayes-ideal perceptual processing is facilitated by decision rules that reflect optimal inference computations and reward-maximizing action selections. A fully Bayesian observer represents its beliefs about the state of world as a posterior probability distribution over possible states; the probability of a particular state is proportional to the brain’s belief that that state is correct (Berger, 1985). This contrasts with other coding methods that represent only the most likely state, or an approximation to the posterior such as the most likely state with an associated confidence measure.

To date little direct evidence exists to support the hypothesis that the brain explicitly represents posterior probability distributions over scene properties, but many of the studies in previous sections are capable of taking advantage of posteriors. Based on the flexibility and robustness of sensorimotor behavior, and evidence of a number of features of Bayesian inference, it is plausible that the brain represents its belief about the state of the world using information-rich posterior probability distributions. Though beyond the scope of this discussion Pouget and Snyder (2000) and Pouget et al. (2003) detail possibilities for instantiating posterior probability distributions in neural code and performing probabilistic computations with them.

1.5 Derivation of explaining-away rules

The experiments in Chapters 2 and 3 provide qualitative evidence of “explaining-away” in visual and haptic size and distance perception. Making optimal quantitative predictions for an observer that explains-away ambiguity in multiple scene properties’ influences on sensory cues (Figure 1.3) is more difficult than predictions for an optimal cue combination observer that uses cues caused by exactly one scene property (Figure 1.2). The main reason is that the optimal combination rule for explaining-away is nonlinear, and requires the observer to have more knowledge of the sensory generative
process and prior information. Still, quantitative predictions are necessary to thoroughly test the computational explaining-away process in the brain. The remainder of this section derives the optimal explaining-away rule for sensory cues that are linearly related to the scene properties and corrupted by Gaussian noise. This is a special case, but is a good approximation of many situations.

Explaining-away inferences are valid for situations in which two or more unknown scene properties generate two or more sensory measurements, and at least one of those sensory measurements is influenced by more than one scene property. I will derive an optimal inference rule for the simplest case, in which exactly two scene properties generate exactly two sensory measurements, but the same logic can be generalized to the many variable case. The unknown scene properties will be represented by $A$ and $B$, the sensory measurements by $X$ and $Y$. The influence among these variables can be represented by the Bayes’ net in Figure 1.4. Scene properties $A$ and $B$ each influence $X$, but $Y$ is only influenced by $B$ (i.e. independent of $A$).

![Figure 1.4: Explaining away Bayes’ net.](image)

$A$ and $B$ are unobservable scene properties. $X$ and $Y$ are observable sensations influenced by $A$ and $B$, according to relationships indicated by arrows. Because of the structure of the probabilistic dependencies (arrows), observation of $Y$ can influence beliefs about $A$.

Beginning with some basic assumptions, it is possible to quantify the relationships among the variables. There are two types of assumptions to follow, those that characterize the actual generative model regarding how $A$ and $B$ generate $X$ and $Y$.
(generative assumptions), and those that characterize the observer’s internal inference process (inference assumptions). For the observer to be Bayes-optimal, the inference computations must precisely reflect the sensory generative process, but for generality we do not assume an optimal observer and only constrain the inference rule to reflect the parametric form of the generative process, not the parameters themselves. The following generative assumptions detail the parametric form; the inference assumptions characterize the observer’s internal parameters that govern the inference process. The observer’s inference parameters will be labeled with a hat, i.e. \( \hat{h} \).

The first generative assumptions are that the relationships between \( A \) and \( B \), and \( X \) and \( Y \), are linear and Gaussian. This means that:

\[
\begin{bmatrix}
X \\
Y
\end{bmatrix} = T \cdot \begin{bmatrix}
A \\
B
\end{bmatrix} + \begin{bmatrix}
\omega_X \\
\omega_Y
\end{bmatrix},
\]

where \( T \) is a 2 x 2 linear transformation and \( \omega_X \) and \( \omega_Y \) are independent, zero-mean Gaussian distribution random variables that represent the sensory noise during a particular observation. Specifically, \( \omega_X \) and \( \omega_Y \) are distributed:

\[
\begin{bmatrix}
\omega_X \\
\omega_Y
\end{bmatrix} \sim N(\bar{0}, \Sigma_{XY}),
\]

where \( N(\cdot, \cdot) \) is a Gaussian distribution, and \( \bar{0} \) (2 x 1 zero-vector) and \( \Sigma_{XY} \) (diagonal matrix) represent the sensory noise mean and covariance, respectively.

A second generative assumption is that \( T \) is upper-triangular, which encodes the aforementioned statement that \( Y \) is conditionally independent of \( A \). A third generative assumption is that the prior distribution of \( A \) and \( B \) in the world is Gaussian such that:

\[
\begin{bmatrix}
A \\
B
\end{bmatrix} \sim N(\bar{\mu}_{AB}, \Sigma_{AB}),
\]

where \( ; \) is the inverse of a matrix.
where $\overline{\mu}_{AB}$ and $\Sigma_{AB}$ represent the prior mean and covariance, respectively. In the case depicted in Figure 1.4, $A$ and $B$ are conditionally independent and $\Sigma_{AB}$ is diagonal. Applying basic rules that describe the transformation of Gaussian random variables, we can represent the marginal sensory measurement distribution as:

$$
\begin{bmatrix}
X \\
Y
\end{bmatrix} \sim N(T \cdot \overline{\mu}_{AB}, T \cdot \Sigma_{AB} \cdot T^T + \Sigma_{XY}).
$$

For Bayesian inference, we need to know the posterior probability distribution (Equation 1.1) of $\begin{bmatrix} A \\ B \end{bmatrix}$ given $\begin{bmatrix} X \\ Y \end{bmatrix}$. Currently we know the marginal distributions of $\begin{bmatrix} A \\ B \end{bmatrix}$ and $\begin{bmatrix} X \\ Y \end{bmatrix}$, and the conditional distribution of $\begin{bmatrix} X \\ Y \end{bmatrix}$ given $\begin{bmatrix} A \\ B \end{bmatrix}$, and we can use these distributions to calculate the joint over all variables and subsequently the posterior. Because all variables are marginally Gaussian and linearly related, the joint is Gaussian.

We know the mean vector of the joint, $\begin{bmatrix} \overline{\mu}_{AB} \\ T \cdot \overline{\mu}_{AB} \end{bmatrix}$, and the block diagonal elements of the covariance, but not the off-diagonal blocks (which are transposes of each other, like all covariance matrices). Thus, we need to find the off-diagonal block of the joint distribution’s covariance matrix, $\Sigma_{OD}$. We can use an identity for conditional distributions of Gaussian random variables which holds that,

$$
T \cdot \begin{bmatrix} A \\ B \end{bmatrix} = T \cdot \overline{\mu}_{AB} + \Sigma_{OD}^{T} \cdot \Sigma_{AB}^{-1} \cdot \begin{bmatrix} A \\ B \end{bmatrix} - \overline{\mu}_{AB},
$$

Equation 1.6

and solve for $\Sigma_{OD}$:
Using the identity in Equation 1.6 again, we can formulate the posterior mean as:

\[
\hat{\mu}_{\text{Post}} = \hat{\mu}_{AB} + \Sigma_{OD} \cdot \left( T \cdot \Sigma_{AB} \cdot T^T + \Sigma_{XY} \right)^{-1} \cdot \left( \begin{bmatrix} X \\ Y \end{bmatrix} - T \cdot \hat{\mu}_{AB} \right)
\]

and the posterior covariance as:

\[
\Sigma_{\text{Post}} = \Sigma_{AB} - \Sigma_{OD} \cdot \left( T \cdot \Sigma_{AB} \cdot T^T + \Sigma_{XY} \right)^{-1} \cdot \Sigma_{OD}^T
\]

All of the terms on the right-hand sides of these two equations have been discussed above. The only aspect that has not been described is the observer’s inference rule. As mentioned, we do not assume a Bayes’-optimal observer, only that the observer makes the correct distributional assumptions. Thus, the observer’s inference rule is trivial:

\[
\hat{\mu}_{\text{Post}} = \hat{\mu}_{AB} + \hat{\Sigma}_{AB} \cdot \hat{T}^T \left( \hat{T} \cdot \hat{\Sigma}_{AB} \cdot \hat{T}^T + \hat{\Sigma}_{XY} \right)^{-1} \cdot \left( \begin{bmatrix} X \\ Y \end{bmatrix} - \hat{T} \cdot \hat{\mu}_{AB} \right)
\]

and:
\[ \hat{\Sigma}_{\text{post}} = \hat{\Sigma}_{AB} - \hat{\Sigma}_{AB} \cdot \hat{T}^T \cdot \left( \hat{T} \cdot \hat{\Sigma}_{AB} \cdot \hat{T}^T + \hat{\Sigma}_{XY} \right)^{-1} \cdot \hat{T} \cdot \hat{\Sigma}_{AB} \]

All terms with hats (i.e. \( \hat{\Sigma}_{AB} \)) reflect the observer’s belief about that term’s role in the generative process; for instance, \( \hat{\Sigma}_{AB} \) is the observer’s belief about the covariance of the prior distribution on \( \begin{bmatrix} A \\ B \end{bmatrix} \).

Thus, beliefs about A and B are represented as a Gaussian with mean \( \hat{\mu}_{\text{post}} \) and covariance \( \hat{\Sigma}_{\text{post}} \), which are computed as functions of the sensations, priors, and generative knowledge.
2. Humans Use Size cues to Improve Distance Perception

2.1 Abstract

This work investigates how humans use object size information to improve distance perception in an interception task. When attempting to intercept a moving object using only monocular visual information, the appropriate interception position may be ambiguous—the observer may be viewing a small object that is near or a large object that is far away. Regardless, humans are quite adept at monocular interception which implies that additional information is incorporated to disambiguate the visual information. We hypothesize that object size information is integrated to accomplish this disambiguation. We conducted an experiment in which we asked participants to intercept a moving ball in virtual reality.

In some trials participants were able to see and touch the ball before intercepting it; in others they were only able to see it. When allowed to touch the ball, participants showed improved interception performance. Effectively, they discounted the influence of the object’s physical size on its image size to obtain more accurate knowledge of the object’s distance. This sort of auxiliary information integration is characteristic of Bayesian information propagation, and is consistent with a broader body of evidence that characterizes perception as Bayesian inference.

2.2 Introduction

2.2.1 Background

Intercepting a moving object requires accurate perception and precise motor control. The actor must determine the time and position at which the object will approach, and coordinate a movement to that position at that time. Interception timing is a well-studied behavior; several studies have reported that information contained in the
visual display of an object is sufficient to determine an appropriate interception time (Lee & Reddish, 1981; Lee & Young, 1985; Peper et al., 1994; Caljouw et al., 2004).

In contrast, selecting an interception position can be quite difficult because the available image data is less reliable, or even ambiguous under monocular viewing. Consider the ambiguity that occurs when relying on an object’s image size to judge its distance: the object may be small and near the observer, or large and far from the observer, yet project to the same image size in both situations (Figure 2.1, top box). Despite this adversity, monocular interception is quite accurate (Servos & Goodale, 1998).

Figure 2.1: Size and distance ambiguity. The top figure illustrates the case in which the image size alone is used to determine an object’s distance; multiple objects project to the same image size and distance judgments are ambiguous. The bottom figure illustrates the case in which image size and object size (horizontal dotted lines) are used to determine the object’s distance; only one object is consistent with the image size and object size information, so distance judgments are unambiguous.

An observer needs information about the object’s size to disambiguate the distance indicated by image size. Unfortunately, object size information is not directly measurable as the object approaches in the moments before an interception. But when an auxiliary object size cue is available before the interception, the observer may incorporate that information with the measured image size to arrive at an unambiguous prediction of
object distance (Figure 2.1, bottom box). This qualitative auxiliary information use has been reported in size perception, where observers may integrate physical distance information to improve size estimates; called Emmert’s law (Boring, 1940; Weintraub & Gardner, 1970).

Several studies have investigated humans’ use of size information for distance perception, and when it develops. Granrud et al. (1985) showed that 7 month-old infants who were allowed to learn the size of different objects by playing with them used the size to judge the distance in post-play test phases. In contrast, 5 month-old infants did not exhibit the use of size information for distance judgments. Yonas et al. (1985) showed that when presented with two objects of different retinal visual angles, infants older than 5 months perceived the larger as nearer, but not 5 month-olds. Yonas et al. (1982) showed that 7 month-old infants’ and adults’ distance judgments are influenced by familiar size information associated with faces, but 5 month-olds are insensitive to familiar size. These results suggest that size information can influence distance judgments, and that such behavior becomes evident after 5 months of age.

Figure 2.2: Sensory generative process and scene property inference.
The graph represents a Bayes’ net that characterizes how an object’s size (S) and distance (D) relate to the retinal image size (I) and auxiliary object size (A) cues. Each circle represents a variable, and the small arrows represent one variable influencing another. The large arrow labeled “gen” indicates the sensory generation process, and the large arrow labeled “inf” represents the inference process. In order for an observer to infer the object’s distance from image size and auxiliary object size cues, knowledge of how scene variables (S and D) generate sensory cues (I and A) must be employed.

Integrating multiple pieces of sensory information (cues) to recover a relevant scene property is non-trivial because there are an infinite number of potential functions to combine the cues, but only a specific few that are useful for interpreting the scene.
Generally the ideal integration strategy reflects the cues’ generative process (Figure 2.2, arrow labeled “gen”): the cues should inform the observer about the scene in exactly the inverse way that the scene generated the cues (Figure 2.2, arrow labeled “inf”). In the case of distance perception introduced above, because an object’s retinal image size is approximately equal to the object’s size divided by the object’s distance (the so-called “small angle approximation”), estimating the object’s distance can be accomplished by multiplying the value of the image size cue by the value of the object size cue. The general principle of inverting the sensory generative process (inference) to use sensations to solve for scene properties is the core assertion of Bayesian models of perception (Knill & Richards, 1996; Kersten et al., 2004); the inference rules reflect the physical stimulation process. In contrast, non-Bayesian models of perception are unconstrained by the sensory generative process and may contain rules that have no physical interpretation.

Using auxiliary cues to disambiguate image size cues is consistent with a specific feature of Bayesian inference termed explaining-away (Kim & Pearl, 1983; Pearl, 1988). In our distance perception case, the term explaining-away refers to the auxiliary cue providing an explanation for how the object size influences the image size cue, which allows the remaining, unexplained component of the image size cue to be fully attributed to the object’s distance. Our experimental hypothesis is that the brain perceptually explains-away the influence of physical size on image size to disambiguate and improve distance judgments.

