Statistical decision theory for human perception-action cycles

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

Erik J. Schlicht

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

Paul R. Schrater & Dan Kersten, Co-advisors

May, 2007

Acknowledgements

I would like to thank Paul Schrater for being an outstanding mentor and friend. His guidance in both academic and personal domains was invaluable. I would also like to thank Dan Kersten and Gordon Legge for their insightful feedback and suggestions on several projects and topics. In addition, I appreciate Sheng He and Michael Wade taking time from their busy schedules to serve on my thesis committee. Finally, I would like to thank Chris, Pete, Nick, Mike and Charlie for providing welcome distractions from my graduate training.

Dedication

To Kari. Your love provides me with purpose and shows me the significance of life beyond academic pursuits. Without you, this would be an empty accomplishment.

To my Mother. Your unconditional love and support has made all of this possible. I would never have been able to endure this process without your help.

To my Father. Thank you for instilling in me the work-ethic and dedication necessary for completing such an intensive project.

In memory of my Grandparents. Thank you for looking-over and guiding me throughout my life.

Abstract

The goal of statistical decision theory (SDT) is to select optimal actions that maximize an agent's expected gain or minimize an agent's expected risk, in the presence of uncertainty. Since most human behavior is goal directed, this paper argues that SDT can be used to provide an understanding of natural movements. However, if people are acting to minimize their expected risk, it necessitates the brain has knowledge of the system's uncertainty, in addition to the loss function that is associated with the task. Therefore, the first part of this paper psychophysically investigates if humans plan for uncertainty when making grasping movements, while the second portion of the paper attempts to reverse-engineer a natural loss function that can predict the contact locations people use when making a grasp. As a means to explore if the brain is aware of its uncertainty, subjects were required to repeatedly reach to the same object while gaze direction was varied, both when the target was in view and when it was occluded. The results show that observers' reach parameters systematically vary as a function of the expected optimal uncertainty. Moreover, it is demonstrated that these changes in reach behavior are predicted by a Bayesian model that quantifies the uncertainty associated with the sources of information in the task: visual, haptic, and eye position information. The second portion of this paper outlines a natural loss function that can predict the contact locations people use when making a grasp. The loss function is based on the physics of object manipulation and assumes that people place their fingers in locations that provide accurate object motion with minimum force. The model predictions were tested by requiring subjects to reach an object at varying orientations, and the results indicate that people are reaching in a manner that minimizes their expected risk for the task. Overall, these results demonstrate that people are acting in a manner consistent with the theoretical principals of statistical decision theory. Therefore, it appears that SDT is a useful tool that can be used to investigate natural human behavior.

Contents

List of Figures

Chapter 1

Introduction to Statistical Decision Theory

Our perceptual system gathers information through several different sources (e.g., vision, audition, haptic - touch). Humans use these sources of information to recreate the scene from which the perceptual data was derived. This is done so effortlessly that it seems trivial; however, the ease with which we carry out this operation disguises its underlying complexity. For example, if we look at a cup of coffee on our desk and wish to grab it, we must turn the 2D visual input from our retina into a 3D representation of the cup and its surroundings. This is called the *inverse vision problem*, as there are an infinite number of 3D objects that can result from any given 2D image [56]. Therefore, what we are doing is *estimating* the type and location of the objects in our environment. These estimates are not perfect - errors are introduced both by the insufficiency of perceptual information and neural processing. This makes the job of perception to *infer* the state of nature in the presence of uncertainty, given the data and our prior knowledge. Since perception involves making decisions in the presence of uncertainty, it lends itself well to statistical decision theory (SDT). There have been numerous studies that have successfully used SDT to better understand how the brain uses perception to arrive at estimates (e.g., [5, 31, 62, 66, 110, 127]). This work has provided valuable insight into how our perceptual system uses sensory information to make decisions.

Similar to the inverse vision problem, there are also an infinite number of ways to

which we can reach and grasp an object. Despite this fact, people tend to reach and grasp objects in an extremely effortless and predictable manner. This suggests that the brain is using some sort of strategy (i.e., loss function) to reduce the ambiguity in the reaching task. This paper argues that SDT can also be used to provide insight into how people perform natural movements. As a result, a major goal of this effort is to reverse-engineer a natural loss function that can predict the contact conditions (i.e., finger locations and velocities) people use when making grasping movements. However, before a loss function can be detailed, some important concepts in sensorimotor control and SDT will be provided in the next two sections.

1.1 Concepts in Sensorimotor Control

1.1.1 Feed-Forward Versus Feedback Control

For the purpose of this thesis, a *perception-action cycle* is defined to be any behavior for which sensory information is used to plan, guide and correct goal-directed actions. Inherent in this definition is the distinction between *feed-forward* and *feedback* control of movements (Figure 1.1).

Forward models can be thought-of as predictions that our brain makes about the consequences of executing a particular force or movement. They are useful in many situations such as state and context estimation, and sensory cancelation [124]. In state estimation, the brain must know about the positions and velocities of body segments. This can be accomplished by using sensory feedback (e.g., visual and proprioceptive) or by making predictions about the future state of a body segment (e.g., hand) based on the motor command sent to the muscles (i.e., efference copy). The advantage of using forward models is that the information is readily available, whereas sensory information is delayed and degraded by noise. The drawback of using a purely feed-forward estimator is that the system's predictions can be inaccurate and often drift over time/space. Therefore, a better solution for state estimation problem is to to begin movements with forward model predictions and then update the state with feedback gained through sensory information during the movement. In fact, it appears that humans are performing actions in a manner consistent with these ideas as there have been studies in sensorimotor control using Kalman Filters to model such behavior [46, 94].

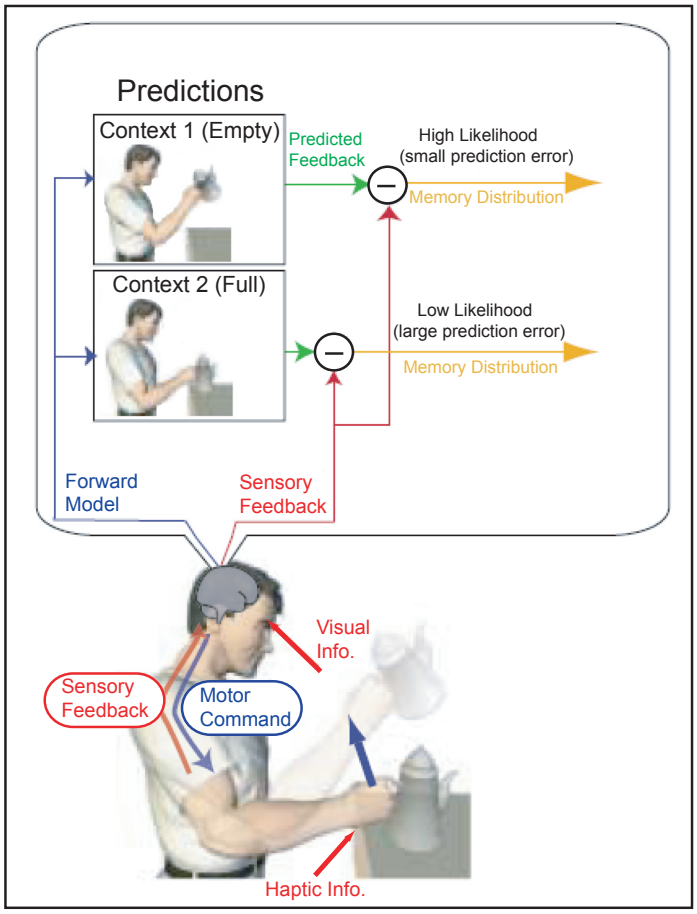

Figure 1.1: Figure adapted from [124]. Diagram illustrating a perception-action cycle. See text for details.

In addition to state estimation, forward models are often useful in sensory cancelation. For example, imagine attempting to jar ketchup loose by smacking the bottom of the bottle. In this example, it is much easier to hold the bottle of ketchup steady when the forces are self-generated (i.e., same person holding and hitting bottle) than if someone else were to deliver the blows. This ability is afforded by the brain using forward model predictions to adjust the grip force of the hand that is holding the bottle. Since self-generated forces can be predicted more accurately than external forces, the result is that the ketchup bottle is more stable when it is held and contacted by the same individual. Indeed, it has also been suggested that these same principals explain why it is more difficult to tickle oneself compared to when someone else tickles you [14].

Another application of forward models in sensorimotor control is that of context estimation (Figure 1.1). In order for an individual to grasp an object (e.g., teapot), their

fingers must be brought to locations that generate and maintain appropriate forces on the object, subject to kinematic constraints imposed by hand and object geometry. Moreover, object manipulation requires knowledge about the physical properties of the object (e.g., mass, center of mass, moments of inertia, surface friction) that are *estimated* through sensory information. For example, Figure 1.1 shows an individual attempting to lift a teapot. In this example, the forward model is estimating the context under which the reach is being performed (i.e., teapot full or teapot empty). Notice that if the teapot were full, it would be beneficial to contact the object with more momentum and grasp the object with greater pinch force than it it were empty. Before a reach is made, the brain simulates each of these possible forward models (blue lines) to compute the predicted feedback under each context. Then, sensory feedback (red lines; visual & haptic information) is compared to each of the forward model predictions to assess which is more likely (Empty, in this example). This information can then be stored and used for future interaction with the object, making it less likely for prediction errors (green lines) to occur over time (yellow lines). Note that this entire process is an example of a perception-action cycle.