Our hypothesis fits into a broader theoretical context that asserts the brain’s perceptual decision-making is consistent with various elements of Bayesian inference (Knill & Richards, 1996; Kersten et al., 2004). This view is supported by a variety of experiments that study linear cue combination (Ernst & Bülthoff, 2004; Knill, 1998; Jacobs, 1999; van Beers et al., 1999; Ernst & Banks, 2002; van Ee et al., 2003; Battaglia et al., 2003; Knill & Saunders, 2003; Alais & Burr, 2004; Shams et al., 2005; Körding et al., 2007; Knill, 2007), prior knowledge use (Mamassian & Landy, 2001; Weiss, et al., 2002; Adams et al., 2004; Körding & Wolpert, 2004; Schwartz et al., 2005; Stocker & Simoncelli, 2006), and other aspects of perception (Knill & Pouget, 2004). Although many perceptual constancy effects can be framed as perceptual explaining-away, only a
few direct experimental tests of perceptual explaining-away have been reported (Knill & Kersten, 1991; Bloj et al., 1999).

One reason that explaining-away phenomena are less frequently measured than other types of perceptual Bayesian inference is that they naturally entail more complex, difficult-to-test models. Consider the situation in which one scene property causes multiple independent sensations that are each exclusively dependent on that property. The Bayesian inference rule for merging the cues is well-approximated by a linear, weighted average over all cues’ values, where the weights are proportional to the cues’ relative reliabilities among each other (Yuille & Bülthoff, 1996). In contrast, situations that require explaining-away inferences are fundamentally more complex and always depend on the specific relationships between the scene properties and cues. With the exception of some artificial, special-case situations, the appropriate rules for most explaining-away inferences are nonlinear and dependent on two or three times as many parameters as the comparable linear cue combination situations. This complexity renders strong, quantitative tests of explaining-away models of perception excessively difficult to execute; they require batteries of careful, integrated experiments to provide sufficient data to evaluate the models.

The experiment presented in this paper is a qualitative test of perceptual explaining-away for distance perception. It does not make strong claims of optimality, but serves as evidence that auxiliary size cues influence distance judgments and improve their accuracy.

2.2.2 Experiment overview

We placed participants in a virtual reality workbench and asked them to intercept a moving ball that they viewed monocularly. Participants’ hands were constrained to move on a fixed line that intersected their eye position and the ball’s movement path. The presence of the constraining line reduced the spatial degrees of freedom along which the participants could select an interception position so that only the distance was under their control. On each trial the ball moved through the scene and crossed the constraining line at some distance from the participant’s viewpoint (called the crossing distance) within the
participant’s reach. We recorded the distances at which the participants attempted to intercept the ball and considered regarded these as their judgments of the balls’ distances (called the judged distance).

Because the ball’s monocular image size is influenced by both its physical size and distance, for a particular image size, if the ball is small then it must be near, and will require a near interception distance. If the ball is large then it must be far, and will require a far interception distance. We tested the impact that size information had on interception performance by allowing the participant to touch the ball and feel its size before intercepting it in the haptic condition. We contrasted participants’ haptic condition judged distances with their judged distances in the no-haptic condition, which did not allow them to touch the ball before the interception. In the no-haptic condition participants’ could only rely on their prior knowledge of the balls’ sizes.

2.3 Methods

2.3.1 Stimuli

As mentioned, experiments were conducted in a custom visual haptic workbench (Figure 2.3) that provides calibrated visual and haptic information about objects in a virtual scene (Figure 2.4). Participants viewed a virtual scene, head-fixed (using a headrest) and monocularly (eye-patch on left eye), displayed on a 21” CRT (pixel resolution of 1600 x 1200, 85 Hz). The display was positioned above and at a 90 deg angle from the participant’s main direction of gaze and was viewed through a mirror to create a virtual image of the scene in front of the participant.
Below the mirror was a reachable space in which participants held a PHANToM force-feedback stylus (Sensible Technologies) that tracked their hand position throughout the experiment. The PHANToM was also used to generate force feedback to simulate stylus-object contact when objects were touchable (haptic condition), and to generate interception-contact feedback during the interception task. A common coordinate frame was established for this virtual space that put the origin at the participant’s right eye, the x- and y-axes lined up with the imaging plane, and gaze direction was along the z-axis.
Figure 2.4: Stimulus screenshot. The picture is an actual screenshot from the experiment, with colored lines overlaid to indicate the constraint line (yellow dotted line), the ball’s true movement path (green solid arrow), and ambiguous movement paths in the no-haptic condition (blue dotted arrows). The point at which the green solid arrow intersects the constraint line is the crossing distance. The points at which the blue dotted arrows intersect the constraint line represent mistaken judged distances that are more likely to result in the no-haptic condition when the distance was ambiguous. The overlaid lines were not visible to the experimental participant, but are illustrative purposes. They are not drawn from the participant’s viewpoint (in that case the constraint line would simply be a point because it lies in the line-of-sight, and the arrows would all fall on top of each other), but rather from a viewpoint elevated above the observer’s head so they can be distinguished from each other. The participant’s hand position was indicated by a 3mm diameter blue sphere (located near the intersection of the upper blue arrow and the constraint line).

Participants seated at this virtual workspace were able to see and feel several objects in the virtual scene. All visible and felt objects were spatially coincident with each other and visual rendering employed accurate perspective projection. The virtual workspace was designed to provide participants with a realistic 3D visuo-haptic experience.

On each trial, the virtual ball’s diameter was randomly selected from between 14 and 42 mm. Each ball had a random texture mapped onto it that did not contain any cues to distance. Figure 2.4 is a screenshot of the actual scene participants viewed, with the constraint line, ball’s true movement path, and ambiguous movement paths overlaid. The visual stimuli included a 3mm diameter virtual sphere that spatially coincides with the
participant’s actual hand position in the virtual workspace. The bottom of the virtual workspace was a visible and touchable ground plane with a checkerboard visual texture. The hand sphere and ground plane were always present throughout every trial.

2.3.2 Trials

Each participant performed four blocks of 80 trials on each of four days for a total of 1280 trials. The first day of trials was not included in the data analysis; those 320 trials were considered to be practice so that the participants could become acquainted with the experimental task.

2.3.3 Exploration phase

Each trial began with an exploration phase that corresponded to the two experimental conditions: haptic and no-haptic. Each block contained an equal number of randomly interleaved haptic and no-haptic trials. During the exploration phase, the ball was always located in the x-z plane and remained stationary. In the haptic exploration condition, participants were able to see and feel a virtual ball for as long as they wanted. Participants were allowed to view the ball for a minimum of one second and no maximum time. The ball’s distance was randomly selected from between 300 and 640 mm. The ball’s eccentricity was randomly selected from between -8.5 and 8.5 degrees, where 0 degrees is in front of the participant along the line of sight.

Before they were allowed to continue to the interception phase of the haptic condition trials, they were required to touch the ball at least once. In the no-haptic exploration condition, the participants’ hands were constrained to move along the constraining line used in the interception phase.

2.3.4 Interception phase

Once the participant was satisfied with the exploration, she depressed the mouse button to signal to the computer that she was ready to begin the interception phase of the trial. Regardless of whether the trial was part of the haptic or no-haptic condition, the interception phase proceeded in the same manner. Figure 2.5 shows a diagram of the interception phase. The ball was always located in the x-z plane during the interception
phase. Once the mouse button was depressed, forces were applied by the PHANToM to the stylus whenever the participant moved the stylus off of the constraint line. This ensured that the hand was positioned on the constraint line for the duration of the interception phase.

![Stimulus scene diagram](image)

**Figure 2.5: Stimulus scene diagram.**
This figure illustrates the components of the stimulus scene and the measurements. The eye is depicted at the bottom of the scene, the hand toward the right, and the ball toward the top-left. The participant adjusts the hand position along the constraint line in an attempt to intercept the ball's movement path. The true crossing distance is the point at which the ball's movement path intersects the constraint line, and the judged distance is the distance from the eye to the hand position at the point the ball crosses the constraint line.

When the mouse button was depressed, the ball rapidly travelled to a random starting position in the x-z plane on the left side of the virtual workspace and then remained stationary. The ball’s starting position was determined in polar coordinates: the distance from the eye was randomly selected from a uniform distribution between 1000 and 1500 mm; the ball’s starting angle (the visual eccentricity) from the z axis was uniformly selected from between -17 and -5 degrees. Once the participant was ready to
begin the actual interception, she depressed the mouse button and the ball began to move
toward the constraining line. The ball’s velocity was randomly selected from between
250 and 375 mm/s. In each trial the ball crossed the constraining line within the
participant’s reach and continued off the right edge of the screen. The crossing distance
on the line was randomly selected from between 340 and 600 mm. The ball’s total travel
time was between 1.3 and 4.8 seconds.

The participant’s task was to position the stylus tip at the point that the ball was
predicted to cross the constraining line. As the ball crossed the constraining line, the
participant received visual and haptic feedback that reflected the accuracy of the
interception. If the stylus tip was positioned within 32 mm of the ball’s center the
moment it crossed the constraining line, a virtual collision was felt (a small force pulse
was applied to the stylus) and the hand position sphere briefly flashed green. Otherwise,
no force-feedback was generated and the hand sphere briefly flashed red. The stylus
position at the time the ball crossed the constraining line was recorded and the distance of
that position from the participant is referred to as the judged distance.

2.3.5 Participants

Six university students, ages 21 to 30, participated in this study. One participant’s
data was excluded due to his inability to perform the task above chance levels, so that
five participants’ data were analyzed and presented. All had normal or corrected-to-
normal vision. Four participants were naive to the purpose of the experiment, and one
was an author (PWB, Participant 1). All participants gave informed consent in
accordance with the University of Minnesota’s IRB standards.

2.3.6 Data analysis

The Results section reports a regression analysis in which we fit a linear model to
predict judged distances from ball sizes. Specifically, for each condition (no-haptic and
haptic) of each participant’s data, we computed the best-fit (minimizing mean-squared
error) slopes and intercepts that outputted the trials’ judged distances based on input ball
sizes; this produces the maximum-likelihood estimate of the linear relationship under the
assumption of Gaussian noise. This quantified the linear statistical dependence between the ball sizes and judged distances.

The accuracies of participants’ judged distances were estimated by computing the standard deviation of the differences between the judged distances and true crossing distances.

Standard errors were estimated using Monte Carlo bootstrap resampling (10000 iterations) of each analysis (Efron & Tibshirani, 1993). Significance levels were estimated using direct hypothesis tests over those bootstrapped samples.

2.4 Results

2.4.1 Disambiguating size from distance

The purpose of this study is to examine whether humans can incorporate physical size information to disambiguate the effects of an object’s size and distance on image size. Our hypothesis was that observers use the haptic size information to explain away the effect of physical size on image size, which predicts that when the haptic size cue is available participants’ judged distances will be less dependent on the value of the ball’s size.
Figure 2.6: Size dependence of judged distances. The bar graph shows the correlation between the ball's size and the judged distance. The x-axis represents each participant, the y-axis represents the magnitude of the correlation, and the grey and white bars represent the no-haptic and haptic conditions respectively. The haptic conditions’ correlations were significantly smaller ($p < 0.0001$) than the no-haptic conditions’. Error bars represent SEMs.

We linearly regressed participants’ judged distances on each trial on the respective ball size to quantify the influence of the size cue on distance judgments (see Data Analysis in Methods). The correlation represents the statistical dependence between the ball size and judged distance. The dependence was expected to be zero if the physical size had no impact on participants’ judged distances, and less than zero if participants’ judged distances were influenced by the size (larger physical size generates larger image size which corresponds to shorter distance); the magnitude of the correlation was proportional to the influence of physical size on judged distance. Thus the quantitative prediction of our hypothesis is that the correlation between size and judged distance in the haptic condition is smaller than in the no-haptic condition. To directly compare the haptic and no-haptic condition correlations, we also computed their ratio for each participant.
Every participant’s judged distances were significantly influenced ($p < 0.0001$) by physical size (Figure 2.6), yet the dependences between size and judged distance were significantly greater ($p < 0.001$) in the no-haptic versus haptic conditions (Figure 2.7). This indicates that the haptic size cue significantly modified participants’ use of image size cues, but was insufficient to fully disambiguate the size from the distance, evinced by the non-zero haptic condition correlations.

### 2.4.2 Size information improves distance judgments

The Bayesian account of size and distance disambiguation using auxiliary size information is motivated by the fact that integrating size information will improve distance judgments. Our hypothesis was that distance judgments would not only bear less dependence on physical size when the haptic size cue was available, but that the judged distances would also improve in accuracy. We predicted that haptic condition
judged distances would have smaller average error (higher accuracy) from the true crossing distance than in the no-haptic condition.

Figure 2.8: Judged distance accuracy. The bar graph shows the accuracies of the participants' judged distances. The x-axis represents each participant, the y-axis represents the accuracy in terms of average error (standard deviation of judged distances from true crossing distances), and the grey and white bars represent the no-haptic and haptic conditions respectively. The haptic conditions' accuracies were significantly improved ($p < 0.03$ for Participant 2; $p < 0.0001$ for all other participants) over the no-haptic conditions'. Error bars represent SEMs.

We subtracted participants’ judged distances from the balls’ crossing distances on each trial, and computed the standard deviation of the differences across all trials to quantify each condition’s average error. Every participant’s interception performance significantly improved ($p < 0.03$ for Participant 2; $p < 0.0001$ for all other participants) when haptic size information was provided (Figure 2.8). This suggests that the brain’s reason for integrating size information is to reap the benefits in visually-guided movements offered by the auxiliary information. Participant 2 showed the least improvement from no-haptic to haptic conditions, but this was likely a ceiling effect due to the fact that her performance in the haptic condition was already so high (participants
1, 3, 4, and 5’s no-haptic performances were nearly identical, while participant 2’s was significantly better than the formers’).

2.5 Discussion

The results clearly indicate that participants integrated auxiliary haptic size information to improve distance judgments. Presumably, the brain has knowledge of how size and distance affect image size, and can discount the impact of physical size on image size when an auxiliary size cue is available. This is consistent with a Bayesian model of perceptual inference in which auxiliary measurements are used to explain away invalid interpretations of a scene and reduce possible perceptual decisions to the remaining valid interpretations (Pearl, 1988; Kersten et al., 2004; Knill & Kersten, 1991; Bloj et al., 1999).

One potential issue with our experimental design is that participants receive a weak auxiliary distance cue during the exploration phase of the haptic condition trials. Because obtaining a haptic size cue by touching the ball also provides a distance cue, in the haptic condition participants could potentially use this information to make their interception distance judgments. This possibility is predicated on the assumption that observers can extrapolate the sensed distance from the exploration phase through distance changes applied to the ball during the interception phase, and use this to improve judged distances. The information gain from the exploration phase distance cue is likely negligible due to the necessary extrapolation through the interception phase.

This study contributes to a growing corpus of evidence supporting the view that the human brain uses probabilistic inference to make perceptual decisions (Knill & Pouget, 2004; Kersten et al., 2004; Ernst & Bülthoff, 2004). In a full Bayesian decision-theoretic treatment of perceptual decision-making, the observer infers a posterior probability distribution (Yuille & Bülthoff, 1996); the posterior reflects the quality of the information that was used to compute it. An open question is whether the brain measures the quality of the information it collects to enable optimal inference (Jacobs, 2002).