This subsection has overviewed the distinction between feed-forward and feedback control of movements. The next subsection will outline some of the major challenges faced by the brain when making natural movements.

1.1.2 Challenges in Sensorimotor Control

In addition to the fact that there are an infinite number of ways that we can reach to and grasp an object, there are many other factors that make reaching a complex problem. For example, the information required by the system to make a successful grasp is not always directly observable. As mentioned above, many of the physical properties of the object cannot be perfectly measured. Instead, the system must rely on sensory feedback to *estimate* the properties of interest. In fact, most parameters that the system requires to perform a successful grasp are not directly observable. Therefore, sensory information plays a crucial role in estimating these parameters of interest.

Since sensory information plays a key component in human perception-action cycles, it is important to note that a second challenge in sensorimotor control arises from the fact that perceptual information is imperfect (e.g., [17, 120, 69]). Similar to how

sensory uncertainty limits our ability to perceive the world, motor variability (e.g., [46]) decreases our accuracy when acting upon targets. As a result, uncertainty in our perception and actions is something the brain must consider in order to make efficient and accurate movements. Indeed, there have been numerous studies investigating how the brain uses uncertainty to perform movements (e.g., [46], [63]), and this paper will extend this work by exploring how knowledge of visual uncertainty (Chapter 2) and coordinate transformation uncertainty (Chapter 3) are used when making grasping movements.

Once the parameters important for reaching have been estimated, an additional challenge surfaces - how does the brain select finger locations to contact object? Finger locations determine the ease at which someone can manipulate or stabilize the object. This problem is difficult as some finger arrangements should be used for stabilizing an object, where others are better for manipulating the object. What's more, is that some manipulations may require different finger arrangements than others (e.g., handing a knife to someone versus cutting with the knife). For all of these reasons, finger location selection is an extremely challenging problem that remains largely unsolved. However, Chapter 4 of this paper will propose a theoretical framework that can predict the finger locations people should use when making different grasping movements.

A final challenge in sensorimotor control is that different actions may have differential rewards/losses that correspond to them. For example, consider the case of someone placing a cookie sheet into the oven. If the sheet is placed into the oven while it's still cool, there is relatively little cost associated with colliding with the oven rack. Conversely, when taking the cookies out-of the hot oven, there is now much greater cost associated with contacting the rack. An intelligent system would compensate for these differences by acting more conservatively when there is greater cost associated with making an error. To that end, Chapter 4 demonstrates how contact locations vary according to the risk associated with the reaching task.

For all of these reasons, challenges in sensorimotor control are abundant and complex. Since sensorimotor control deals with planning and acting in uncertain environments under differential levels of reward/loss, this paper argues that statistical decision theory (SDT) can be used to provide solutions to these challenges. Before efforts done to this end are described, the next section will first overview some important

concepts in SDT.

1.2 Concepts in Statistical Decision Theory

1.2.1 Signal Detection Theory for Perception

A specialized area of SDT called signal detection theory has a long tradition in the study of human perception (e.g., [8, 60, 61, 67, 73, 96]). Signal detection theory not only has ties with SDT, but also with information theory [98], and is concerned with the process of detecting signals in the presence of noise. This is important for human perception because it is known that our perceptual and motor systems have noise attached to them (e.g., [69, 46]). Signal detection theory provides a formal framework from which to investigate how humans arrive at decisions in the presence of perceptual noise.

There are four ideas that are essential to signal detection theory [68]. First, the notion of *sensitivity* can be thought of as the ability of the system to detect a signal. If the signal-to-noise ratio (d') is high, then the system can better detect signals than when it is low. For example, in a classroom setting it is difficult to hear the instructor (i.e., signal) if your classmates are talking (i.e., noise). Signal detection theory provides a way to quantify this intuition.

A second important component of signal detection theory is that of *efficiency*. Signal detection theory allows the information in the task to be quantified, enabling researchers to theoretically derive optimal performance. The use of ideal observer models have a lengthy history in the study of human perception. For example, early use of an ideal observer by [8] compared human photon detection to an ideal observer model. They found that humans were suboptimal in photon detection as they needed about ten-times the number of photons as the ideal model. Ideal models have also been compared to humans for slant discrimination from texture [60, 61], reading [67], object classification [73], and motion detection [96]. These models have provided valuable insight into how efficiently humans use perceptual information to complete various tasks.

The third component of signal detection theory is the *response threshold*. The threshold can be intuited as a person's tendency to report detecting a signal. Notice that thresholds tend to change from person to person, even with constant sensitivity.

Moreover, individuals can change their own threshold across time. For example, a pathologist may be more likely to diagnose disease after looking at a specimen on one day compared to the next. As a result, thresholds can change both across and within people, over time.

A final component of signal detection theory is called *efficacy* and is concerned with the consequences of making an incorrect decision. For example, in medical diagnostics it's more costly to miss a diagnosis of cancer if someone has the disease than to incorrectly test positive. The consequences of making incorrect decisions can be formalized through loss functions, which will be described in detail later in this chapter.

Despite the success of signal detection theory in investigating human perception, there are limitations inherent to this theory. For example, classic signal detection theory assumes linearity [58]. That is, it assumes that the input is the sum of the signal and noise. However, in natural tasks this is seldom the case.

A limitation in experimental approaches to signal detection theory is that the prior probabilities and loss functions are usually manipulated by the researcher [58]. Ideally, the prior should be determined naturally through image/task statistics and loss functions should result from the task. This paper will attempt to remedy this limitation by proposing a loss function that directly results from the task demands.

These above limitations bound the utility of signal detection theory to simple and/or artificial tasks. In order to model more complex and natural behaviors, robust techniques from SDT must be employed. The next subsection will overview general concepts of SDT and relate them to a classification example. The following subsections will contrast differences between maximum likelihood and Bayesian approaches to SDT.

1.2.2 Statistical Decision Theory

In general, statistical decision theory provides a formal method for determining optimal actions that either maximize an agent's expected gain, or minimize an agent's expected risk, in the presence of uncertainty. SDT has traditionally been used in areas such as economic modeling (e.g., [30, 53]) and game theory (e.g., [83]). More recent applications of SDT include medical science (e.g.,[38, 75]), robotics (e.g., [108, 55]), pattern recognition (e.g., [83]), and human perception (e.g., [5, 31, 62, 66, 110, 127]). This

paper maintains that SDT is well suited for better understanding human perception-action cycles. Figure 1.2 shows an example of how a loss function can be derived from a natural task (i.e., pressing an elevator button). From this example, it is clear that optimal actions for humans should be those that minimize the expected risk (or maximize the expected gain), since there is uncertainty involved with our perception and actions (e.g., [69, 46]). The rest of this section will formally distinguish loss and risk, in addition to overviewing differences in Bayesian and maximum likelihood approaches.

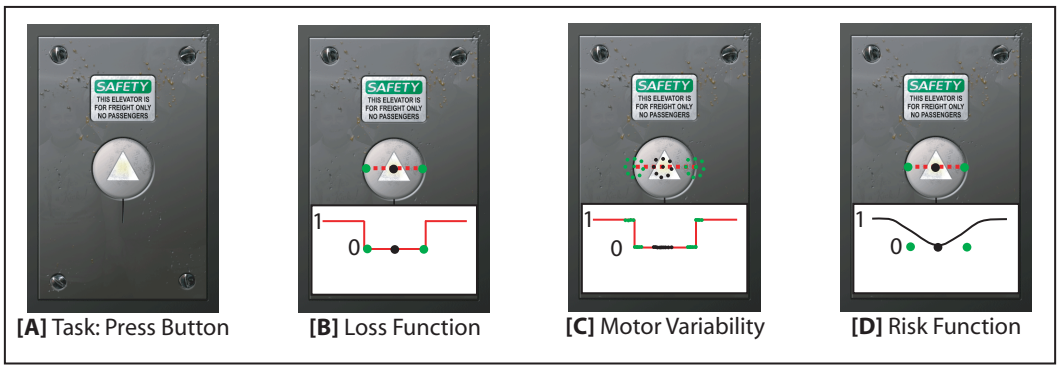

Figure 1.2: Diagram illustrating how SDT can be applied to a simple task, such as pressing an elevator button [**A**]. [**B**] The mid-line slice (red dashed line) of the button is only considered in this example. The button-pressing task is most easily represented by a heaviside loss function, such that the agent receives no loss for contacting the button (valley of loss landscape), and a loss of one is accrued for missing the button (ridge of loss function). To successfully complete the task, the agent must minimize the loss (i.e., contact the button). However, it should be noted that as far as the loss function is concerned, peripheral strategies (green dots) are equally good as a central strategy (black dot). [**C**] This assumption is problematic since the human motor system has signal dependent noise, and repeated attempts to contact the central and peripheral strategies will result in a distribution of points. Notice that some of the points on the peripheral strategies fall outside of the loss function's minimum (i.e., they miss the button). Therefore, loss functions are not optimal for agents with uncertainty in their perception and/or actions. [**D**] Risk functions take into account the agent's uncertainty. Therefore, optimal actions for agents with uncertainty should minimize the *risk* associated with the task, not the *loss*.