For example, in the interception task presented in this study, if the auxiliary ball size cue was of poorer quality, a Bayesian observer would vary its use of the haptic size
cue in proportion to the cue’s reliability. Experimentally human participants could be tested for this prediction by presenting “squishy”, or otherwise unreliable, balls during haptic exploration. The Bayesian prediction is that the explaining-away effect we recorded (Figure 2.7) would be diminished due to the poverty of the auxiliary cue.

Although our experiment was conducted in a laboratory setting with controlled presentation of object size information, the fact that humans can make interceptions and judge distances to objects in the real world suggests that other sources of size information exist. One likely source of size information is prior knowledge of familiar object sizes. For instance, all of our experience reinforces the fact that soccer balls are larger than tennis balls, so it is possible that perceptual processing mechanisms take advantage of such prior experiences to improve distance judgments. Another future direction of this work is to study the role of familiar size in distance perception, and how humans learn from prior size experiences to build knowledge that can serve distance perception.
3. Humans Use Distance-Change Cues to Improve Object Size-Change Perception

3.1 Abstract

Many studies have shown that the perceptual system optimally integrates multiple pieces of sensory information (cues) when they are independently and exclusively generated by a single scene property (Ernst & Bültzoff, 2004). In contrast, there have been few reports of direct tests of how multiple cues are integrated when they are generated by multiple scene properties. Because an object’s size and distance both influence its projected retinal image size, it is impossible to unambiguously estimate the object’s size without integrating additional information. We investigated whether distance cues could disambiguate and improve size perception.

We presented participants with a virtual ball that simultaneously changed in size (by inflating or deflating) and distance (moving forward or backward). Participants were asked to discriminate whether the ball’s physical size increased or decreased by either looking at the scene monocularly or binocularly, and while either touching the ball with a finger or not. Under monocular viewing, image size cues were provided, physical size discrimination was poor. When additional haptic and/or stereoscopic cues about the ball’s distance were provided, participants’ judgments about the ball size improved significantly. Stereoscopic information had a significantly stronger effect than haptic, even though the reliabilities of haptic and stereoscopic cues were similar. Thus, cue reliability was insufficient to fully account for the distance cues’ discrepant influences on physical size-change perception.

Our results suggest the brain combines information about object distance and retinal image size to unambiguously estimate physical object size. We conclude that the brain exploits its knowledge of the relation between image size and physical distance to improve physical size judgments.
3.2 Introduction

A variety of perception experiments have studied how the brain integrates multiple pieces of uncertain or ambiguous sensory information (cues) to infer scene properties with higher precision than the individual cues allow on their own (Ernst & Bülthoff, 2004). Those studies have been restricted to humans’ perception of physical situation in which one scene property generates multiple independent cues. The following study also examines integration of multiple cues, but for perception of physical scenes in which two scene properties cause two sensory cues; in our experiment, an object’s physical size and distance generate retinal image size and distance sensory cues. The ideal perceptual solution to each type of situation reflects the cues’ generative process: the cues should inform the observer about the scene in exactly the inverse way that the scene generated the cues. The present study evaluates whether observers can use retinal image size-change cues and distance-change cues to judge an object’s changing physical size.

Judging the size of an object is important for many tasks, involving for example grasping and manipulation, but the object’s projection onto the retina generates an image size cue (sometimes called the “proximal size” in the perceptual literature) that is insufficient to uniquely determine the physical size of the object (sometimes called the “distal size”). For instance, a small, near object and a large, far object may project to the same image size at the retina. Without information about the object’s distance, the relative impact of the physical size and distance on the image size cue cannot be determined. Fortunately, sensory information about the object’s distance (distance cues) are often available: if two eyes are viewing the object then a stereo disparity cue is present, if the object is within reach and is touched, then a haptic cue is obtained. In effect, an auxiliary distance cue can rule out potential size-distance combinations that are inconsistent with that cue and has been termed Emmert’s Law (Boring, 1940; Weintraub & Gardner, 1970). But, incorporating distance cues in the perceptual estimate of the object size requires knowledge of how physical distance and physical size are related to image size and distance cues.
Our experiment investigates an extension of the size perception problem, specifically, perception of objects’ physical size changes. Imagine a balloon that floats toward you while deflating, the image of the balloon may remain relatively constant in size because the increase in image size caused by the nearing distance is balanced by the decrease in image size caused by the balloon’s diminishing physical size. If you view this scene with one eye, it is difficult to judge the rate at which the balloon is deflating. If you open both eyes or touch the balloon with your finger as it moves, these distance changes cues may help disambiguate the size- and distance-change rates. Our experiment provides image size-change and distance-change stimuli of a ball changing its size, and asks the participants to judge whether the physical size of the ball is inflating or deflating. We measured whether participants are able to take advantage of the auxiliary distance-change information to improve size-change judgments.

Few studies have investigated size-change perception, but physical size perception research has a long history (Holway & Boring, 1941). Previous examinations of size and distance perception concluded that humans often misjudge the size and distance of objects despite being provided cues that offer potentially disambiguating information. Holway & Boring (1941) demonstrated the role of distance information when making size judgments. They found size constancy (in which perceived size is proportional to object size) was best facilitated by providing many, strong distance cues, and the “law of visual angle” (in which perceived size is proportional to retinal image size) was prevalent when distance cues were removed. Epstein et al. (1961) surveyed the “size-distance invariance hypothesis” which holds that retinal visual angle constrains perception of an object’s size and distance ratio to a constant value. A consequence is that when a distance cue is available, the object size may be unambiguously estimated, resulting in size constancy across distances. Epstein et al. (1961) concluded that size constancy was subject to a variety of failures, and that the size-distance invariance hypothesis did not fully characterize the nature of size perception (Gogel et al., 1963; Ono, 1966). Epstein et al. (1961) and other authors (Heinemann & Nachmias, 1965; Gruber & Dinnerstein, 1965; Ono et al., 1974; Brenner & van Damme, 1999) acknowledge that distance judgments are not always veridical (apparent distances do not
always match physical distances) and accounts for some size misperceptions, and specific experimental design choices and task demands often contribute to the nature of the experiment’s recorded failure of size constancy (Kaufman & Rock, 1962; Kaufman & Rock, 1989; Blessing et al., 1967; Mon-Williams & Tresilian, 1999). Despite considerable effort, no unified account has emerged to generally explain size and distance perception, though myriad descriptive relationships among physical object properties and their respective induced percepts have been posited.

Figure 3.1: Sensation generation and scene inference. The graph represents a Bayes’ net that characterizes how an object's size- (S) and distance-change (D) rates relate to the sensed image size-change (I) and distance-change (A) cues. Each circle represents a variable, and the small arrows represent conditional probability relationships. The large arrow labeled (i) indicates the sensory generation process, and the large arrow labeled (ii) represents the inference process. In order for an observer to infer the physical size-change from image size-change and auxiliary distance-change cues, knowledge of the relationships between scene variable (S and D) and sensory cues (I and A) must be employed.

Many recent studies have explained various perceptually-guided behaviors as consistent with Bayesian inference (Knill & Pouget, 2004). The Bayesian framework is well-suited to formally characterize perception’s goal of recovering scene properties from sensory cues (Kersten et al., 2004). We adopt a Bayesian model to qualitatively describe how the brain uses sensory information for size-change perception. Figure 3.1 is a graphical model, called a Bayes’ net (Pearl, 1988), that represents the sensory generative process (Figure 3.1, arrow labeled “gen”) as well as the perceptual inference process (Figure 3.1, arrow labeled “inf”). The circles are termed nodes and represent scene properties and sensory cues: object size- and distance-change rates, and retinal image size-change and distance-change cues (“S”, “D”, “I”, and “A”, respectively). The arrows
represent the influence, or probabilistic dependence relationship, between the scene properties and the cues they generate. The generative process describes how sensory cues are caused by particular scene properties, according to their dependence relationships. The inference process prescribes how to compute beliefs about the scene properties from the sensory cue, according to their inverted dependence relationships.

Our observers’ task is to judge the size-change rate of a ball based on image size- and distance-change rate cues. Because recovering the size-change rate from the image size-change rate is ambiguous, the distance-change rate cue is important for making good judgments. This particular type of inference process is termed explaining-away (Kim & Pearl, 1983; Pearl, 1988). Effectively, the distance-change cue explains the role the ball’s distance-change rate plays in generating the image size-change cue, so the unexplained component of the image size-change cue’s value can be fully attributed to the ball’s size-change rate. In this terminology, our experimental hypothesis is that the brain uses distance-change cues to explain-away the effect of the ball’s distance-change on the image size-change cue, to disambiguate and improve perceptual ball size-change judgments.

For both classic cue combination situations that contain only one relevant scene property and explaining-away inferences, the Bayesian observer uses knowledge of the sensory generation process to compute beliefs based on cues. The distinction between our task and classic cue combination situations can again be illustrated by Figure 3.1. If you imagine removing the S node and its descendant arrow, and replacing the image size cue, I, with an additional distance cue, the graph then represents the case in which one scene property, object distance, causes multiple cues. Because of the more complex relationship between scene properties and sensory cues in our task, no simple combination rule, like reliability-weighted averaging (Yuille & Bülthoff, 1996), exists to prescribe the optimal inference process. Thus the purpose of this paper is to provide qualitative evidence that the brain uses auxiliary distance-change cues to disambiguate image size-change cues, and improve physical size-change perception.

3.3 Methods
3.3.1 Participants

Eight naïve volunteers (4 male, 4 female), ranging in age from 18 to 35 years old, participated in the experiment. Participants had normal or corrected-to-normal vision (Snellen-equivalent of 20/25 or better; Ferris et al. 1982), normal stereopsis of 60 s of arc or better (Stereotest circles; Stereo Optical, Chicago), and were right-handed (Oldfield 1971). They were recruited from the MPI Tuebingen Subject Database and in return for their participation they received payment of 8 €/h. Written, informed consent was obtained from the participants prior to their inclusion in the study and the rights of the participants were protected according to the 1964 Declaration of Helsinki.

3.3.2 Apparatus

Participants were seated in a virtual workbench that presented both graphical and haptic stimuli, described in detail in Ernst and Banks (2002). Participants looked toward the center of the workspace, tilting their heads 45 degrees forward. The head was stabilized with a chin-and-forehead rest. Participants could see the viewing monitor (21" GDM-F500R SONY monitor, 38.2 x 29.8 cm, with a resolution of 1280 x 1024 pixels and refresh rate 100 Hz) whose center was 50 cm from the eye reflected on a first-surface mirror, and which was tilted 22 degrees backwards from the fronto-parallel plane (top further than bottom). For stereoscopic stimuli, participants wore CrystalEyes TM (StereoGraphics) liquid-crystal shutter glasses which allowed different images to be presented to each eye. The haptic stimuli were presented using a Premium PHANToM force-feedback device (SensAble Technologies). The index finger was attached to the PHANToM device with a thimble and elastic band, allowing six degrees of freedom movements in a 20 cm³ workspace. The 3D position of the tip of the finger was monitored, and the computer applied simulated normal forces when the tip reached the positions of the virtual haptic objects. The apparatus was calibrated to spatially align the visual and haptic stimuli, and thus simulate a single scene.

3.3.3 General procedure
There were two experiments, referred to as *Main Experiment* and *Distance-change Reliability Experiment*. There were two distance-change cue types, *haptic* and *stereo*, in which ball’s distance-change rates were presented by touch and binocular disparity, respectively. At the start of each trial, a 35 mm diameter red ball was placed between 443 mm and 455 mm from the observer. If the trial contained a stereo distance-change cue, the ball was presented binocularly to the observer’s two eyes with disparity between each eye’s ball positions consistent with an interocular distance of 58 mm; otherwise the ball was presented monocularly. The participant signaled he or she was ready to begin the trial by reaching out and contacting the ball with the index finger (attached to the PHANToM device). Once contact was made, the PHANToM device applied forces to the fingertip to guide it to the center of the ball.

![Figure 3.2: Size- and distance-change rates schematic.](image)

Each box contains a qualitatively different physical situation presented in the experiment. The top-left box contains a ball receding and deflating, the bottom-left box contains a ball approaching and deflating, the top-right box contains a ball receding and inflating, and the bottom-right box contains a ball approaching and inflating. The top-left situation will always project to a decreasing image size and the bottom-right situation will always project to an increasing image size. The bottom-left and top-right situations will project to either decreasing or increasing image sizes depending on the magnitudes of the size- and distance-change rates.
At this point the experimental phase of the trial began: the ball began moving in depth with respect to the participant, while simultaneously changing in size, for a duration of 1000 ms. If the trial contained a haptic distance-change cue, forces were applied to the fingertip as the ball moved to maintain its position at the center of the ball throughout the movement; otherwise no forces were applied to the fingertip once the ball began to move. The ball also oscillated in the observer’s fronto-parallel plane following a sinusoidal displacement (with amplitude between 5.0 and 15.0 mm) in a random direction and at a random frequency (between 0.35 and 0.5 Hz). This was intended to both increased the uniqueness of the visual and haptic trajectories, increasing their perceptual fusion, as well as obviate local edge motion information as a direct indicator of image size-change. After 1000 ms, the ball disappeared. In the Main Experiment, only a single stimulus interval was presented. In the Distance-Change Reliability Experiment two stimulus intervals were presented, so after the first interval a new ball appeared and the second interval proceeded just as the first. Once the stimulus interval(s) were finished, two buttons appeared on the left side of the scene and participants were instructed to press the button that corresponded to his or her judgment of the scene. The trial ended once the button was pressed, and the subsequent trial began immediately.

3.3.3.1 Main experiment

In the Main Experiment the buttons were labeled “inflating” and “deflating”. The participant was instructed to press the button that corresponded to his or her perception of the ball’s physical size change. Participants’ choices constituted ball size-change absolute value discriminations.

3.3.3.2 Distance-Change Reliability Experiment

Each trial was designed as a two-interval forced-choice (2IFC) task. In every trial, both balls moved in the same direction with respect to the participant. Also, the balls never changed in size (equivalent to 0 mm/s size-change rate in the main experiment). In haptic distance-change cue trials, the ball disappeared as soon as it began to move. Following the two intervals participants were instructed to press one button
among two choices, labeled “1st” and “2nd”, indicating which interval contained the faster ball.

3.3.4 Design

The following independent variables were under experimental control: the type of distance-change cue(s) presented, the ball’s distance-change rate, and the ball’s size-change rate.

3.3.4.1 Main Experiment

Four distance-cue conditions were run, distinguished by the type(s) of distance-change cue that were presented: no-haptic/no-stereo (H-/S-), haptic/no-stereo (H+/S-), no-haptic/stereo (H-/S+), and haptic/stereo (H+/S+). The haptic and stereo distance-change cues are described above in the Procedure subsection; each provided a compelling sensation of the ball’s changing distance.