Formally, uncertainty results from our lack of understanding of some unknown parameter θ that may affect the decision being considered. You can think of θ as the *state of nature*, and Θ as all the possible states of nature. In that respect, θ is the *parameter* and Θ is the *parameter space*. In SDT, decisions are commonly referred to as actions a, such that $a \in A$. The goal of SDT is to find the action that minimizes a loss function:

$$d^* = \arg\min_a E[L(\theta, a)], \tag{1.1}$$

where $L(\theta, a)$ is the *loss function* that is defined for all $(\theta, a) \in \Theta \times A$, and d^* is the *optimal decision* - the one that minimizes the loss function.

Loss functions are important because in certain situations the consequences of erring in a decision are drastically different. For example, when approaching a stoplight while driving, if you decelerate too slowly you'll crash into the car in front of you. On the other hand, if you decelerate too quickly you may be hit from behind. The loss function for braking your vehicle in this example would have extreme negative values and appear quadratic, with a minimum around the optimal deceleration location. Conversely, the loss function for breaking for a small bump in the road is likely not to be as costly, and may even be ignored.

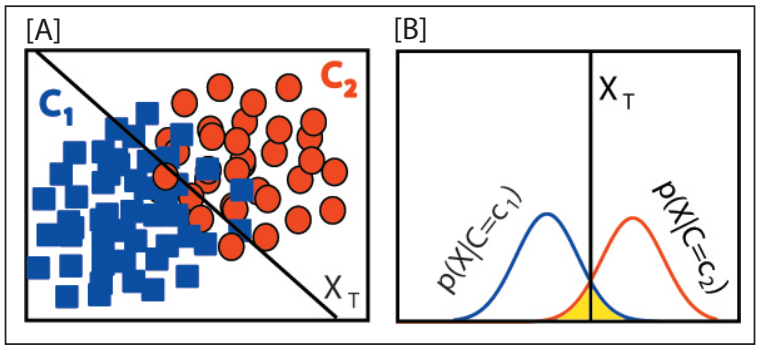

Figure 1.3: Figure of a classification task. **[A]** Demonstrates two different classes of objects (c_1 and c_2). The optimal decision boundary x_T (diagonal black line) is the decision criterion that minimizes classification error (yellow). **[B]** The likelihood distributions for each class ($p(x|C = c_1)$ and $p(x|C = c_2)$). It is apparent that x_T minimizes the classification error (yellow).

In order to learn more about the state of nature θ we make observations (x_i), such that $x_i \in X$. Many times our data (D) is made-up of several sets of observations, such that $D = \{X_1, X_2, \ldots, X_n\}$. Our observations allow us to develop a likelihood distribution $p(D|\theta)$ that tells us how likely the data is given the state of nature. For example, if our task is to decide whether an object belongs to one class c_1 or another c_2, then we are doing a classification task (Figure 1.3). Suppose we make several observations $X = \{x_1, x_2, \ldots, x_n\}$ and must decide which observations belong to c_1 or c_2. Figure 1.3 shows that making errors in the classification task is unavoidable as there is overlap in the distributions. Our goal is to derive a decision boundary (i.e.,

classifier) that minimizes the loss function for this task ($L(\hat{c}, c) = 1 - \delta_{\hat{c},c}$), where $\delta_{\hat{c},c}$ is a Kronecker delta such that it assigns a one to a correct decision ($\hat{c} = c$) and a zero otherwise. Remember, the goal of this task is to find the *optimal* decision boundary for classifying the objects (i.e., minimize the classification error). It is important to understand that the optimal decision boundary may change as a function of the approach used. There are two approaches to optimality have been commonly used in SDT: maximum likelihood approaches (ML) and Bayesian approaches. The next two subsections will contrast ML and Bayesian approaches to SDT.

1.2.3 Bayesian Approaches

Bayesian analysis allows us to make informed decisions given our prior knowledge $\pi(\theta) = p(\theta)$ about the state of nature, θ. For example, in an object recognition task we use our knowledge of what objects are more likely given the room we are currently occupying [82]. If it is known that we are occupying a classroom, then we wouldn't expect to see a car in the room. On the other hand, it would be very likely that a desk would be present. Therefore, we are using our prior knowledge to make an informed estimate as to the identity of objects in a given location. Bayesian analysis allows us to formally integrate our prior knowledge about the state of nature. Bayes' Theorem takes on the following form:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}, \tag{1.2}$$

where $p(\theta|D)$ is called the *posterior distribution* - the probability of the state of nature given the data, $p(D|\theta)$ is the *likelihood* - the probability of the data given the state of nature, and $p(\theta)$ is the *prior distribution* - the probability of the state of nature in the absence of any observations.

Equation (1.2) shows how the posterior is derived in a static case. When it's necessary to consider how information is updated in a dynamic environment, then Bayesian learning provides a useful framework. Bayesian learning provides a way to quantitatively update and store experience across time. Given the conditional independence of the data ($p(D|\theta) = \prod_i^N p(X_i|\theta)$), then Equation (1.2) can be recursively defined as:

$$
\begin{aligned}
p_t(\theta|D) &= \frac{p_t(D|\theta)p_t(\theta)}{\int p_t(D|\theta)p_t(\theta)d\theta}, \\
&= \frac{p_t(D|\theta)p_{t-1}(\theta|D)}{\int p_t(D|\theta)p_t(\theta)d\theta}
\end{aligned}
\tag{1.3}
$$

Notice that the posterior $p_{t-1}(\theta|D)$ at time $t-1$ is defined to be the prior $p(\theta)$ at time t. This technique is important for formalizing how experience is accumulated across time.

Bayesian techniques can easily be integrated into SDT by allowing the *prior*, $\pi(\theta)$ to influence the loss function. This is accomplished by adjusting our loss by the posterior. It follows that the *posterior risk* of an action a is

$$
\begin{aligned}
R_p(\pi, a) &= E^{\pi}[L_p(\theta, a)] \\
&= \int_{\Theta} L(\theta, a)p(\theta|D)d\theta \\
&= \int_{\Theta} L(\theta, a)\frac{p(D|\theta)p(\theta)}{p(D)}d\theta \\
&= \int_{\Theta} L(\theta, a)\frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}d\theta
\end{aligned}
\tag{1.4}
$$

Equation (1.4) is essentially the expected risk for making an action a for fixed data D. As a result, Posterior risk takes into account our *prior knowledge* of the state of nature θ. Notice that the notion of risk differs from loss in that it assumes uncertainty in θ. This is important because in human perception and action there is uncertainty attached to our estimates of the state of nature.

Another concept that needs to be defined is called *Bayesian risk*. Similar to posterior risk, Bayesian risk takes into account the uncertainty about the state of nature θ through the posterior distribution $p(\theta|X)$. The resulting Bayesian risk expression is:

$$
\begin{aligned}
R_b(\pi, a) &= E^{\pi}[R_b(\theta, a)] \\
&= \int_{\Theta} R_b(\theta, a) d\theta \\
&= \int_{\Theta} \int_{D} L(\theta, a) p(\theta|D) dD d\theta \\
&= \int_{\Theta} \int_{X_1} \int_{X_2} \dots \int_{X_n} L(\theta, a) p(\theta|X_1, X_2, \dots, X_n) dX_1 dX_2 \dots dX_n d\theta
\end{aligned}
\tag{1.5}
$$

From Equation (1.5) it follows that Bayesian risk is the theoretical average loss across all possible data sets. In the case where the loss function for the task can be defined as $L(\theta, a) = \delta(\theta - a)$, then the posterior risk can be used for a MAP decision criterion:

$$
\arg\min_a R(\pi, a) = \arg\min_a \int_{\Theta} L(\theta, a) \frac{p(D|\theta) p(\theta)}{p(D)} d\theta \tag{1.6}
$$

If we assume that $p(D|\theta)$ is *i.i.d.* and marginalize over θ to get $p(D)$, we obtain

$$
\arg\min_a R(\pi, a) = \arg\min_a \int_{\Theta} L(\theta, a) \frac{p(\theta) \prod_i p(X_i|\theta)}{\int p(\theta) \prod_i p(X_i|\theta) d\theta} d\theta \tag{1.7}
$$