In the Main Experiment the ball’s movement rate was selected from between -104.0 and 104.0 mm/s, where a negative velocity corresponds to the ball moving toward the observer and a positive velocity corresponds to the ball moving away, in the line of sight of the participant. Specifically, we selected 3 pedestal distance-change rates, {-71.5, 0.0, 71.5 mm/s}, and varied the distance-change around these pedestal values by adding satellite values {-32.5, -26.0, -19.5, -13.0, -6.5, 0.0, 6.5, 13.0, 19.5, 26.0, 32.5 mm/s}, for a total of 33 possible distance-change values.

Concurrent with the ball’s distance change, its size changed at a rate selected from between -11.0 to 11.0 mm/s, where negative rates correspond to the ball deflating and positive rates correspond to the ball inflating. For each pedestal distance-change, we paired each of the satellite distance-change values with a particular size-change rate from the set {-11.0, -8.8, -6.6, -4.4, -2.2, 0.0, 2.2, 4.4, 6.6, 8.8, 11.0 mm/s}. The reason we varied distance-change rates around the three pedestal rates was to increase the number of unique exposed distance-change rates so as to prevent participants from learning the possible distance-change rates.

In total there were 33 unique distance and size-change rate pairs, each repeated 10 times. Figure 3.3 plots all unique distance- and size-change rate combinations (black
The colored lines in Figure 3.3 indicate three general distance-change directions: *approaching*, *neutral*, and *receding*.

![Figure 3.3: All trials' size- and distance-change rates.](image)

Because the ball’s distance and size changed during each trial, the image of the ball projected onto the observer’s retina varied in visual angle as well. But the rate of change of visual angle alone was insufficient to uniquely specify the ball’s rate of physical size change because it was not apparent to what degree the visual image size change was due to the ball’s changing distance. This ambiguity was a central feature for
the experiment and the following trial types presented different distance-change cues that could potentially disambiguate the physical size change.

### 3.3.4.2 Distance-Change Reliability Experiment

Two distance-cue conditions were run, distinguished by the type of distance-change cue that was presented: haptic and stereo. The experiment was 2IFC and the two intervals were called the standard and comparison, the order in which they were presented was randomly selected before each trial. For each distance-cue condition, two standard distance-change rates were used, {-55.0, 55.0 mm/s}. The comparison distance-change rates differed from the standard by a value from the set {-54.0, -36.0, -18.0, 0.0, 18.0, 36.0, 54.0 mm/s}. Each possible standard and comparison pair was repeated 14 times.

### 3.3.5 Data analysis

#### 3.3.5.1 Main Experiment

The experiment was designed to determine what effect distance-change cues had on physical object size-change perception. Therefore we compared size-change discrimination performance across the four distance-cue conditions. We used maximum-likelihood estimation (MLE) to fit participants’ size-change discrimination performance with psychometric functions that input size-change rates and outputted probability of responding “inflating”; the specific psychometric function was a modified cumulative normal function that allowed asymptotic choice probabilities to be greater than 0% and less than 100%, and was robust to noisy responses at extreme size-change rates (“psignifit” Matlab toolbox, see Wichmann & Hill [2001a; 2001b]).

To compare size-change discrimination performance across the four distance-cue conditions, we measured how participants used the distance-change cues across the three distance-change directions, approaching, neutral, and receding. Specifically, we measured whether size-change judgments in the different distance-change conditions were biased by the distance-change, which suggests participants’ failure to account for the true distance-change. We defined a metric that represents the bias in participants’
size-change judgments, called Point of Constant Subjective Size (PCSS). The PCSS corresponds to the physical size-change at which the participant was 50% likely to judge the stimulus as increasing in physical size, as indicated by the condition’s psychometric function fit (the point at which the horizontal gray line in Figure 3.4 intersects each condition’s psychometric function indicates the condition’s PCSS); this is analogous to the “point of subjective equality”, or “PSE”, used in other experiments. We estimated the best-fit line, termed discrimination boundary, between the PCSSs for the different distance-change direction trials (different colored lines- approaching, neutral, receding, in Figure 3.3) as a function of distance-change rate. The best-fit slope and intercept estimates corresponded to the line with the minimum mean-squared error (MSE) between itself and the PCSSs. If a participant’s PCSSs did not vary with distance-change rate, the discrimination boundary slope would be 0; the magnitude of the slope is proportional to the influence of distance-change rate on size-change judgments.

To find the discrimination boundary slope for a hypothetical participant who strictly relied on the image size-change to judge the physical size-change (image-only discriminant), consider the following formula that relates the ball’s physical properties to its retinal image size change:

\[
\dot{I}(t) = \frac{\dot{S} \cdot D(t) - \dot{D} \cdot S(t)}{D(t)^2},
\]

where \(\dot{S} , \dot{D} , S(t) , D(t) , \dot{I}(t)\) are size-change rate, distance-change rate, instantaneous size at time \(t\), instantaneous distance at time \(t\), and instantaneous image size-change at time \(t\), respectively.

We can find the set of \(\dot{S}\) and \(\dot{D}\) values for which the image size-change is zero by setting \(\dot{I}(t) = 0\),
\[
0 = \frac{\dot{S} \cdot D(t) - \dot{D} \cdot S(t)}{D(t)^2}
\]

\[
0 = \dot{S} \cdot D(t) - \dot{D} \cdot S(t)
\]

\[
\dot{D} \cdot S(t) = \dot{S} \cdot D(t)
\]

Because \(\dot{S}\) and \(\dot{D}\) are constant in our experiment, we can substitute:

\[
S(t) = \dot{S} \cdot t + S_0
\]

\[
D(t) = \dot{D} \cdot t + D_0
\]

where \(S_0\) and \(D_0\) are the starting size and distances, respectively. This leads to a convenient relationship between \(\dot{S}\) and \(\dot{D}\) that defines the values for which the image size-change is zero:

\[
\dot{D} \cdot \left(\dot{S} \cdot t + S_0\right) = \dot{S} \cdot \left(\dot{D} \cdot t + D_0\right)
\]

\[
\dot{D} \cdot \dot{S} \cdot t + \dot{D} \cdot S_0 = \dot{S} \cdot \dot{D} \cdot t + \dot{S} \cdot D_0
\]

\[
\dot{D} \cdot S_0 = \dot{S} \cdot D_0
\]

\[
\dot{S} = \frac{\dot{D} \cdot S_0}{D_0}
\]

As mentioned above, \(S_0\) was always 35 mm and \(D_0\) is approximately 450 mm (between 443 and 455 mm), so the function becomes:

\[
\dot{S} \approx \dot{D} \cdot \frac{35}{450} = \frac{7}{90} \dot{D}
\]

This linear relationship defines the discrimination boundary in size- and distance-change rate space that corresponds to an observer who exclusively relies on the image size-change to judge the physical size-change (represented by the diagonal dashed line in Figure 3.3). A participant that takes the influence of distance-change rate into account
when judging the size-change rate will instead have an unbiased discriminant boundary collinear with the axis, $\hat{S} = 0$ (represented as vertical dotted line in Figure 3.3).

We divided each participants’ best-fit discrimination boundary slope by the image-only discriminant slope ($\frac{7}{90}$) to quantify how biased participants’ size-change judgments were by distance-change rate, and termed this ratio PCSS bias. A PCSS bias near 1 indicates the participant’s discrimination boundary was consistent with the image-only discriminant, and thus was biased by the distance-change rate. A PCSS bias near 0 indicates the participant’s discrimination boundary was consistent with the unbiased discriminant, and thus explained-away the effect of distance-change on image size-change to make size-change judgments.

In order to quantify the effects of haptic distance-change cues on PCSS bias, we subtracted the PCSS biases in conditions with haptic distance-change information from their respective PCSS biases in conditions without haptic information. For instance, we subtracted H+/S- PCSS bias from H-/S- PCSS bias, and H+/S+ PCSS bias from H-/S+ PCSS bias, and termed these values Haptic bias reduction (stereo absent, S-) and Haptic bias reduction (stereo present, S+), respectively. The effect of the stereo distance-change cues on PCSS bias were computed as H-/S+ PCSS bias minus H-/S- PCSS bias, and H+/S+ PCSS bias minus H+/S- PCSS bias, and were termed Stereo bias reduction (haptic absent, H-) and Stereo bias reduction (haptic present, H+), respectively.

We used nonparametric bootstrapping to quantify the confidence in the experimental measurements (Efron & Tibshirani, 1993). Specifically, we resampled each condition’s responses with replacement 2000 times, and fit psychometric functions to each of these 2000 resampled datasets. We conducted each reported analysis on all 2000 PCSSs for each condition, and evaluated our hypotheses using Monte Carlo “hypothesis-testing”; a “p-value” of less than 0.05 corresponds to less than 100 out of 2000 analyses failing to satisfy the experimental hypothesis. 95% confidence intervals around a particular estimate were computed by sorting all 2000 estimates by their values and selecting the 50th and 1950th elements of the sorted list as the boundaries that contain 95% of the estimates.
Error bars on some figures were computed using the “median absolute deviations with finite sample correction factors” (MADC) from the LIBRA Robust Statistics toolbox for Matlab (Verboven & Hubert, 2005). MADC is a robust alternative to the sample standard deviation, and returns nearly identical values as std. dev. for normally-distributed data. For data that is highly skewed and/or kurtotic, MADC resists extreme data values and gives a more interpretable measure of a dataset’s variation scale.

3.3.5.2 Distance-Change Reliability Experiment

The Distance-change Reliability Condition assessed the reliability of the haptic and stereo distance-change cues for each participant. By varying the comparison interval’s distance-change stimulus around the standard, we were able to measure a psychometric function that reflected discriminability performance. The participants’ responses were indicative of which interval contained the greater distance-change rate stimulus. By fitting cumulative normal functions to participants’ psychometric functions using robust MLE (psignifit toolbox, Wichmann & Hill [2001a; 2001b]), we interpreted the fitted just-noticeable-difference (JND) as $\sqrt{2}$ times the standard deviation of the noise which corrupted a single distance-change cue (Ernst, 2005). This single cue standard deviation, which we refer to as noise, was an estimate of how reliable each distance-change cue was (reliability is inversely proportional to the noise’s standard deviation). Again we used nonparametric bootstrapping with 2000 bootstrap resamples to compute confidence measurements (Efron & Tibshirani, 1993).

3.4 Results

3.4.1 Main experiment

We provided observers with stereo and/or haptic cues to the ball’s distance-change rate and asked them to judge whether the ball’s physical size was inflating or deflating. The four distance-cue conditions (H-/S-, H+/S-, H-/S+, H+/S) contained each possible combination of haptic and stereo cues. Generally, participants’ size change judgments were more accurate when more distance change cues were available,
indicating that the brain exploits knowledge about how distance change, image size change, and physical size change are related to improve perceptual estimates.

As described in the Methods, we fit psychometric functions to each distance-direction (different colored lines in Figure 3.3) in each distance-cue condition. Figure 3.4 depicts the fitted psychometric functions for participant 3, where the height of the function at a particular size-change rate indicates the percentage of trials judged “inflating”. Figure 3.5 illustrates how the psychometric functions relate to the discrimination boundaries; the 3D surfaces correspond to participant 3’s psychometric functions interpolated across the entire size-change / distance-change rate space (imagine placing Figure 3.4’s psychometric functions on Figure 3.3’s surface, and interpolating across the empty regions), for each distance-cue condition. Again, the height represents the percentage of trials in which the ball was judged “inflating”. In the H-/S- condition, the participant’s discrimination boundary is aligned along the image-only discriminant (diagonal dashed line in Figure 3.3). As stereo and haptic cues are added (H+/S-, H-/S+, H+/S+), the discrimination boundary aligns with the veridical physical size-change of 0 (vertical dotted line in Figure 3.3).
Figure 3.4: Psychometric functions for all distance-cue conditions. Each graph depicts the approaching (left blue curve), neutral (middle green curve), and receding (right red curve) size-change psychometric functions in one distance-cue condition (inset boxes), for participant 3. The x-axis represents size-change rate (mm/s) and the y-axis represents the percent of trials judged “inflating”. The ‘X’s represent actual data and the curves represent psychometric fits (cumulative Gaussian) to those data. The horizontal gray lines represent the point at which the participant judged the ball “inflating” 50% of the time, and each curve’s intersection with that line indicates the curve’s PCSS. The vertical colored dashed lines indicate the size-change rate that corresponds to zero image size-change for each distance-change direction condition.
Figure 3.5: 3D psychometric functions for all distance-cue conditions (participant 3). The surface plots depict participant 3’s choice probabilities for each distance-cue condition. The x-axis represents size-change rate, the y-axis represents distance-change rate, and the z-axis represents the percentage of trials in which the participant judged the ball as “inflating”. The lines overlaid on the surfaces are similar to those in Figure 3.3 because this figure is for illustrative purposes, we interpolated the middle, neutral curve from Figure 3.4 to all size- and distance-change combinations in these graphs according to the measured discrimination boundary. This participant’s discrimination behavior changed as more distance-change cue information was made available, as evinced by the changing orientation of the discrimination slopes in the different distance-cue conditions.
Figure 3.6: Discrimination boundaries for all distance-cue conditions (all participants). This figure depicts all participants’ discrimination boundaries in all distance-cue conditions, on the same axes as in Figure 3.3. Each quartet of boxes indicates a single participant (labeled with bold-typed number in upper-left corner of upper-left box). Each box in a quartet is a single distance-cue condition (indicated by “H” / “S” on left side of each box). The solid colored lines are the same as those depicted in Figure 3.3. Gray regions represent size- and distance-change combinations judged “inflating” more than 50% of the time; white regions represent combinations judged “deflating” more than 50% of the time. The borders between white and gray regions are shaded in proportion to the confidence of those discrimination boundary estimates (as estimated by bootstrap resampling, see Data Analysis). Participants 1-6 show substantial orientation changes in their discrimination boundaries across distance-cue conditions.

Figure 3.6 shows all participants’ discrimination boundaries in all distance-cue conditions with Figure 3.3 overlaid. White regions represent size-change and distance-change combinations for which participants judged the physical ball size as “deflating”, gray represents “inflating” judgments. The borders between white and gray regions are shaded in proportion to the confidence of the discrimination boundary estimates (by bootstrap resampling, see Data Analysis). It should be noted that the confidence in the boundary corresponds to the variation in our estimates of the line slope through the psychometric functions’ PCSSs, in contrast to the psychometric function slopes.