Notice that we can disregard the normalizing factor $p(D)$, when finding the minimum. In the context of the classification example provided above, the Bayesian decision criterion would be based on the posterior estimate. Since the loss function for the above example was $L(\hat{c}, c) = 1 - \delta_{\hat{c},c}$, a MAP approach would state that c_1 should be selected if $P(c_1|X) \geq P(c_2|X)$ and c_2 otherwise. However, in order to obtain the posterior distributions, we need to estimate the likelihood and prior distributions. In this example, the prior distributions can be assumed by what is known about the classes and the likelihood distributions are assumed to be Gaussian:

$$
p(x|c_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/2\sigma_1^2} \tag{1.8}
$$

$$
p(x|c_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x-\mu_2)^2/2\sigma_2^2} \tag{1.9}
$$

Gaussian distributions can be summarized by two parameters - their mean and variance (μ, σ^2), respectively. The task in Bayesian decision theory is to perform density estimation on the likelihood and derive the prior in order to obtain the posterior. As a result, Bayesian approaches use a distribution on the classes $(P(C))$ and not just a single value:

$$\begin{aligned} p(c_i|X) &= \frac{p(X|c_i)P(c_i)}{p(X)} \\ &= \frac{p(x_1, x_2, \ldots, x_n|c_i)P(c_i)}{p(x_1, x_2, \ldots, x_n)} \end{aligned} \tag{1.10}$$

In MAP, the classification task can be rephrased such that c_1 should be selected if the observation x is greater than the decision threshold (x_T), which can be found if the following is solved for x [127]:

$$\log \frac{p(x|C = c_1)P(c_1)}{p(x|C = c_2)P(c_2)} = 0 \tag{1.11}$$

Since the likelihoods are Gaussian, x_T becomes:

$$\left(\frac{1}{2}\right)(\mu_1 + \mu_2) + \frac{\sigma^2}{(\mu_1 - \mu_2)} \log \frac{P(c_2)}{P(c_1)} \tag{1.12}$$

From the above example, it's easy to see how Bayesian decision theory uses decision thresholds that take into consideration prior knowledge about the state of nature. This results in a distribution on the parameter space allowing posterior and Bayesian risk to capture the variability of the parameters. In general, Bayesian Decision Theory has a few qualities that are important to keep in mind.

1.2.3.1 Important Features of Bayesian Decision Theory

- First, it should be mentioned that the variance of the posterior $(p(\theta|D))$ typically indicates the reliability of the estimate. However, in most cue combination

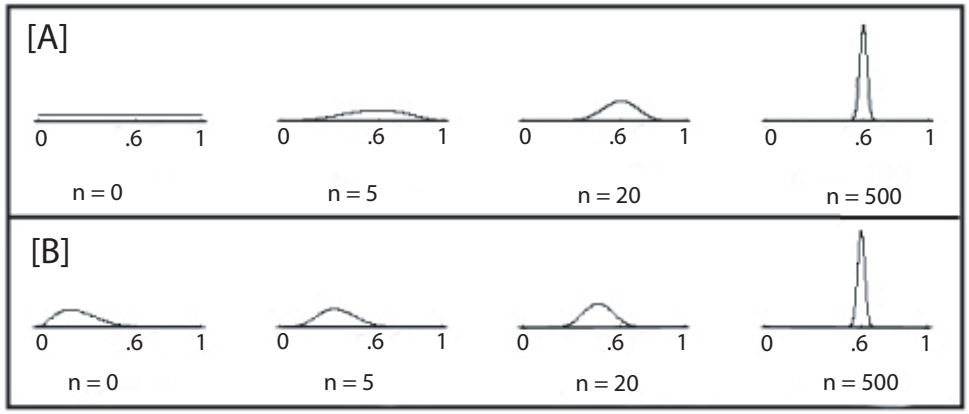

Figure 1.4: Figure adapted from [27]. Effect of prior choice over n observations. **[A]** It is assumed that the prior is uniform (similar to ML). You can see that it starts to center around the true estimate (.6) between 5-20 observations. By 500 observations, it is strongly spiked at the true value. **[B]** Here a prior of .2 is assumed. By 20 trials, it's difficult to distinguish from the uniform assumption ([A]). After 500 trials there is no difference between the two, suggesting for a large n the choice of prior is insignificant.

research the likelihood is used to assess reliability (e.g., [31, 32, 60, 66]), which is sufficient in large n cases.

- The second quality of a Bayesian approach is that as the number of observation sets X approaches ∞, observations have more of an impact on the posterior than the prior distribution, $\pi(\theta)$. This means that selecting a prior (in the large n case) is an arbitrary decision as the posterior will still converge to the true distribution (Figure 1.4).

- A third feature of Bayesian Decision Theory is that as the number of observation sets X approaches ∞, the variance of the posterior ($p(\theta|D)$) decreases. Rapid convergence can actually be a problem when modeling learning in dynamic environments. This is because after only a few trials the posterior has converged so tightly, that current observations no longer affect the posterior. Therefore, early experience has a greater impact than current observations which is not optimal for adaptive systems. This problem can be ameliorated by "discounting" information after each time step. Essentially, the discounting can be thought-of as a form of "forgetting" between trials.

Although Bayesian methods are useful, one drawback of this approach is that any nonlinear transformation of the parameter space Θ will lead to a change in the den-

sity possibly causing the MAP solution to no longer be valid. A more practical concern is that Bayesian methods tend to be more difficult to evaluate than ML methods. Bayesian methods require the likelihood and prior distributions to be estimated, whereas ML approaches only require the likelihood be estimated. Many times estimating the prior is not an issue (e.g., large n case), but it can sometimes be problematic when a prior must be derived through a small n data set.

The general form and concepts in Bayesian approaches to SDT have been overviewed. Most of the existing work using SDT for perceptual investigation have adhered to a ML approach (e.g., [31, 32, 60, 66]). Therefore, it's important to understand the important features of ML decision theory. The next section will overview these features and contrast them with the Bayesian approach described in this section.

1.2.4 Maximum Likelihood Approaches

Maximum Likelihood (ML) approaches differ from Bayesian approaches in that they do not incorporate a prior when computing the risk. Rather, they treat the prior as a uniform distribution over θ, the possible states of nature. This approach is appropriate when there is no good reason to infer a prior (e.g., a novel task) or when the data is sufficient to overwhelm any prior. The ML approach attempts to assess the expected loss of each decision rule (d_b), if it were used repeatedly with varying X, for each θ. As a result, the *risk function* of any decision rule d_b is defined as:

$$\begin{aligned} R(\hat{\theta}, \theta) &= E_{ML}[L(a, \theta)] \\ &= \int_X L(a, \theta) p(x|\theta) dX \\ &= \int_{x_1} \int_{x_2}, \ldots, \int_{x_n} L(a, \theta) p(x|\theta) dx_1 dx_2 \ldots dx_n, \end{aligned} \tag{1.13}$$

where $a = \hat{\theta}$ is an estimate given by the decision function (d_b). Comparing the notation between Equation (1.5) and Equation (1.13), the differences between Bayes risk and ML risk become clear: Bayes risk incorporates *prior knowledge* and attempts find the action that minimizes the loss across parameters. Whereas, ML risk attempts to define a decision that minimizes loss across *different values of* X for *a fixed state of nature*, θ [11].

Using the ML approach, decisions d_{ML} are chosen that maximize the likelihood $p(X|\theta)$ of the data. If the data are i.i.d. then the result is:

$$\begin{aligned} d_{ML} &= \arg\max_a p(X|\theta) \\ &= \arg\max_a p(x_1, x_2, \ldots, x_n|\theta) \\ &= \arg\max_a \prod_{i=1}^{n} p(x_i|\theta), \end{aligned} \tag{1.14}$$

where $a = \hat{\theta}$. In practice, it is usually easier to minimize $-\log p(X|\theta)$ than to maximize $p(X|\theta)$. The minimization results in:

$$\begin{aligned} d^*_{ML} &= \arg\min_a -\log \prod_{i=1}^{n} p(x_i|\theta) \\ &= \arg\min_a \sum_{i=1}^{n} -\log p(x_i|\theta) \end{aligned} \tag{1.15}$$

For a Gaussian distribution, minimization leads to $\hat{\mu} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} x_i$ and $\hat{\sigma} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} (x_i - \hat{\mu})$. In the above classification example, a ML model would select the class that maximized the likelihood of the data. More specifically, it would state that we should choose c_1 if $p(X|c_1) \geq p(X|c_2)$, and we should choose c_2 otherwise. Stating this another way, it would mean that we should select c_2 if $x \leq \left(\frac{1}{2}\right)(\mu_1 + \mu_2)$ and select c_1 otherwise.

The ML approach suffers from the major limitation that it does not take into consideration any prior knowledge. Prior information is important for reducing the state space and in cases when the likelihood may have resulted from spurious data. Moreover, ML doesn't account for the fact that some types of estimation errors are worse than others [127]. For example, suppose you are being tested for lung cancer. It is common knowledge that people who smoke are more likely to develop lung cancer. Therefore, an optimal strategy for interpreting a positive test would be one that takes

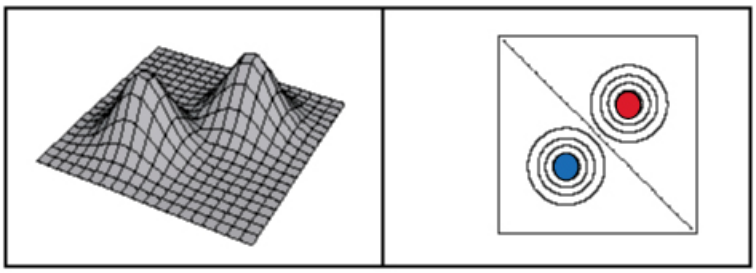

Figure 1.5: Figure adapted from [27]. Classification task when the priors are equal (ML approach). The red and blue posterior distributions resulted from uniform priors. The decision line running diagonally represents the optimal decision line to discriminate the two classes. Notice that the line runs through the middle of the two distributions.

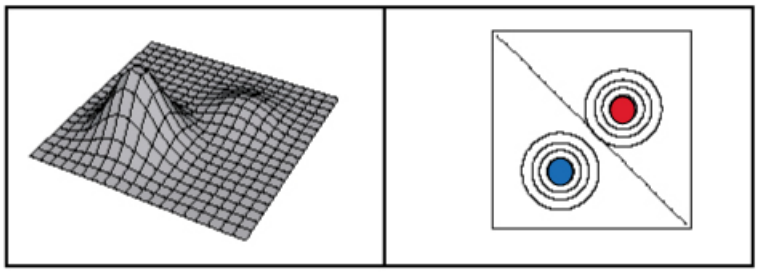

Figure 1.6: Figure adapted from [27]. Classification task when the priors were different (Bayesian approach). The red distribution resulted from a different prior (.2), than the blue distribution (.8). The decision line running diagonally represents the optimal decision line to discriminate the two classes. Notice that the line now runs closer to the red distribution - the one with the less probable prior.

into account the prior information when making a decision about what the test results mean. In the classification example, the change in decision criterion resulting from different approaches can be graphically seen in Figures 1.5 and 1.6. As a result of these fundamental limitations, the work in this paper will adopt a Bayesian approach to SDT.

This first section overviewed important aspects of SDT. Although these are abstract statistical concepts, there is evidence that our brain may be computing information in a manner consistent with these ideas. The next section will overview evidence that the human brain computes information similar to the theoretical ideas of SDT, such as expected loss/reward and uncertainty. The subsequent section will outline studies that suggest people are *behaving* according to their perceptual and motor uncertainties.

1.3 Statistical Decision Theory and the Brain

Although statistical models work well for modeling human performance, they do not provide insight into information processing unless they are based on mechanisms in the brain. The language of statistics formalizes how probabilities are derived and integrated. In order for the brain to compute probabilities, in a parametric sense, there must be some way to encode these distributions using neurons. Work in *population coding* has suggested that, not only does the brain compute probabilities, but it also computes things such as expected reward and uncertainty [33, 97]. Population coding has been used to describe sensorimotor transformations through basis functions [91], code image velocities [96, 100] and code for reach planning [36]. There is also evidence that population codes may be representing probability distributions [4, 93, 128, 129] and even doing Bayesian inference by coding the posterior distribution [93, 129, 35].

Moreover, population coding is a promising technique for investigating how the brain represents statistical concepts such as expected reward. Rather than merely being a theoretical abstraction, there is mounting evidence suggesting that the brain may be directly estimating expected reward (i.e., negative loss). The next subsection will overview work involving population coding in dopamine neurons that suggests these neurons are responsible for coding expected reward and uncertainty in appetitive behaviors.

1.3.1 Dopamine Activity for Appetitive Behaviors

Dopamine neurons in the ventral tegmental area (VTA) and substantia nigra have long been associated with rewarding behavior. For example, cocaine exerts its pleasurable sensations by prolonging the effects of dopamine on the neurons [123]. It has also been demonstrated that rats will suppress a lever to receive electrical stimulation to these areas. They will even choose to press the lever over other rewarding stimuli like sex or food [81]. Therefore, it seems that dopamine neurons are a good candidate for encoding reward in goal directed movements. The current section will overview work that supports this notion.

Work by [97] demonstrated that dopamine neurons code for errors in the predicted reward of appetitive behaviors. In this seminal piece of work, [97] showed that monkeys' dopamine neurons fired when a visual (or auditory) stimulus (US/CS) was paired with a reinforcing stimulus (e.g., piece of apple, or juice). The response in

dopamine neurons can be seen in Figure 1.7. Notice when the monkey is beginning training with the visual or auditory stimulus (US), the dopamine neurons fire at the time of the reward (Figure 1-6, top). The authors maintain that the activity results from a positive prediction (TD) error, which will be described below.

However, as the animal trains with the reinforcing pair the dopamine neurons begin to fire at the time of the visual or auditory stimulus (CS). The authors believe this to result from the CS perfectly predicting the reward, leading the two pieces of information to become redundant. This redundancy allows the dopamine neurons to only respond to the piece of information that occurs first (i.e., the CS).

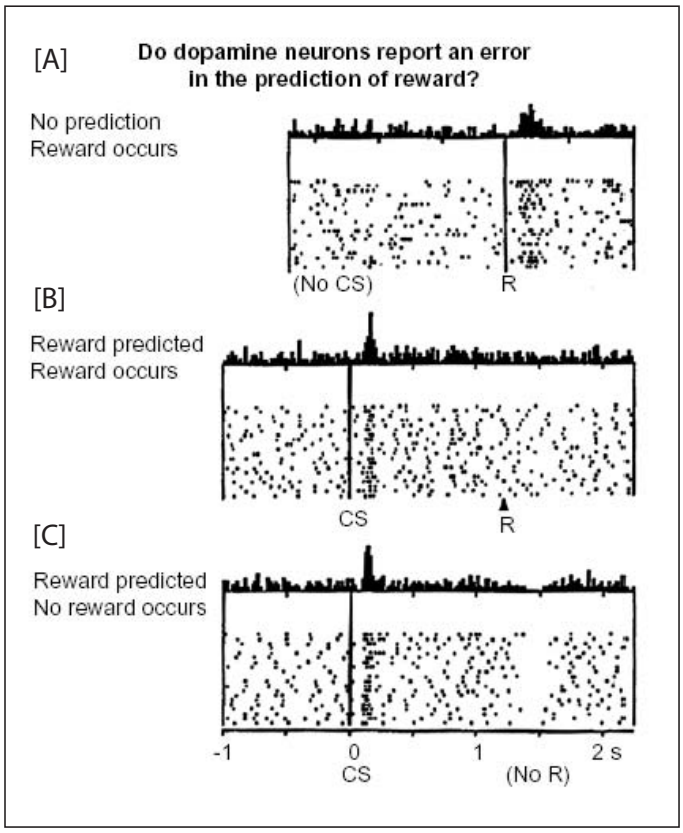

Figure 1.7: Figure adapted from [97]. Dopamine neuron response in monkeys for appetitive behaviors. **[A]** Before learning, the US the does not predict the reward which leads to a positive TD error. This error causes dopamine neurons to fire at the time of the stimulus. **[B]** After learning, the CS predicts the reward (no TD error), so the dopamine neurons fire at the time of the CS (see text). **[C]** CS is given but is not followed by a reward. This leads to a negative TD error and a sub-baseline firing rate in the dopamine neurons at the time of the reward (see text).

Finally, if the reward is removed from the CS it leads to a reduction in the firing of dopamine neurons at the time of the reward. Reduction in the firing of dopamine

neurons is now believed to result from antagonistic activity in the serotonin system [25].

These results are impressive not only for their empirical merit, but also becayse they can be predicted by theoretical ideas in Reinforcement Learning (RL). RL has been a powerful tool commonly used by the AI community for robotics applications and machine learning. More recent applications in computational neuroscience have been demonstrated in work such as [97]. They argue that the observed results in dopaminergic response can be fully predicted by the TD error - the difference between the predicted reward and the actual reward. In RL, the expected value for the task takes on the following form:

$$V_t = E[\gamma^0 r_t + \gamma^1 r_{t+1} + \ldots + \gamma^n r_{t+n}], \tag{1.16}$$

where r_t is the expected reward at time t and $E[\cdot]$ is the sum of future rewards until the end of the trial n. Notice that $0 \leq \gamma \leq 1$ is a discount factor that makes important rewards arrive sooner than less important rewards.

Relating these ideas to the earlier discussion on SDT, a relationship between the reward in Equation 1.16 and loss can be expressed as:

$$r_t^a = E_{\theta, x_t}[-L(a, \theta)] \tag{1.17}$$

You can see from this relationship that reward is simply the negative loss involved with the task. As a result, it is plausible that dopamine cells could encode both the expected reward and loss for the task, depending on the formalization.