Figure 3.7 shows the PCSS bias in each distance-cue condition for every participant. The height of each bar is proportional to the degree to which the size-change judgments were biased by the distance-change. To quantify the effect of distance-change cues on size-change judgments, we subtracted the PCSS bias when each type of distance-change cue was present from the PCSS bias when the distance-change cue was absent. Because we had two distance-change cues, stereo and haptic, we could assess the influence of each in the company of the other, and in isolation. Figure 3.8 depicts these PCSS bias differences, or bias reductions, caused by the presence of the distance-change cues.
Figure 3.7: Main experiment PCSS bias, all conditions (all participants). The figure depicts the PCSS bias for each participant. Each bar is a single distance-cue condition’s PCSS bias, with 1 MADC error bars (see Data Analysis). The distance-cue condition is indicated by the bar’s shading and referenced in the legend. The horizontal dashed line indicates the predicted PCSS bias for an observer that relies exclusively on the image size-change cue to make physical size-change judgments; this is why the H-/S- condition bars, in which only image size-change cues were available, all overlap the horizontal dashed line.
Figure 3.8: Distance-cue bias reduction (all participants).
The left and right graphs represent the haptic and stereo cues’ bias reductions, respectively, for each participant. The crosses are centered on the median bias reduction amount (negative numbers indicate reduced bias, positive number indicate increased bias), the lines represent 95% confidence intervals around the medians, and the numbers refer to the participant (see Figure 3.6 and Figure 3.7). The x-axis represents the bias reduction when the other distance-change cue was absent; the y-axis represents the bias reduction when the other cue was present. The solid grey lines and dots centered at (0, 0) represent zero bias reduction (no effect of the distance-change cue): crosses that lie on the negative sides of the (0, 0) point have reduced size-change bias when a distance-change cue was presented. Participants who used the distance-change cue to significantly reduce (p < 0.05, one-tailed bootstrapped hypothesis test) their size-change judgment bias in a single condition have a solid 95% confidence interval line, dotted lines represent non-significantly reduced bias.

Participants #1, 2, and 4 use haptic cues to significantly reduce their bias when stereo distance-change cues are absent, participants #4 and 5 use haptic cues to significantly reduce their bias when stereo distance-change cues are present, all participants except #8 use stereo cues to significantly reduce their bias when haptic distance-change cues are absent, and all participants use stereo cues to significantly reduce their bias when haptic distance-change cues are present (p < 0.05, one-tailed bootstrapped hypothesis test, see Data Analysis). The solid lines in Figure 3.8 represent significant bias reduction. Additionally, all participants except #8 have significantly reduced bias when both distance-change cues are present as opposed to when both are absent (p < 0.05, one-tailed bootstrapped hypothesis test).
These results indicate participants use stereo distance-change cues more than haptic cues for disambiguating the image size-change cue and improving physical size-change judgments. To more closely examine this discrepancy, we used data from the Distance-Cue Reliability Experiment.

3.4.2 Distance-Cue Reliability Experiment

We conducted a follow-up experiment to assess whether the reliability of each distance-cue was predictive of its bias reduction magnitude. We asked participants to perform a 2IFC discrimination task between two moving balls, in haptic and stereo distance-change cue conditions, to measure the magnitude of their sensory noise (Ernst, 2005).

The haptic and stereo distance-change noise magnitudes for each participant are shown in Figure 3.9. Each pair of bars represents the noise magnitude of the haptic and stereo cues for each participant. In general the haptic and stereo cues have similar noise magnitude for each participant; we failed to reject the null hypothesis that the noise magnitudes were identical for any participant (p < 0.05, one-tailed bootstrap hypothesis test). Because the haptic and stereo noise magnitudes were comparable, we cannot explain the pattern of more participants using the stereo cue to reduce PCSS bias than the haptic cue.

To make this point more clear, we measured the differences in the participants’ PCSS biases in the H+/S- and H-/S+ conditions from the Main Experiment, and compared these to the respective differences between the haptic and stereo distance-cue reliabilities. Figure 3.10 depicts a scatterplot between each of these differences. We conducted a correlation analysis using Spearman’s rank-coefficient method, and failed to reject the null hypothesis that there was no correlation among these quantities (p > 0.5).
Figure 3.9: Distance-Change Reliability Experiment distance cue noise std. dev.
The figure depicts the reliability of the haptic and stereo distance-change cues for each participant, and pooled across all participants. Each bar represents the standard deviation of the noise that corrupts a distance-change cue, with 1 MADC error bars (see Data Analysis). The haptic cue is indicated by the light bars, the stereo cue by the dark bars.
Figure 3.10: Noise magnitudes differences versus PCSS bias differences (all participants). The figure depicts relationship between the difference among distance-cue noise magnitudes and the difference among PCSS biases for the isolated distance-cue conditions (H+/S- and H-/S+). The x-axis represents the haptic noise magnitude minus the stereo noise magnitude measured in the Distance-cue reliability experiment. The y-axis represents the H+/S- PCSS bias minus the H-/S+ PCSS bias. Each dot is the estimate of each quantity for one participant. We failed to reject the null hypothesis that there was no correlation among these quantities (p > 0.5, Spearman’s rank-coefficient correlation).

3.5 Discussion

Our study supports the hypothesis that the brain uses sensory distance-change cues to disambiguate image size-change cues and improve physical size-change judgments. Specifically when retinal size is the only available cue, participants judged an object as inflating or deflating based on whether the image size is increasing or decreasing, because there is no sensory information specifying the object’s distance. When cues to the object’s distance-change are provided, participants take this auxiliary information into account to explain-away the influence of changing distance on image
size-change, resulting in more accurate judgment of the physical size-change. Binocular distance-change cues were more effective than haptic distance-change cues in reducing the distance-change-induced bias in participants’ size-change judgments. Both cues together allowed nearly unbiased physical size-change judgments in 7 of 8 participants.

The reason for the discrepant influences of haptic and stereo distance-change cues for reducing bias in size-change judgments remains unknown. One potential reason may be that the haptic cue is more weakly coupled with the image size-change cue than the stereo cue, perhaps reflecting the causal structure of the world. For instance, is it very difficult to contrive a situation in which stereo and image size cues are caused by independent objects in the world because the information contained in each cue is transmitted by the same photons to the eyes, and carried to the brain by the same nerve fibers. But, there are many situations in which the felt position of an object may differ from its visual position, like manipulating a tool, playing with a yo-yo, or touching an object that is occluded by a nearer object. In such situations, it is inappropriate to combine the visual and haptic cues because they do not share the same physical position.

This study qualitatively confirms that distance-change cues influence physical size-change judgments; there are several difficulties in constructing and testing quantitative models for this task that must be overcome to formally characterize participants’ perceptual judgments. One challenge is assessing what prior beliefs the perceptual system holds about the scene, which may be inferable from the data. For example, an interesting feature is that in the H-/S- condition, participants’ PCSSs fall around the physical size-change that is consistent with zero distance-change, and the slopes of their psychometric functions are quite steep. This can be characterized as prior belief that the distance-changes are distributed about zero, or the assumption of the same object distance throughout the trial. This has been noted in other studies and referred to as “specific distance tendency” by Gogel et al. (1963), and attributed to the presence of monitor flatness cues.

If participants assume fixed, non-zero distance-changes, their PCSSs would fall around some non-zero physical size-change rate. If they assumed a uniform distribution of possible distance-changes, their psychometric slopes would be very shallow. Neither
of these patterns is evident in the data, so we can infer that participants believe distance-changes are typically distributed around zero. This claim is bolstered by the pattern of participants’ physical size-change judgments in the H+/S- and H-/S+ conditions, which seem to be based on a mixture of the H-/S- and H+/S+ conditions’ judgments, and is expected if the sensory distance-change cue is combined with a prior expectation of zero distance-change, to form a hybrid belief of distance-change between zero and the sensory-indicated value. But without a quantitative model for the predicted size-change judgments in the presence of prior distance-change beliefs, it is difficult to draw strong conclusions about prior knowledge. Though it is usually not possible to direct measure prior beliefs, indirect methods of measuring these prior distance-change beliefs are being developed (Stocker & Simoncelli, 2006).

A future direction on which we are currently embarking includes placing the haptic and stereo distance-change cues in conflict (so that the haptic distance-change cue indicates 10 mm/s movement while the stereo cue indicates 30 mm/s, for instance) and measuring to what degree people rely on each cue’s indicated distance-change. One possibility is that the brain employs a winner-take-all strategy in which one cue’s information gets exclusive privileges in disambiguating the image size-change cue. Another possibility is that the cues’ indicated distances are averaged, perhaps in proportion to their respective reliabilities. This possibility is plausible given previous accounts of cue combination behaviors (Ernst & Bülthoff, 2004) but may not be the case given our result that haptic and stereo cues do not disambiguate the image size cue in proportion to their relative reliabilities.

Another future direction involves giving participants a direct physical size-change cue, such as a changing grip-aperture, and measuring its use in proportion to the indirect distance-disambiguated image size cue. Assessing the relative contributions of the direct and indirect cues can be accomplished by again placing the cues in conflict, and measuring to what degree participants’ judgments reflect each cues’ indicated physical size-change.

In order to use distance information to improve size judgments, the brain must have knowledge of how an object’s physical size and distance generate projected retinal
image size and distance cues. Our finding is consistent with a growing body of evidence that suggests perceptual processing uses auxiliary information to explain-away ambiguity in visual cues when making scene inferences (Knill & Kersten, 1991; Bloj et al., 1999; Kersten et al., 1996; Kersten et al., 1997; Battaglia et al., 2005). In addition, these findings extend the view that multiple source of independent information are fused to improve the perception of single scene properties (Ernst & Bülthoff, 2004) by showing that cues caused by more complex arrays of scene properties can be integrated as well.

In summary, our results indicate that the brain uses distance-change cues to disambiguate image size-change cues and improve physical size-change judgments. Haptic and stereo distance-change cues are effective, stereo more than haptic. This discrepancy is not due to differences in the cues’ relative reliabilities. The results support the view that perceptual processing employs knowledge of the sensory generative process to infer scene properties.
4. Humans Balance Perceptual and Motor Uncertainty to Maximize Reach Accuracy

4.1 Abstract

Previous research has shown that the brain uses statistical knowledge of both sensory and motor accuracy to optimize behavioral performance. Here we present the results of a novel experiment in which participants could control both of these quantities at once. Specifically, maximum performance demanded the simultaneous choices of viewing and movement durations, which directly impacted visual and motor accuracy. Participants reached to a target indicated imprecisely by a 2D distribution of dots within a 1200 ms time limit. By choosing when to reach, participants selected the quality of visual information regarding target location as well as the remaining time available to execute the reach. New dots, and consequently more visual information, appeared until the reach was initiated; after reach initiation no new dots appeared. However, speed-accuracy trade-offs in motor control make early reaches (much remaining time) precise and late reaches (little remaining time) imprecise. Based on each participant’s visual- and motor-only target-hitting performances, we computed an “ideal reacher” that selects reach initiation times that minimize predicted reach endpoint deviations from the true target location. Participant’s timing choices were qualitatively consistent with ideal predictions: choices varied with stimulus changes (but less than the predicted magnitude) and resulted in near-optimal performance, despite the absence of direct feedback defining ideal performance. Our results suggest visual estimates and their respective accuracies are passed to motor planning systems, which in turn predict the precision of potential reaches and control viewing and movement timing to favorably trade off visual and motor accuracy.

4.2 Introduction

Sensorimotor decisions involve three distinct components: world states, potential actions, and the rewards associated with different combinations of the first and second
components. Inferring the state of the world is an implicitly uncertain process due to incomplete and noisy sensory data, and the outcome of intended actions is inherently variable for reasons that include neural firing variability and imprecise muscle responses to motor commands. Recent studies reveal the brain contains a rich representation of these components of sensorimotor decisions including perceptual uncertainty, expected motor output variability, and monetary rewards (Trommershäuser et al. 2003a, 2003b; Wu et al., 2006). As states of the world are often dynamic, and actions unfold in time, decisions must also take time into account. Trommershäuser et al. (2005; 2006a; 2006b) showed that such near-optimal target selection may occur under tight time-constraints and even in cases when the target or reward information vary during the reach. Our study asks: does the brain represent visual and motor variability as functions of time? Moreover, when task performance is degraded by both visual and motor errors, can the brain manipulate the sources of these errors to maximize task performance?

Many visually-guided motor behaviors have time constraints to completion that induce a tradeoff between time allotted to gathering sufficient visual information for action planning and time allotted for action execution. Often there is exactly one action to be performed within a finite time; competition for the limited time allowed by the task requires the simultaneous choice of viewing and movement durations. Longer viewing durations improve the quality of visual information while longer movement durations decrease motor errors, thus the choices of these durations directly impact performance.

For example, consider driving on a mountain road during a snowstorm. As visual information accrues suggesting a possible obstacle ahead, like a boulder or car, a plan must rapidly be formed to continue on course, slam the brakes, or perform a risky swerve. Viewing and action time intervals are controllable and their durations directly affect task performance. Waiting lets you base your decision on greater sensory information, but leaves less time to execute the action. Reacting immediately may allow sufficient time to execute the action, but affords little information to decide which action is best. Thus time-constrained visuomotor tasks require balancing visual and motor timing to minimize visual and motor errors. The goal of this paper is to psychophysically test how well
people can achieve the optimal performance tradeoff in an experimentally-controlled visuomotor task.

We asked participants to reach to a target within a fixed time limit. Our experiment was designed to allow participants to trade visual accuracy for motor accuracy by choosing when to initiate their reaches. We evaluated how participants balanced this tradeoff to infer whether they knew, and were able to control, their visual and motor variability.

4.3 Methods

4.3.1 Apparatus

Participants performed trials in a virtual workbench consisting of the graphical display of a scene rendered under accurate perspective projection and a haptic interface (PHANToM force-feedback device, Sensable Technologies, Woburn, MA) that simulated the feel of objects in the scene. The graphics were displayed on a 21” CRT monitor (pixel resolution of 1600 x 1200, 85 Hz). The monitor’s image was reflected off a full-silvered mirror so that the graphics appeared co-aligned with the haptic workspace. The graphics and haptics were calibrated to always be consistent with the 3D virtual scene. The virtual scene consisted of a fronto-parallel task surface with contour information to enrich perspective cues, a start button, a countdown sand-timer, and the target stimulus. Participants wore eye-patches over their left eyes to remove potential stereo-disparity cue conflict. Participants placed their fingertips in a thimble attached to the haptic interface, which tracked the fingertip position. Visual feedback of fingertip position was continuously provided as a 1.3 mm diameter sphere.
4.3.2 Target

On each trial our participants’ task was to place his or her fingertip on a 2.5 mm diameter “start button” and reach to a 2.75 mm arc-length target (Figure 4.1). The target itself was invisible, but always lay collinear with a visible, gray “guide arc” which was an arc-length at a radius 8 cm from the start button (arc interval between 176 and 236 degrees counterclockwise from the positive x-axis). Also, the target’s position was indicated by a dynamic visual stimulus that appeared at the start of each trial. The stimulus was composed of very small dots scattered around the target position by sampling their positions from a 2D normal distribution with a mean at the target position and with a standard deviation (termed dot scatter level) that was varied across conditions. Three dot scatter levels were used: 4, 7, and 11 mm, or low, medium, and high, respectively.
At the trial start there were always 5 dots visible. As the trial time elapsed, the number of visible dots increased quadratically until a reach was initiated. The number of dots at any particular elapsed trial time was given by:

\[ N(t) = \text{floor} \left( t \cdot \left( \frac{5.47}{1200} \right) + 2.45 \right)^2 - 1 \]  

Equation 4.1

where \( \text{floor}(x) \) rounds \( x \) down to the nearest integer, \( t \) is elapsed trial time, and \( N(t) \) is the number of dots as a function of \( t \).