The authors maintain that since the following holds:

$$V_t = E[r_t + \gamma r_{t+1}], \tag{1.18}$$

an error that is based on the information at successive time steps can be developed, that is called the *TD error*:

$$\delta_t = r_t + \gamma \hat{V}_{t+1} - \hat{V}_t \tag{1.19}$$

The authors suggest that dopamine neurons seem to encode information about the reward for appetitive behaviors through TD error. The key point is that if the brain is shown to compute TD error, then it is also likely to be able to compute risk.

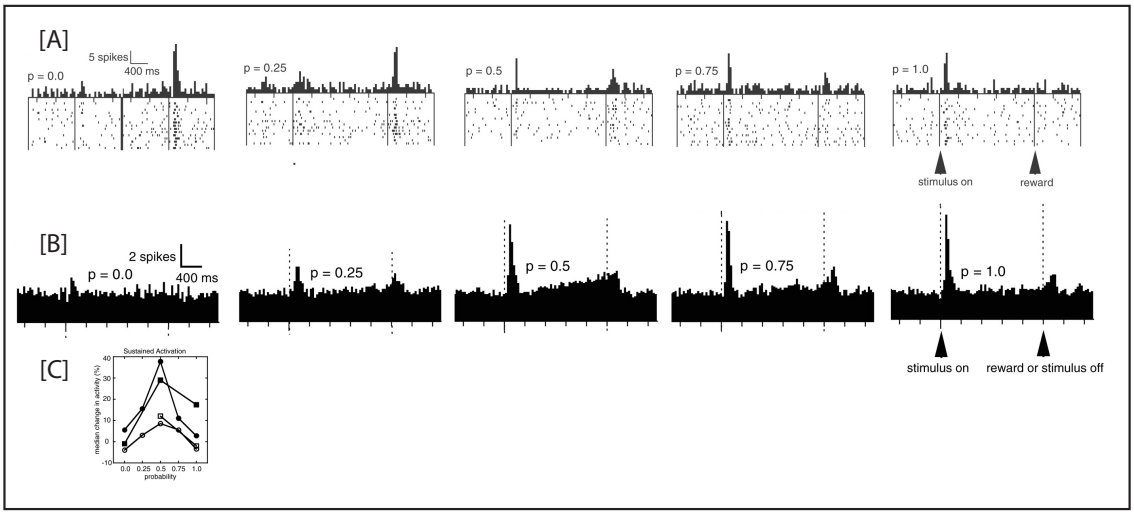

Figure 1.8: Figure adapted from [33]. Demonstrates the phasic activation of dopamine neurons (top) and the sustained activation of dopamine neurons in monkeys (middle, bottom) for appetitive behaviors. **[A]** Results demonstrate that dopamine neurons are encoding the expected reward in the task in addition to the probability of the reward (see text). **[B]** Sustained activation appears to encode the uncertainty in the task (see text). **[C]** Sustained activation as a function of the probability of the reward. Results imply that sustained activation encodes uncertainty as sustained activation is greatest when entropy is maximal ($p = .5$).

Another study by [33] demonstrated that dopamine neurons not only encode the expected reward for the task, but also the uncertainty involved with the task. To exemplify this idea they employed a similar paradigm as [97], except they added uncertainty to the reward structure (Figure 1.8). It can be seen from Figure 1.8A that the response of the dopamine neurons shifts from responding maximally at the time of the reward (p=0.0), to responding maximally at the time of the CS (p=1.0). For intermediate probabilities the dopamine neurons seemed to take distributed values across the two. Therefore, it appears that phasic activation of the dopamine neurons not only encode the expected reward but also the probability of the reward structure.

Another interesting finding was that the sustained activation (Figure 1.8B,C) encoded the uncertainty in the reward structure. Uncertainty can be measured nonparamet-

rically through entropy:

$$H(x) = -\sum_{x \in X} p(x) \log_2 p(x), \tag{1.20}$$

where $H(x)$ is the entropy associated with an event x. Notice that entropy is maximal when the $p(x)$ is uniform. Consequently, in this task entropy is greatest when $p(x) = .5$ because in their task reward was a binary outcome (i.e., either the monkey received a reward or they did not).

Figures 1.8B and 1.8C show that sustained activation is greatest when $p = .5$, which is when entropy is theoretically maximal. The above work implies that dopamine neurons encode uncertainty in the task through sustained activation.

This section overviewed how dopamine neurons encode both expected reward and uncertainty. If the brain encodes this information in appetitive responses, it's not unreasonable that humans would use this same information for different tasks. Now it will be demonstrated that humans *behave* as if they are taking into account uncertainty in psychophysical tasks.

1.4 Uncertainty Estimation in Human Behavior

Most work concerned with behavioral measures of uncertainty estimation have employed cue combination paradigms. Cue combination is a psychophysical task that requires subjects to make perceptual judgements or motor responses using the sources of information available (e.g., visual, haptic, auditory). The reliability of the information is often manipulated to assess how people integrate perceptual information in the presence of uncertainty. Cue combination has a long tradition in perceptual research and has been used to explore how people integrate visual cues (e.g.,[62, 63, 65]), visual and haptic cues (e.g.,[5, 31, 32]), visual and auditory cues (e.g.,[37, 10]) and even visual and proprioceptive cues (e.g.,[113, 114]). The next sections will review two cue combination studies demonstrating that people optimally integrate information both within a perceptual modality (i.e,. visual information) [63], and between perceptual modalities (i.e., visual and haptic information) [31].

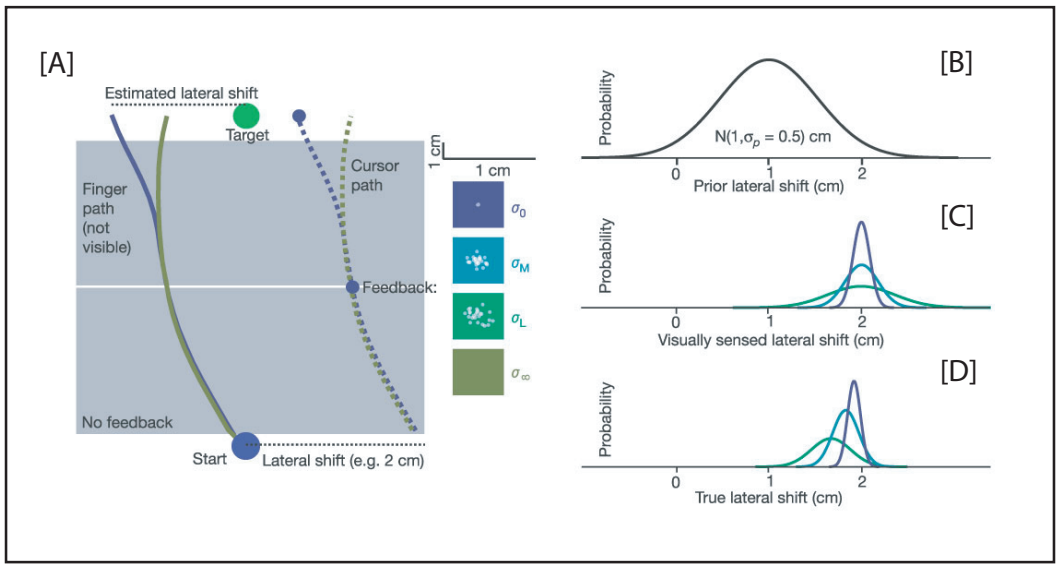

Figure 1.9: Figure adapted from [63]. Graphic demonstration of the task [A] and the Bayesian model [B-D]. **[A]** Subjects were required to make a pointing movement from the starting location (purple dot) to the target location (green dot). Once the movement was initiated, view of the cursor was occluded until half way into the reach and the cursor was displaced by a gaussian $\mathcal{N}(1, .5)$. This displacement Gaussian can be thought-of as the prior. Once view of the cursor returned, it did so with different levels of visual uncertainty: σ_0 represented veridical visual information; σ_M and σ_L represented medium and large "blurring" of the visual information, respectively; and σ_∞ indicated that visual information was not given. **[B]** Prior distribution on lateral shifts; **[C]** Distributions for the different types of visual feedback. **[D]** Expected shifts (as a function of visual uncertainty) if optimal Bayesian integration occurs.

1.4.1 Unimodal Cue Combination: Visual Uncertainty

Visual information is a dominant perceptual cue that specifies the properties and locations of objects in our environment. However, the quality of the visual information varies as a function of several factors. For example, peripheral viewing of an object linearly increases uncertainty about the object's location in space ([17, 18, 69, 120]). Therefore, changing the eye's position with respect to the target will change the observer's ability to estimate where the target is located in space. As a result, it is important that researchers understand how people use knowledge of their uncertainty when performing daily activities, such as reaching.

Research by [63] attempted to explore if people optimally estimate and act according to the visual uncertainty in a pointing task. Subjects were simply required to make a pointing movement from one location to the next (Figure 1.9). In this task, the subject's hand was never visible. Rather, they received feedback in the form of a cursor

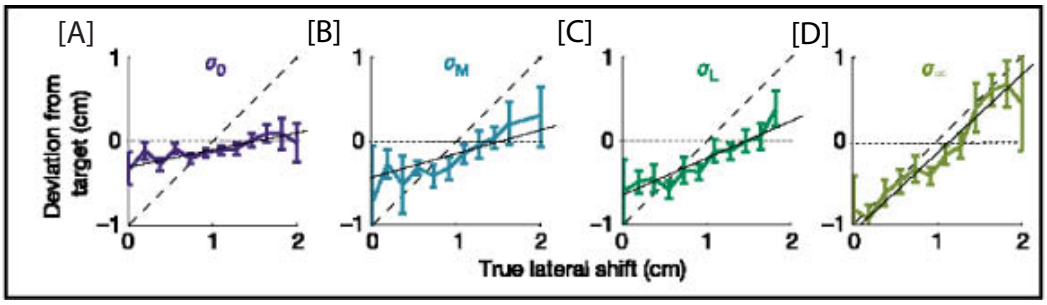

Figure 1.10: Figure adapted from [63]. Represents the deviation of the target (cm) as a function of true lateral shift, across visual information. Notice as the visual uncertainty increases (from **[A]** - **[D]**), that the deviation from the target shifts towards what the prior would predict (diagonal dotted line). However, when visual information is good **[A]**, then the deviation from the target is small, as Bayesian integration would predict.

that notified the subject as to the location of their finger. Once the subject began a reach, view of the cursor was occluded until half-way into the reach and the cursor was displaced by an amount randomly selected from a Gaussian ($\mathcal{N}(1, .5)$). Once the cursor returned, uncertainty in the visual feedback of the cursor was manipulated (See Figure 1.9). It was predicted when visual information is reliable, the subjects should use their visual information more heavily and shift their estimates toward the visual estimate (See Figure 1.10). If subjects are using Bayesian integration, and they receive unreliable visual information, they should not shift their estimate as much as when visual information is precise.

The results demonstrate (Figure 1.10) that subjects reached as if they were estimating the visual uncertainty in the task and optimally combining it with the target prior. These results are consistent with the Bayesian integration of perceptual information.

The current section outlined how people estimate visual uncertainty and integrate it in an optimal manner when making a reach. The next section will explore how both haptic (i.e., touch) information and visual information can be used in making psychophysical estimates.

1.4.2 Multimodal Cue Combination: Visual and Haptic Uncertainty

In addition to visual information, whenever a person is able to interact with an object through touch, they also receive haptic information that specifies the object's location and type. This brings-up the question of how people integrate these two forms of perceptual information when performing a task. Work done by [31] suggests

that the brain optimally integrates perceptual information from different modalities.

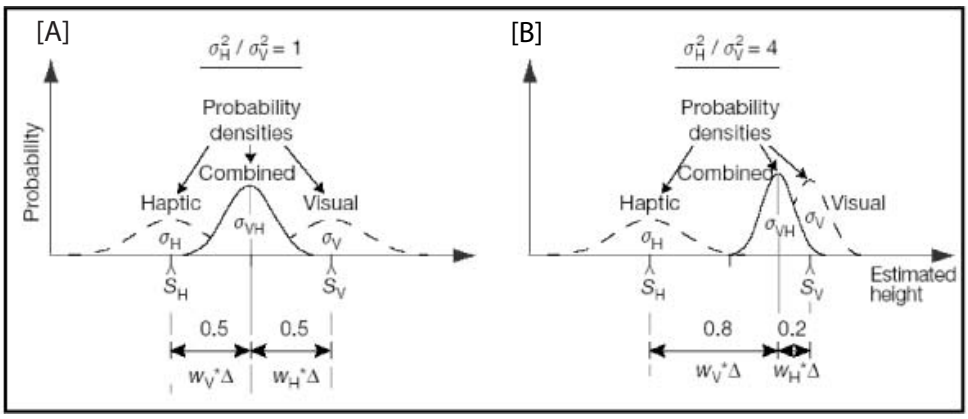

Figure 1.11: Figure adapted from [31]. ML integration of visual and haptic estimates. In this figure, the visual and haptic estimates of the block's height differs by Δ. Notice when the uncertainty in the visual and haptic estimates are equal [**A**], the combined density weighs each equally and takes a middling position. However, when the haptic estimate is less reliable than the visual information [**B**], the combined distribution gives greater weight to the visual estimate and is shifted toward the visual estimate.

ML integration in cue combination studies maintains that an estimate is weighted towards the cue with the most reliable likelihood (Figure 1.11). Therefore, if people are combing information optimally, their response should be biased towards the more reliable cue.

To test this hypothesis, subjects were required to make visual and haptic estimates about the height of a block. The estimates could be made using visual (stereo) information or through haptic information provided by "touching" the virtual block using a PHANToM arm (Figure 1.12). Head-fixed subjects viewed the stereogram through shutter glasses that allowed for binocular disparity. The surfaces of the stimuli were perpendicular to the line of sight. Haptic information was provided through interaction with the virtual object using a PHANToM arm that provided force feedback. Position of the subject's finger and thumb were given through two small 3D virtual markers.

In order to assess if ML combination was occurring, random displacements (errors) were introduced to the visual information but no manipulation to the haptic uncertainty was made. It was predicted that subjects should adjust their estimate of the object's height in a manner consistent with ML integration (Figure 1.11). More specifically, it was predicted that estimates would be weighted towards the haptic es-

timate when visual uncertainty was added, but weighted towards the visual estimate in the absence of additional visual uncertainty. The results of this study supported these predictions and it was concluded that the brain is optimally integrating visual and haptic information in a ML manner.

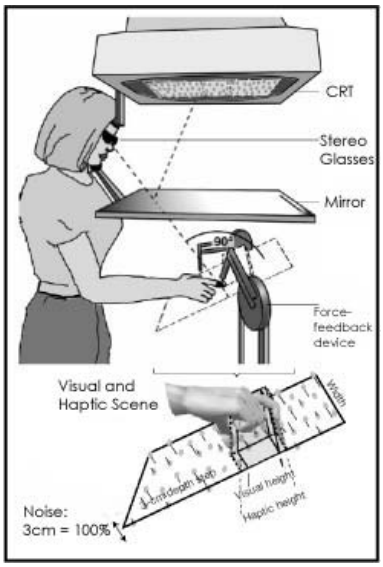

Figure 1.12: Figure adapted from [31]. Display of the apparatus used in the study. See text for details.

Although there's an obvious discrepancy between the results of [63] and [31] as to the form of integration that takes place in the brain (ML versus Bayesian). It could be that the only reason work similar to [31] found results consistent with a ML model is that the subjects were performing a novel task, therefore no prior information (in the Bayesian learning sense) could be incorporated. Although [63] used a novel task for the study, subjects were forced to "learn" the prior thereby allowing it to be integrated later. When people perform everyday actions such as reach and grasp, they can integrate their prior knowledge since they are well practiced at the task. Bayesian integration will be used throughout this paper as subjects in the experiments described later will be required to perform well-practiced reach and grasp movements.

The overall goal of this work is to provide evidence that people are grasping objects in a manner that minimizes the expected risk for the task. However, this necessitates that the brain is aware of: 1) the perceptual and motor uncertainty associated with the system; and 2) the loss function resulting from the task demands. Therefore, the paper is organized such that Chapters 2 & 3 will provide evidence that people

are adjusting for the uncertainty both within (Chapter 2) and across (Chapter 3) perceptual modalities when making a grasp, respectively. Then, Chapter 4 will detail and test a loss function for a grasping task that is based on the physics of object manipulation. Finally, Chapter 5 will offer suggestions for other perception-action cycles that could benefit by using SDT. Together, these efforts will provide evidence for the utility of using SDT to better understand natural movements.

Chapter 2

Effects of Visual Uncertainty in Grasping Movements

2.1 Introduction

Several studies in motor control have demonstrated that the sensorimotor system incorporates estimates of motor uncertainty to improve the success of point-to-point reaching movements. For example, previous research has shown that people use knowledge about their motor uncertainty to minimize end-point variance [46], avoid obstacles [92, 45], and maximize the expected payoff of a reach [110, 111]. Other studies suggest that the sensorimotor system also uses estimates of visual uncertainty and optimally combines them with motor uncertainty, both across reach trials [63, 19] and during feedback control of a reach [125, 94]. However, these studies do not address how visual uncertainty is used in grasping movements, where the goal of the task is to manipulate an object. The purpose of this chapter is to investigate the effects of visual uncertainty on grasping behavior.