An ideal observer simply computes the sample mean of the dot positions as the optimal target location estimate. Thus the ideal observer’s standard deviation from the true target location decreases linearly with the square root of \( N(t) \). We chose the quadratic relationship between \( N(t) \) and \( t \) so that the ideal observer’s standard deviation from the true target location decreased linearly with \( t \).

4.3.3 Procedure: Experiment 1

Three conditions were run: combined test condition (CC), visual baseline (VB), and motor baseline (MB). The participant’s task varied slightly between conditions. Before each block, participants performed a series of practice trials to familiarize them with the task.

4.3.3.1 Practice sessions

Before every block in every condition, participants were given about 50 practice trials that were identical to the trials of the subsequent block but were not recorded or further analyzed. The purpose of these practice trials was to allow the participant to warm up and familiarize him or herself with the subsequent block’s particular trial procedure.

4.3.3.2 Combined test condition (CC)

In each trial, the participant held the fingertip on the start button for 500 ms (Figure 4.1) to signal he or she was ready to begin. When the computer detected the
ready state, the trial started. At the trial start, the invisible target was placed at a random location on the guide arc. Simultaneously, a rectangular “sand-timer” appeared above the start button and began counting down the fixed trial time (1200 ms). The sand-timer’s width was always proportional to the remaining trial time, thus providing a visual cue to the remaining trial time.

In order to complete a reach before timing out, the participant had to move the fingertip from the start button to the guide arc within 1200 ms of the trial start. The time at which the fingertip first contacted the guide arc was called the trial end. If the participant did not complete his or her reach within the time limit, the trial was considered a “time-out”, and repeated later with a novel stimulus and target location. In order to make a successful reach, the participant had to be in contact with the target at the trial end. No matter what, the trial always ended once the fingertip first contacted the guide arc, meaning the participant could not cross the guide arc then return to cross it again at another location. The computer recorded the position at which the fingertip first crossed the guide arc; this was considered to be the participant’s indication of the target position for that trial. The distance, measured along the guide arc in mm, between the center of the target and the position at which the finger first crossed the guide arc is termed the *reach endpoint offset*, and signifies the spatial inaccuracy of a reach. Because the target was a 2.75 mm arc-length, any reach endpoint offset less than or equal to 1.375 mm was considered successful while any offset greater than 1.375 mm was considered unsuccessful.

The computer also recorded the duration of two time intervals during each trial, viewing time ($t_v$) and movement time ($t_m$). $t_v$ was the interval between trial start and the reach initiation time, when the fingertip left the start button to move toward the target. $t_m$ was the interval between the reach initiation time and the trial end, when the fingertip first contacted the guide arc. The computer recorded the $t_v$ and $t_m$ durations with a temporal resolution of 11.8 ms.

Participants ran 6 blocks of 150 trials each. There were 3 dot scatter level sub-conditions (low, medium, and high), with 2 blocks in each sub-condition. Each block
took about 20 minutes to complete, and was divided into 30 trial sub-blocks. Between sub-blocks, participants were allowed to rest for a few minutes before continuing. Also, the participant was informed of the cumulative number of successful trials, in which the fingertip had contacted the target, for that block. This was the only form of feedback about task performance provided to the participants during the CC condition of experiment 1.

4.3.3.3 Visual baseline condition (VB)

The VB condition was designed to quantify the relationship between the number of visible dots and visual variability for each participant, without any effect of motor variability. The visual stimulus presentation that indicated the target location was similar to the CC condition, but instead new dots continued to appear after the reach was initiated until a pre-determined time that was varied between sub-blocks. Thus participants could freely move their fingertips without affecting the appearance of the visual stimulus. Also, to indicate the target location, the participant positioned the fingertip on his or her estimate of the target location and depressed the left mouse button. Participants in the VB condition were given 5000 ms to complete the task (instead of the 1200 ms in the CC condition), which was more than adequate time to precisely position their fingertip. Participants rarely used the full 5000 ms; typically they made target location selections within 2000-3000 ms. Effectively, these alterations eliminated target localization imprecision due to motor variability so that we could isolate the effect of the number of dots on visual variability. The computer recorded $t_v$ as the pre-determined duration of the interval over which new dots appeared as well as reach endpoint offsets, which were analyzed to determine their relationship.

Participants ran 3 blocks of 150 trials each. There were 3 dot scatter level sub-conditions (low, medium, and high), with 1 block in each sub-condition. Each block took about 20 minutes to complete, and was divided into 30 trial sub-blocks. The time intervals over which new dots appeared varied across sub-blocks. From first to fifth sub-blocks, the time intervals over which new dots appeared were: 0, 212, 425, 637, or 850 ms, or a total of 5, 11, 19, 28, or 39 dots, respectively. Between sub-blocks, participants
were allowed to rest for a few minutes before continuing. Also, the participant was informed of the cumulative number of trials in which he or she had completed a successful reach for that block. This was the only form of feedback about task performance provided to the participants during the VB condition of experiment 1.

4.3.3.4 Motor baseline condition (MB)

The MB condition was designed to quantify the relationship between movement duration and motor variability for each participant, without any effect of visual variability. The MB condition was similar to the CC condition, with several differences. Instead of scattered dots, the visual stimulus that indicated the target location was now a small 2.75 mm white arc-length that perfectly specified the target location. Also, the trial time limit was not constant across the MB condition, it was varied between sub-blocks, and was always less than or equal to 1200 ms. The countdown sand-timer’s initial width was shortened to match the sub-block’s trial time limit. Once the trial started, the sand-timer shortened at the same rate as in all other conditions so that the width was proportional to the remaining trial time. Also the trial start was determined by the reach initiation time, that is, trial time began to elapse once the reach was initiated. The participant was instructed to use as much time as possible to complete his or her reach. Effectively, these alterations removed the effects of visual target localization variability to isolate the relationship between movement duration and motor variability. The computer recorded $t_M$ and reach endpoint offsets, which were analyzed to determine their relationship. Note that in the analysis, we related reach precision to the measured $t_M$, not the cued movement time.

Participants ran 2 blocks of 150 trials each. Each block took about 20 minutes to complete, and was divided into 30 trial sub-blocks. The cued movement time varied across the 5 sub-blocks. From first to fifth sub-blocks, the cued movement times were: 1200, 988, 775, 563, and 350 ms. Between sub-blocks, participants were allowed to rest for a few minutes before continuing. Also, the participant was informed of the cumulative number of trials in which he or she had completed a successful reach for that
block. This was the only form of feedback about task performance provided to the participants during the MB condition of experiment 1.

4.3.4 **Procedure: Experiment 2**

We conducted Experiment 2 to assess the effect of performance feedback on reaching behavior. Exp. 2 was similar to Exp. 1 with several key differences. First, participants received performance feedback after every trial, in addition to the sub-block success summaries. Performance feedback consisted of the haptic sensation of a bump when the target had been contacted, as well as visual presentation of an illuminated arc-length at the target’s location that was green if the target had been successfully contacted and red if the target had been missed. Second, we did not include the medium dot scatter level condition, thus only the 4 and 11 mm dot scatter levels were used. Third, two Exp. 1 participants were unavailable, so only 5 of the original 7 participants participated in Exp. 2.

In Exp. 2 we repeated the VB and MB conditions identically to Exp. 1 (no performance feedback) for each participant and compared performance to Exp. 1.

4.3.5 **Participants and compensation**

7 (4 female, 3 male) naïve, right-handed University of Minnesota students with normal or corrected-to-normal vision participated in this experiment. Participants gave informed consent in accordance with University of Minnesota Institutional Review Board standards, were compensated $8 per hour of participation, and received bonus money depending on performance.

Participants could earn bonus money in all conditions. Whenever the participant’s reach contacted the target, the trial was considered a success. For each successful trial, the participant earned $0.02 bonus. This bonus money was awarded in addition to the $8 per hour compensation for participation time.

4.3.6 **Model**
4.3.6.1 Visual and motor accuracy tradeoff

In each CC trial, once the reach was initiated no new dots appeared, meaning the visual stimulus no longer improved. Thus, the reach initiation time effectively divided the trial into two distinct intervals: viewing time and movement time. Because visual localization accuracy improved with increased viewing time, $t_v$, and reach precision improved with increased movement time, $t_M$, the reach initiation time implicitly imposed a tradeoff between visual and motor accuracy. Ideal CC condition performance required the participant to select $t_v$ and $t_M$ to jointly maximize visual and motor accuracy.

![Figure 4.2](image)

Figure 4.2: Visual, motor, and predicted CC performance vs. elapsed time (participant 4). Visual accuracy, motor accuracy, and predicted CC condition reach endpoint offsets (dotted, dashed, and solid lines, respectively) are plotted as functions of reach initiation time (x-axis) for participant 4. For the VB and CC curves, reach initiation time is equivalent to $t_v$. For the MB curve, reach initiation time is equivalent to $(1200 - t_M)$ and assumes that the participant used all of the remaining trial time after leaving the start button as $t_M$, so that he or she finishes the reach at a trial time of 1200 ms. The y-axis represents the standard deviation of the reach endpoint offsets from the target position, so lower values mean higher accuracy and better performance. The 3 boxes represent low, medium, and high dot scatter levels. Visual accuracy curves were using VB condition data (Equation 4.3), motor accuracy curves were estimated using MB condition data (Equation 4.4), and CC endpoint accuracy was predicted by combining the visual and motor variability functions (Equation 4.5). The solid, vertical line in each box indicates the reach initiation time that minimizes the CC condition performance, $\sigma_C$, i.e. those $t_v$ that are predicted to yield maximum performance. Across dot scatter levels, the visual accuracy curves differ, and the $t_v$ that is predicted to yield maximum performance increases (numbers with arrows).
To make predictions for participants’ maximal CC performances, we computed an “ideal reacher” by combining participants’ individual predicted visual and motor variability, measured in their VB and MB conditions, respectively. Figure 4.2 depicts ideal predicted CC performance. Each box in Figure 4.2 shows a different dot scatter level condition. For illustrative purposes, Figure 4.2 constrains $t_M$ to be equal to $(1200 - t_V)$ so that $t_V$ and $t_M$ can be represented on a single axis. Notice how the tradeoff between visual and motor accuracy produces predicted CC performance curves that vary with viewing time, $t_V$. The viewing time that minimizes predicted CC performance function (arrows with numbers) increases as dot scatter level increases.

4.3.6.2 Ideal reacher derivation

CC condition reaches were modeled as follows. The process of planning and executing a reach included two components. First, the participant visually estimated the target’s location. Let $X$ represent the true target location, and $\hat{X}$ represent the visually estimated target location. We assumed there was visual variability that contributed to additive errors, $\varepsilon_V$, between $X$ and $\hat{X}$ such that:

$$\hat{X} = X + \varepsilon_V$$

Second, the participant directed a reach toward the estimated location. Let $Z$ represent the reach endpoint. We assumed there was motor variability that contributed to additive errors, $\varepsilon_M$, between $\hat{X}$ and $Z$ such that:

$$Z = \hat{X} + \varepsilon_M$$

Taken together:
resulting in a total error, \( \varepsilon_c = \varepsilon_v + \varepsilon_M \), that constitutes the reach endpoint offset between \( Z \) and \( X \).

Visual errors, \( \varepsilon_v \), varied with \( t_v \) and \( \sigma_d \), had mean equal to 0, and variance, \( \sigma_v^2 \), such that:

\[
\sigma_v(t_v, \sigma_d)^2 = \left( \rho^2 + 1 \right) \frac{\sigma_d^2}{N(t_v)} + 2 \rho \omega \frac{\sigma_d}{\sqrt{N(t_v)}} + \omega^2
\]

Equation 4.3

where \( \rho \) and \( \omega \) are free parameters that were fit by maximum likelihood estimation to the VB data. The derivation of Equation 4.3 and the fitting procedure are described in Appendix A. Note, for the rest of the paper we will suppress explicit dependency expressions after their initial presentation, so for instance, \( \sigma_v(t_v, \sigma_d)^2 \) will simply be written \( \sigma_v^2 \).

Motor errors, \( \varepsilon_M \), varied with \( t_M \), had mean equal to 0, and variance, \( \sigma_M^2 \), such that:

\[
\sigma_M(t_M) = D \cdot \exp(\alpha - \beta t_M) + \gamma
\]

Equation 4.4

where \( \alpha \), \( \beta \), and \( \gamma \) are free parameters that were fit by maximum likelihood estimation to the MB data, and \( D \) was the distance between the start button and the guide arc (8 cm). The derivation of Equation 4.4 and the fitting procedure are described in Appendix A.

As mentioned above, in each CC condition trial we allowed participants to choose a combination of \( t_v \) and \( t_M \) values. We will refer to the combination of \( t_v \) and \( t_M \)
chosen on a single CC condition trial as \([t_v, t_m]\). Because the variance of the sum of two independent random variables is the sum of their variances, the variance of \(\varepsilon_c\), denoted by \(\sigma_c^2\), is the sum of \(\sigma_v^2\) and \(\sigma_m^2\):

\[
\sigma_c^2(t_v, t_m, \sigma_d) = \sigma_v^2 + \sigma_m^2
\]

Equation 4.5.

Note that the combined variance depends on both timing choices, \(t_v\) and \(t_m\), as well as the dot scatter level, \(\sigma_d\).

Since the VB and MB conditions measured how \(t_v\), \(t_m\), and \(\sigma_d\) affected participants’ visual and motor variability independently, we used these relationships (Equation 4.5) to formulate an ideal reacher model to predict how combining these sources of variability would affect participants’ reach endpoint offsets in the CC condition. Specifically, by fitting the free parameters of \(\sigma_v^2\) and \(\sigma_m^2\) (see Appendix A), we were able to compute predicted \(\sigma_c\) landscapes that depended on \(t_v\) and \(t_m\), for each of the three experimental \(\sigma_d\) levels. Errors in fitted parameters (estimated by bootstrapping the data across MLE fits) were propagated by computing predictions for each set of bootstrap parameter fits. The ideal reacher is subject to the constraint that \(t_m\) is less than or equal to the remaining trial time after reach initiation \((1200 - t_v)\), just as in the CC condition. We assumed that participants learned the three \(\sigma_d\) levels from the practice and VB conditions, and immediately implemented that knowledge in the CC condition.

For each \(\sigma_d\) level, we predicted ideal timing choices, \([t_v^*, t_m^*]\), by numerically minimizing Equation 4.5. Note, our ideal model’s predictions are based on the assumption that timing choices will be executed without error. However we expect that humans viewing times and reaching times will differ from their choice intentions due to noise. Although we investigated the effects of noisy execution of timing choices on
theoretical predictions, we found a full analysis provides only a modest increase in our ability to predict performance in the CC condition, so for simplicity we ignored it.

4.4 **Results**

4.4.1 **Experiment 1**

We compared each human participant’s performance in the CC task to their ideal predicted performance. A critical test is whether participant’s choices vary with systematic changes in the quality of visual information (dot scatter level) as they do for the ideal predicted choices. To ensure our results reflected normal behavior and not learned associations, no performance feedback was given to participants in the reach task except a cumulative score presented every 30 trials (we later conducted Exp. 2 with performance feedback provided to assess what, if any, impact this had on performance).

4.4.1.1 **Baseline conditions**

The VB condition measured participants’ $\sigma_y$ functions. Since the number of dots, $N(t_y)$, increased with viewing time, $t_y$, and more dots gave more information about the target location, visual offsets should decrease as $t_y$ increases. The dashed curves in Figure 4.2 show one participant’s $\sigma_y$ function, as described by the visual variability model (Equation 4.3) with best-fit parameters. Figure 4.2 illustrates that as $t_y$ is increases, the participant’s visual variability decreases.
Figure 4.3: Dot centroid estimation noise (all participants).
Each line represents the average error (quantified as a standard deviation) of one participant's estimates of dot centroid locations in the VB condition. The x-axis represents different numbers of dots, and each box represents a different dot scatter level. Perfect estimation of dot centroids would correspond to noise standard deviations of 0. In Appendix A, the dot centroid estimation noise was referred to as $\sigma_c$.

Figure 4.3 depicts participants’ errors in estimating dot centroid locations for various numbers of dots, $N(t_v)$, and dot scatter levels, $\sigma_d$, in the VB condition. These errors are computed relative to the mean, or centroid, of each trial's dot scatter (rather than the underlying target location), to discount visual variability due to centroid misestimation. CC condition performance depends on the overall visual error, $\varepsilon_v$, which combines the errors shown in Figure 4.3 with errors due to deviations between the centroid of the dots and the true target location. The relationship between the different components of participants’ visual errors and $\sigma_v$ is provided in Appendix A. Tassinari et al. (2006) reported a similar analysis of human dot centroid mislocalizations, and although they used slightly different dot scatter levels and numbers of dots, their results are consistent with ours when extrapolated to our particular conditions.

The MB condition measured each participant’s motor precision (quantified as expected motor offset from the target location) as a function of $t_M$. Remember that $t_M$ is the remaining trial time following a reach initiation, so shorter viewing times, $t_v$, allow greater movement times, $t_M$. The dotted curves in Figure 4.2 show one participant’s $\sigma_M$. 

98
function, as described by our motor precision model (Equation 4.4) with best-fit parameters. Figure 4.2 illustrates that as $t_M$ increases (earlier reach initiations), participants’ motor variability decreases. $\sigma_M$ was used to predict the contribution of motor variability to $\sigma_C$ in the CC condition. Table 4.1 represents mean (± SEM) values for the fitted parameters from our motor precision model (Equation 4.4).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean value</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>-1.213</td>
<td>±/±0.36</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.271</td>
<td>±/±0.062</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.327</td>
<td>±/±0.202</td>
</tr>
</tbody>
</table>

Table 4.1: Motor precision model parameters (all participants). The table represents the parameter values of the motor precision model fit using MLE (Equation 4.4). The SEM column represents SEMs computed by bootstrapped resampling of the raw data.

4.4.1.2 CC condition: timing choices

Different dot scatter levels, $\sigma_d$, produce different $\sigma_C$ functions (Equation 4.5). Because the minimal $\sigma_C$ varies with dot scatter level, the ideal performer adjusts its $[t_v^*, t_M^*]$ choices for different values of $\sigma_d$ (solid, vertical lines in Figure 4.2 with arrows and numbers indicating the $t_v$ value). Specifically, the ideal performer should increase $t_v$ and decrease $t_M$ as dot scatter level increases. We compared each participant’s actual timing choices to his or her respective ideal choice predictions for different dot scatter levels to assess how well humans chose to trade off visual and motor variability.
Figure 4.4: $\sigma_C$ and measured $[t_Y, t_M]$ choices (participant 4).

Each of the 3 boxes depicts participant 4’s CC condition predicted $\sigma_C$ function for different dot scatter levels (grayscale contours), as well as measured $[t_Y, t_M]$ choices (points). The x-axis represents $t_Y$ in ms, the y-axis represents $t_M$ in ms. $\sigma_C$ is quantified as the standard deviation of reach endpoint offsets from the true target location (in mm). Dark contours represent large values of $\sigma_C$, light contours represent small values of $\sigma_C$. Each point is a $[t_Y, t_M]$ choice measured on a single trial.

Figure 4.4 shows participant 4’s measured $[t_Y, t_M]$ choices from every CC condition trial (points) superimposed over the predicted $\sigma_C$ landscape as a function of $[t_Y^*, t_M^*]$ (grayscale contours). Lighter shades represent smaller values of $\sigma_C$ from target location; darker shades represent larger values of $\sigma_C$. We interpreted the mean of the observed timing choices as the participant’s estimate of the timings that would maximize his or her performance (minimize reach endpoint offsets), and the covariance as resulting from variability in both decision-making and motor output.

To simplify the analysis and provide an intuitive measure of participants’ timing choices, we computed the perpendicular projections of each $[t_Y, t_M]$ vector onto the nearest point on the line with slope -1 and y-intercept 1200ms (termed the total time axis), and called these values $t_C$. The projections of the ideal performer’s $[t_Y^*, t_M^*]$ choices onto the total time axis are called $t_C^*$. The total time axis can be thought of as
the set of timing choices for which \( t_v + t_M = 1200 \) ms. Each \( t_C \) represents the nearest point on the total time axis to the participant’s \([t_v, t_M]\) choice. This axis is meaningful because the minimum of \( \sigma_c \) for ideal predictions shifted along that axis as a function of \( \sigma_d \).

**Figure 4.5:** Timing choices and their corresponding \( t_C \) values (participant 1). This figure is similar to Figure 4.4, but for participant 1. Instead of separating dot scatter levels into different boxes, we superimposed the \([t_v, t_M]\) choices from all dot scatter levels on each other and labeled them using different colors. The histogram represents the \( t_C \) values onto the total time axis (diagonal line). The x-axis represents \( t_v \) in ms, the y-axis represents \( t_M \) in ms. The left blue, middle green, and right red bars are \([t_v, t_M]\) choices from the low, medium, and high dot scatter levels, respectively. The shifts between the histograms illustrate how the participant’s timing choices varied across dot scatter levels.
Figure 4.5 shows a scatterplot one participant’s timing choices and their respective $t_C$ values. The x-axis represents viewing time, the y-axis represents movement time, each dot represents a $[t_v, t_M]$ choice from one trial, and the diagonally-aligned bar graph represents a histogram of $t_C$ on the total time axis. Left blue, middle green, and right red bars represent trials from the low, medium, and high dot scatter level conditions, respectively. Notice that as dot scatter level was increased, $t_C$ choices shifted along the total time axis as $t_v$ increased and $t_M$ decreased.

Figure 4.6: Regression slopes of $t_C$ values across dot scatter levels (all participants). Each bar represents the slope of the 1st-order linear regression of $t_C$ values on dot scatter level (with 95% confidence intervals) for each participant. Every participant shifted his or her $[t_v, t_M]$ choices positively along the total time axis as dot scatter level was increased. This is qualitatively consistent with the shift predicted by the ideal performer model (see solid, vertical lines in Figure 4.2).

We performed 1st-order linear regression analysis on the $t_C$ values as a function of dot scatter level. Figure 4.6 summarizes the regression slopes (with 95% confidence intervals) for each participant. All slopes were significantly greater than zero ($p<.05$), indicating all participants shifted $t_C$ in the appropriate direction given $\sigma_C$ across dot scatter levels.
Figure 4.7: Measured vs. predicted $t_C$ means (all participants).
Measured $t_C$ means (y-axis) are plotted against predicted $t_C$ means (x-axis), for all
participants. Each point represents one participant’s $t_C$ mean (with 95\% confidence
intervals) from a CC block with a particular dot scatter level. The various color and style
combinations for the lines correspond to individual participants, as labeled by the legend.
The heavy, solid, diagonal line represents perfect coherence between predicted and
measured $t_C$ means.

Qualitatively all participants adjusted their timing choices in concert with the
ideal performer. A quantitative comparison between participants’ actual $t_C$ choices and
their ideal performers’ predicted $t_C$ choices is shown in Figure 4.7. The means of
measured $t_C$ values were plotted against ideal performer’s $t_C$ values that would
minimize $\sigma_C$, for each participant.
Figure 4.7 confirms that all participants shifted their $t_c$ values in the direction of the ideal predicted shift, but sometimes with a lesser magnitude than predicted (see Discussion). One notable pattern in this figure is that all participants but one (participant 2) show a greater $t_c$ shift between high (rightmost points) and medium (middle points) dot scatter levels versus between medium and low (leftmost points) dot scatter levels. One potential explanation for this pattern is that participants may be less sensitive to their own internal visual target certainty at lower uncertainty levels. Another potential explanation may be an artifact of our methodology: we conducted the high dot scatter level condition several days after the low and medium dot scatter level conditions, whose blocks were interleaved. Perhaps interleaving the low and medium scatter level blocks promoted strategy generalization among the participants in which they chose $[t_L, t_H]$ that were good for both scatter levels.

![Figure 4.8: Measured vs. predicted offset standard deviations (all participants). Measured reach endpoint offset standard deviations (y-axis) are plotted against the predicted offset standard deviations (x-axis).](image)

Figure 4.8: Measured vs. predicted offset standard deviations (all participants). Measured reach endpoint offset standard deviations (y-axis) are plotted against the
predicted offset standard deviations at the predicted \([t_r, t_M]\) choices (x-axis). Each point represents the standard deviations of all offsets (with 95% confidence intervals) from a CC block with a particular dot scatter level. The various color and style combinations for the lines correspond to individual participants, as labeled by the legend. The heavy, solid, diagonal line represents perfect coherence between predicted and measured offset standard deviations.

4.4.1.3 CC condition: task performance

In order to validate our predictive model of \(\sigma_c\) given baseline condition measurements, we compared participants’ measured reach endpoints to their ideal performers’ \(\sigma_c\) function. Figure 4.8 represents all participants’ measured endpoint offset standard deviations plotted against the ideal predicted offset standard deviations. Each measured endpoint offset standard deviations was computed by taking the standard deviation of a participant’s reach endpoint offsets for a particular dot scatter level, and the ideal predicted offset standard deviation is simply the value of the ideal performer’s predicted \(\sigma_c\) at its minimum. Figure 4.9 characterizes the relation of predicted and measured performance by “combination efficiency”. The percentages in Figure 4.9 were computed as follows: for each CC condition reach, the measured reach endpoint offset was divided by the value of \(\sigma_c\) at the participant’s measured \([t_r, t_M]\) choice in that trial. We took the root mean square of these ratios and multiplied it by 100 to compute average efficiency, in percent, of the participant’s performance with respect to the model.
Figure 4.9: Measured reach endpoint accuracy with respect to predictions (all participants). Each triplet of bars is a set of efficiency scores for reach accuracy for a single participant (with +/- 95% confidence intervals). White, gray, and black bars are efficiency scores for the low, medium, and high dot scatter levels, respectively.

Generally, measured performances slightly exceeded predicted performance despite actual timing choices shifting slightly less across dot scatter levels than the ideal prediction. This slight over-performance may be due to increased participant motivation in the CC condition due to the task’s increased difficulty over the baseline conditions (Figure 4.2) and greater allowance for participants’ choices, since they chose $t_v$ and $t_M$ in the CC task as opposed to being cued to particular values of each, as in the MB and VB conditions.

4.4.2 Experiment 2

A final question addressed whether providing direct performance feedback to participants would substantially change their timing choices, perhaps helping them choose $[t_v, t_M]$ nearer to the minimum of their ideal performer’s $\sigma_C$ function.
As described in the Methods section, Exp. 2 provided performance feedback indicating whether reaches were successful or not in the CC condition. If participants’ timing choices remained unchanged between Exp. 1 and 2, we could conclude that participants did not require performance feedback to select $[t_v, t_M]$ with low values of $\sigma_c$.

![Figure 4.10: CC task performance comparison between Exp. 1 and 2 (all Exp. 2 participants). Performances are expressed as percent of total trials in which the participant successfully reached the target. The x-axis represents Exp. 1 performances; the y-axis represents Exp. 2 performances. Each point represents the Exp. 1 and 2 performances for one dot scatter level, for one participant. The ellipses are SEM ellipses in both x and y directions, as computed by bootstrapped resampling of the raw data. The correlation between the points is 0.96 (p<0.00001).](image)

We found no significant difference in either CC task performance or $[t_v, t_M]$ distributions when direct feedback was provided. Figure 4.10 scatterplots CC task performance between Exp. 1 and 2. To quantify performance, we used the percentage of
CC task trials in which the participant successfully reached the target. The x-axis represents Exp. 1 performances, and the y-axis represents Exp. 2 performances. Each point represents the performances for one dot scatter level, for one participant. The correlation between the points is 0.96 (p<0.00001). This supports the view that each participant performed consistently across Exp. 1 and 2.

Figure 4.11: Timing distributional area comparison between Exp. 1 and 2 (all Exp. 2 participants).
Timing distributional area was quantified as the square-root of the trace of the covariance matrix of each dot scatter level’s \([t_V, t_M]\) choice distribution. The x-axis represents Exp. 1 \([t_V, t_M]\) distributional area, the y-axis represents Exp. 2 distributional area. Each point represents the two experiments’ distributional areas for one dot scatter level, for one participant. The ellipses are SEM ellipses in both x and y directions, as computed by bootstrapped resampling of the raw data. The point with the “X” over it (participant 4, low dot scatter level condition) can be considered an outlier.

Figure 4.11 scatterplots timing choices’ distributional areas between Exp. 1 and 2. To quantify distributional area, we used the square-root of the trace of the covariance
matrix of each participant’s 2D \([t_v, t_m]\) distribution. We plotted the distributional areas for Exp. 1 against those for Exp. 2 to assess the consistency of participant's timing choice variability across the two experiments. The correlation between the points is 0.41, which shows participants’ \([t_v, t_m]\) distributions had similar areas, with and without direct feedback. If we consider the point with the “X” over it (participant 4, low dot scatter level condition) an outlier and remove it from this analysis, the correlation rises to 0.67 and becomes significantly different from zero (p<0.05). Because there was little difference between distributional areas in Exp. 1 and 2, we conclude the variability in \([t_v, t_m]\) choices was unrelated to participants’ knowledge of their own performance. Potential explanations for the timing variability are explored in the Discussion.

4.5 Discussion

We interpreted our model and participants’ behaviors in the context of statistical decision theory, whose application to sensorimotor neuroscience stems from a long tradition of treating perception and action as statistical computation problems (Attneave, 1954; Barlow, 1961; Fitts, 1954). Our results suggest that when performing visually-guided motor behaviors, the brain represents both the quality of the visual information and potential motor output. Moreover, the brain understands how visual and motor variability depend on time, and selects viewing and movement durations to minimize consequent errors.

These results are not necessarily surprising in isolation, as previous studies have shown human performance of various sensorimotor tasks reflects key elements of statistically optimal decision-making (Maloney, 2002; Körding and Wolpert, 2006), including near-optimal use of sensory information (Kersten, 1987; Geisler, 1989; Legge et al., 1997; Knill, 1998; Geisler, 2003; Kersten et al., 2004), reliability-weighted sensory information combination (Landy et al., 1995; Knill, 1998; Jacobs, 1999; Ernst and Banks, 2002; Battaglia et al., 2003; Alais and Burr, 2004; Ernst and Bültthoff, 2004; Shams et al., 2005), knowledge of sensory inputs’ generative processes (Knill and Kersten, 1991; Bloj et al., 1999; Battaglia et al., 2005), use of prior information (Mamassian and Landy,
2001; Weiss et al., 2002; Adams et al., 2004; Körding and Wolpert, 2004; Tassinari et al., 2006), internal motor output variability representations (Harris and Wolpert, 1998; Todorov and Jordan, 2002; Todorov, 2004), and selection of gain-maximizing actions (Schrater and Kersten, 2000; Trommershäuser et al., 2003a, 2003b, 2005, 2006a, 2006b; Wu et al., 2006). What is surprising is that these elements cooperate to allow human performance to approach optimal performance for a novel visuomotor task. Minimally, visual estimates and their respective accuracies are passed to motor planning systems, which predict the precision of potential reaches and control viewing and movement timing to favorably trade off visual and motor variability.

Based on participants’ isolated visual (VB) and motor (MB) performances, we predicted timing choices expected to minimize reaching errors in the combined visuomotor condition (CC). Participants clearly adjusted their timing choices in a manner predicted to improve task performance. Because participants chose viewing and movement durations predicted to yield low offsets with nearly zero performance feedback, we conclude that this behavior is not merely an association between timings and success, but in fact an internal representation of task structure. This conclusion is necessary to explain the different timing choice strategies across dot scatter levels. Because Exp. 2’s additional direct performance feedback did not substantially improve (or otherwise modify) participants’ behavior, it seems that inherent knowledge of visual and motor variability are sufficient to support optimal task performance, and little learning is required to optimize that knowledge. A potential future experiment could impose an alternative manipulation of the optimal trade-off time by varying the rate at which new dots appear. If carefully controlled, this manipulation could help expose the relationship between neural integration windows and information accumulation periods for perceptual decision-making (Mazurek et al., 2003; Uchida et al., 2006).

One puzzling feature of our data is the variability in participants’ measured timing choices in the CC condition. Across participants, the standard deviations of $t_M$ measured in the MB condition were consistent with Zelaznik et al. (1988). The average standard deviation of $t_M$ measured in the CC condition across participants was 98.9 ms ($\pm$ 18.0 ms 95% CI) which was 1.8 times (with 1.3 as the 95% lower confidence interval) what
our MB measurements, as well as Zelaznik et al.’s (1988) measurements, predict as participants’ minima. This means participants allowed significantly more (p<.05) variability in movement time, $t_M$, in the CC condition than in the MB condition. We conclude participants did not deem $t_M$ variance minimization to be critically important in the CC condition. Likewise, CC condition viewing times, $t_V$, had high variability, with an average standard deviation of 122.1 ms (+/- 22.7 ms 95% CI) across participants.

We could not concretely explain these high degrees of variability, but there are several possibilities. First, participants’ choices may reflect a principle from learning theory called the exploration/exploitation tradeoff. This holds that when learning to improve task performance, optimal task behavior may be deliberately sacrificed in order to test whether novel behaviors may potentially yield greater performance. In our task, participants may choose novel $[t_V^*, t_M^*]$ values that are not consistent with their estimates of the optimal $[t_V^*, t_M^*]$ choices to investigate whether these novel timings may improve performance. Despite the lack of performance feedback in Exp. 1 beyond the 30-trial cumulative scores, participants may internally monitor their reach performance to provide supervision for learning, akin to “bootstrapped-learning”. A second possibility is that participants have some uncertainty about the exact $t_V$ and $t_M$ combinations that minimize $\sigma$. When people are uncertain which timing is best, they may simply consider a set of timings to be “good enough”, and thus explicitly allow $[t_V, t_M]$ choices to vary within that set. This is qualitatively consistent with the “minimum intervention principle” posited by Todorov and Jordan (2003). A third possibility is that because $\sigma$ have relatively flat minima, the predicted difference in monetary reward is so small, perhaps as small as a few cents, that the cost of controlling $[t_V, t_M]$ choices is higher than the small payoff such control may yield.

A potentially related issue was the smaller-than-predicted timing shifts across dot scatter levels. One possibility again relates to participants devoting some trials to task exploration, as described above. In this case, the extreme dot scatter levels (e.g. low and high) would have some $[t_V, t_M]$ choices distributed toward the “middle” of the available
timing range, thus lowering the measured shifts across scatter levels. A second possibility involves a potential mismatch between the assumptions in the ideal reacher model and human behavior. Ideal timing choices contain no temporal scatter and have no restrictions on $t_M$ (e.g. $t_M = 0$ is possible). Perhaps in the presence of temporal scatter and additional unmodeled costs associated with extremely rapid movements, the range of admissible $t_M$ choices was reduced, consequently reducing the range of $t_V$ choices as well. A third possibility is that again, because of the relatively flat minimum of $\sigma_C$, participants may not expect sufficient reward to warrant fully changing their choice strategy with dot scatter level.

One limitation of our model is that it only places cost on missing the target. The model does not acknowledge any cost for timing out. Thus our model’s optimal $[t_V^*, t_M^*]$ choices always lie on the total trial time axis (slope: -1, y-intercept: 1200 ms). Timing choices that lie on the total trial time axis imply that the participant used all 1200 ms of the total trial time (i.e. $t_V + t_M = 1200$ ms). This means that if the participant’s $t_V$ or $t_M$ lasted even 1 ms longer, the trial would time out. Since people cannot control the duration of $t_V$ or $t_M$ to within 1 ms, the optimal decision is to shorten $t_V$ and/or $t_M$ to avoid timeout costs. In our experiment, participants often shortened $t_V$ and/or $t_M$ such that $t_V + t_M < 1200$ ms. We believe implicit timeout costs explain why many timing choices are displaced away from the total time axis. Although we did not impose a monetary penalty for timeouts, and each timeout trial is repeated later, it is reasonable to assume participants would prefer to avoid repeating timeout trials. We applied simple estimates of timeout costs to our ideal performer model, but the estimates were inaccurate and participants’ consistency with these timeout-penalized optimal timing choices only improved marginally, so we did not report those analyses.

Though our experiment was purely psychophysical, neurophysiological studies support the existence of neural representations and computations required for optimal task performance, including time, probability, and reward-driven decision-making (Schall, 2001; Platt, 2002; Sugrue et al., 2005). Recent studies have provided evidence
for temporal probability representations in lateral intraparietal (LIP) areas and parietal cortex in general. Leon and Shadlen (2003) reported evidence of neurons in macaque LIP that code the value as well as uncertainty of a remembered temporal duration (or at least monkeys’ judgments of such quantities) while Janssen and Shadlen (2005) reported evidence of macaque LIP neurons encoding the probability of an event’s occurrence as a function of time. These results highlight potential cortical substrates for processing components our task requires, specifically representing trial time to decide when, and for how long, to execute reach movements.

In our task, the brain represents visual and motor uncertainty relationships and combines them to perform a joint task. A natural question is whether such behavior relies on cortically separate visual and motor representations, or a unified representation of visuomotor uncertainty. Compelling arguments exist for both of these views. Coordinate transformations, temporal syncing, and the propagation of task goals to individual visual and motor decisions may be better served by a unified representation of visuomotor uncertainty. Conversely, independent noise corruption, sensory or motor recalibration, and general organizational simplicity may favor separate representations (see Pouget et al., 2002 for a review of similar issues).

In conclusion, our results supported the view that people can represent their visual and motor variability as functions of time. Moreover, they can combine these components to predict their performance in a task that depends on both, and select viewing and movement durations to minimize reaching errors. This behavior is consistent with Bayesian Decision Theoretic performance of visuomotor tasks.
5. APPENDIX A

5.1 Visual variability model

We assumed that participants computed $\hat{X}$ by estimating the centroid of the dot positions by taking their mean. The overall deviation of $\hat{X}$ from $X$ was given by $\varepsilon_v$, as mentioned, but this term can be further split into two discrete sources of error such that:

$$\varepsilon_v = \varepsilon_\mu + \varepsilon_e$$

$\varepsilon_v$ has mean equal to 0, and variance $\sigma_v^2$ which is the sum of the variances of $\varepsilon_\mu$ and $\varepsilon_e$ (by conditional independence) such that:

$$\sigma_v^2 = \sigma_\mu^2 + \sigma_e^2$$

Equation 5.1

The first source of error, $\varepsilon_\mu$, results from deviations between the mean of the dot positions, $\mu_d(t_v)$, and $X$, due to randomness in sampling the dot positions. Because the dot positions were normally distributed, $\varepsilon_\mu$ was normally distributed with mean 0 and standard deviation $\sigma_\mu$. Formally, $\sigma_\mu$ was a function of viewing time $t_v$, and dot scatter level $\sigma_d$ such that:

$$\sigma_\mu(t_v, \sigma_d) = \frac{\sigma_d}{\sqrt{N(t_v)}}$$

Equation 5.2

using $N(t_v)$ from Equation 5.1. Notice there are no free parameters to fit from data.

The second source of uncertainty, $\varepsilon_e$, results from participants’ misestimates of individual dots’ positions, which we assumed were corrupted by independent, mean 0, normally distributed positional uncertainty. Thus $\varepsilon_e$ was normally distributed with mean 0 and standard deviation $\sigma_e$. For simplicity, $\sigma_e$ was assumed to be a first-order linear function of $\sigma_\mu$ and thus dependent on $t_v$ and $\sigma_d$ as well such that:

$$\sigma_e(t_v, \sigma_d) = \rho_e \cdot \sigma_\mu + \omega_e$$

Equation 5.3

where $\rho_e$ and $\omega_e$ were free parameters.

By substituting Equations 5.2 and 5.3 into 5.1, $\varepsilon_v$ has mean equal to 0, and variance $\sigma_v^2$ such that:
\[ \sigma_v(t_v, \sigma_d)^2 = \left( \rho_e^2 + 1 \right) \frac{\sigma_d^2}{N(t_v)} + 2 \rho_e \omega_e \frac{\sigma_d}{\sqrt{N(t_v)}} + \omega_e^2 \]

as given in Equation 5.3 in the Methods. We fit the free parameters, \( \rho_e \) and \( \omega_e \), by maximum likelihood estimation (MLE) using the VB data separately for each participant, for each dot scatter level, as described below.

### 5.2 Motor variability model

We assumed that errors between \( \hat{X} \) and \( Z, e_M \), were due to motor noise, and normally distributed, with mean 0, and standard deviation \( \sigma_M \). Classically, Fitt’s law expresses a relationship between average movement duration and target width that has several variants (MacKenzie and Buxton, 1992). It has been used to model reach endpoint standard deviations as a function of movement durations (Schmidt et al., 1978; Harris and Wolpert, 1998). We modified the simplest form of the equation:

\[
T = a + b \cdot \log\left( \frac{D}{W} \right)
\]

where \( T \) is expected movement duration, \( D \) is target distance, and \( W \) is target width. \( a \) and \( b \) are free parameters that vary across tasks and participants.

In particular we characterized \( \sigma_M \) as a function of \( t_M \), or \( \sigma_M(t_M) \). We related \( W \) to \( e_M \) by observing that 95% of reach endpoints will lie within the target width when:

\[
\sigma_M(t_M) = \frac{W}{1.96}.
\]

Note that we have equated the expected movement time, \( T \), with the movement time, choice, \( t_M \). To formulate endpoint variability as a function of \( t_M \), we solve the above expression for \( W \), replace \( W \) by \( \frac{\sigma_M(t_M)}{c} \), where \( c \) appropriately scales the endpoint variability to match success criterion for acquiring the target (i.e., the target width). Also, we add a constant, \( \gamma \), that represents the minimum achievable movement endpoint offset (\( \gamma \) can be thought of as the offset achievable if given a very long time to complete the reach). The resulting expression is fit to individual participant data:

\[
\sigma_M(t_M) = D \cdot \exp(\alpha - \beta t_M) + \gamma,
\]

as given by Equation 5.4 in the Methods, where \( \beta = \frac{1}{b} \), \( \alpha = \frac{a}{b} + \log(c) \) absorbs the
effects of \( c \). We fit the free parameters, \( \alpha \), \( \beta \), and \( \gamma \) using MLE as described below.

### 5.3 Maximum likelihood estimation of model parameters

Maximum Likelihood estimates (MLE) of model parameters were separately fit for each participant using the baseline condition data. The log-likelihood of the data was computed for each reach by evaluating the probability of each trial’s offset using the model, evaluated at various model parameter values. Those model parameters that produced the maximum likelihood were considered to be best-fit parameters. For example, the likelihood of \( \{\alpha, \beta, \gamma\} \) is given by the following expression, where \( \phi(\cdot) \) is a Gaussian density function, \( Y_i \) is the offset for the \( i^{th} \) reach, and \( k \) is the total number of reaches:

\[
L(\alpha, \beta, \gamma) = \sum_{i=1}^{k} \log \left[ \phi(Y_i; 0, D \cdot \exp(\alpha - \beta \cdot t_M) + \gamma)^2 \right]
\]

The likelihood for the visual data is similar. These expressions were numerically maximized to find optimal parameters.

### 5.4 Bootstrapped confidence intervals/SEMs

Reported results are accompanied by 95% confidence intervals or SEMs. When necessary for predictions, these confidence intervals or SEMs were computed by bootstrapped resampling of the raw data (Efron and Tibshirani, 1994). Specifically, we sampled the original data with replacement 50 times (or more if computational costs were not prohibitive) and performed the reported analysis on all 50 resampled data sets. From the set of 50 results, we computed the mean and standard deviation; the mean is the value reported and the bootstrap standard deviation represents the SEM. 95% confidence intervals were computed as 1.96 times the SEM.
6. Bibliography


Goodale MA (1998) Visuomotor control: Where does vision end and action begin? 
*Current Biology* 8(14): R489-R491.


Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-
conflict studies with stereoscopic and figure cues to slant. *Journal of Vision* 7(7):
1-24.


and computation for perception and action. *Trends in Neuroscience* 27(12): 712-
719.

University Press.

Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture


of Neurophysiology* 92(5): 3161-3165.

427: 244-247.