The consequences of perceptual uncertainty for producing a successful grasp are complex and currently unknown. In contrast to point-to-point tasks where the objective of the movement can be expressed in terms of the final state of a point on the finger (e.g., Section 1.3.1), the goal of grasping movements is to generate and maintain appropriate forces on the object, subject to kinematic constraints imposed by hand and object geometry [23, 102]. While the goals of grasping are generally complex, sufficient conditions can be characterized for simple movements like grasping a vertically

oriented cylinder. In particular, maximum pinch force can be achieved by placing finger contact normals parallel and at the same height on the cylinder (thus producing maximal frictional contacts). Research has shown that minimum jerk paths to such a parallel arrangement requires the fingers to open wider before contact to achieve an approach that is parallel to the surface normals [102]. However, parallel approach and contact are clearly more difficult to obtain if the object's and hand's locations are not known precisely. Peripheral viewing of an object linearly increases uncertainty about the object's location in space [18, 69, 120]. In addition, peripheral viewing reduces our understanding of the hand's position relative to the target [94]. Because uncertainty increases the difficulty of successfully completing a reach and grasp, this chapter tests if people adjust for their visual uncertainty when making such movements.

It is expected that people will display more conservative grasping behavior with increased visual uncertainty. Since it is known that maximum grip aperture (MGA - maximum distance between the thumb and finger) increases without visual information [122], the current effort explores if subjects adjust their MGA according to the amount of visual uncertainty in the task due to peripheral viewing. The intuition is that if people are more uncertain about the object's location, they simply open their hand wider to assure their fingers don't collide with or miss the object (Figure 2.1).

More generally, it is anticipated that uncertainty should change finger trajectories. For visual sensing, there is usually more uncertainty in the depth-plane than in the picture-plane [113]. When performing precision grasps, the fingers typically contact an object at different depths leading to differential uncertainty at the contact location of each finger. Moreover, one of the fingers may contact the object at an occluded point where there is more visual uncertainty about surface location. For both these reasons, the effects of uncertainty may impact one of the fingers more than the other, and compensation for uncertainty may require independent adjustments between the fingers. To address these possibilities, principal changes in finger trajectories and coordination due to visual uncertainty will be analyzed.

Finally, increased uncertainty about object location may result in more variability in object-finger contact. To test this prediction, variability in finger positions near object contact will be measured as a function of visual uncertainty. Together, these results provide a more comprehensive description of the effects of visual uncertainty

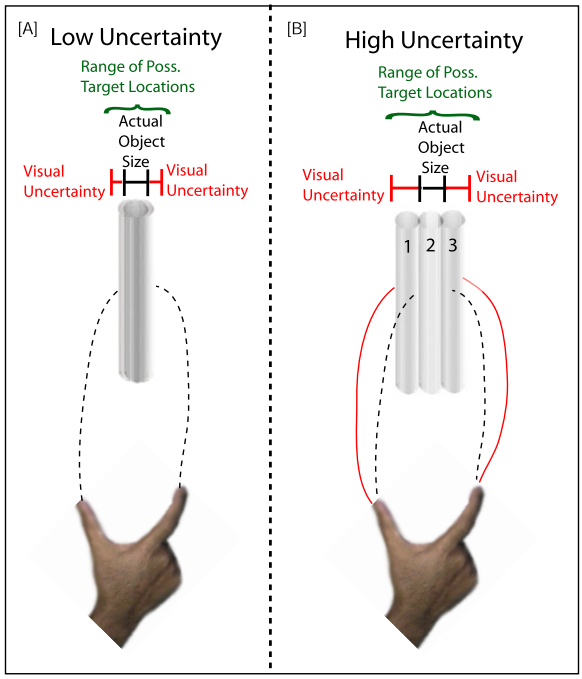

Figure 2.1: Figure demonstrates how MGA is predicted to vary with uncertainty. It is assumed that MGA is proportional to the target size plus the overall uncertainty. For simplicity, the diagram only depicts target location uncertainty. **[A]** In conditions when uncertainty is low there are fewer locations at which the target could be located. As a result, MGA can be kept smaller compared to when uncertainty is high. **[B]** In high uncertainty conditions, there are a number of possible target locations. Notice when target location uncertainty is high, a larger MGA should be employed to decrease the chance of colliding with (location 3) or missing (location 1) the target. See text for details.

on precision grasping behavior.

2.2 Materials and Methods

2.2.1 Subjects

Five right-handed subjects (4 males, 1 female) from the University of Minnesota participated in the study for monetary compensation. The subjects ranged from 19 - 23 years of age, and all had normal or corrected to normal vision.

2.2.2 Apparatus

Trajectory data were acquired by attaching three infrared emitting devices to the fingernails of both the forefinger and the thumb that were tracked via an Optotrak 3020 sampling at 100Hz (Figure 2.2C). Forward view distance from the fixation location

("E") was 62 cm. Fixation letters (A-I) corresponded to degrees away from the cylinder (80-0, respectively). Subjects began each reach from a starting location located 35 cm away from a 2.2 cm diameter cylinder (3D distance from the center of the starting box finger indentations to the center of the cylinder), 11.5 cm in length. The starting location was located 95 cm off the ground plane and approximately 10 cm below the cylinder plane. In terms of the apparatus, the x-dimension is the dimension that primarily captures the transport component and extends between the observer and the fixation marks. The y-dimension primarily corresponds to the grip-width and is in the depth plane on Figure 2.2C. Finally, the z-dimension is the dimension parallel to the cylinder axis and extends from the floor to the ceiling with respect to the set-up. To complete a trial, subjects had to lift the cylinder approximately 5 mm off the target resting cradle to trip a switch on the bottom of the target resting cradle. The cylinder's position was maintained using a 2 cm tall clear plastic tube that was just large enough to allow smooth cylinder movement.

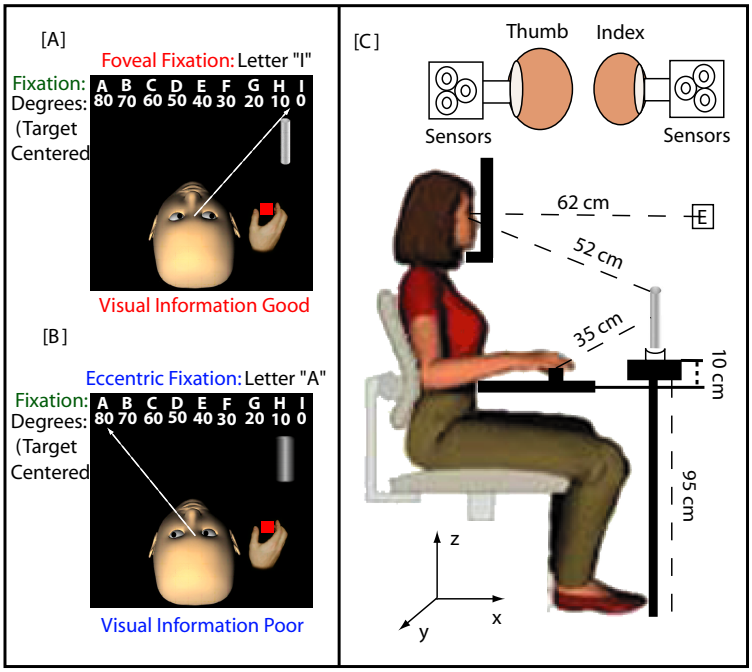

Figure 2.2: Experimental set-up used in psychophysical reaching task. See Methods Section for details. Subjects were allowed to view both their hand and the target throughout the duration of the reach. Visual information was manipulated by requiring subjects to fixate different points while making their grasp. This changed the visual uncertainty without affecting the kinematic demands of the task. When subjects fixated the target ([**A**]), the visual information was more reliable than when they fixated an eccentric point ([**B**]). [**C**] Side view of the reach apparatus displaying the key set-up parameters.

2.2.3 Procedure

Head-fixed subjects made repeated reaches to the same spatially fixed target located 40 degrees to the right of forward view (in head centered coordinates, 0 degrees in cylinder coordinates - see Figure 2.2). The cylinder was held fixed to assure that any changes in reaching behavior resulted from changes in viewing conditions, and the location was selected such that there was the maximal range of eye positions away from the target (0-80 degrees) to allow for the greatest range of visual uncertainty. Subjects were allowed to view the target and their hand throughout the reach. Varying the subjects' eye position without changing the target location allows for the manipulation of visual uncertainty without changing the kinematic demands of the task.

At the beginning of each trial, the fixation point was announced (e.g., "A"). After subjects fixated, they were allowed to make their reach to the target. Once subjects moved either finger more than 2mm from the starting location, the trial was initiated. The starting location was specified by indentations in the starting box that were approximately 1cm in diameter and located 0.75cm below the top of the box. Subjects could choose comfortable finger positions within the indentations at the beginning of each trial. Subjects had 1200 msec to successfully lift the cylinder. The trial ended when the subject lifted the cylinder high enough to trip a lever located on the bottom of the cradle. Once subjects successfully lifted and replaced the target, each finger had to be returned within 0.66 mm of its starting position before the next fixation point was announced. If subjects took too long, or didn't lift the cylinder correctly, an error message was played and the trial was repeated. Fixation points were randomly assigned throughout the trial and each fixation was announced seven times per block. Subjects ran in six blocks per day over three days, for a total of 1134 (7 repetitions x 9 Fixations x 6 Blocks x 3 Days) trials.

2.2.4 Analysis

Cubic interpolating splines were used to fill-in missing data segments that resulted from occlusion, if less than $40msec$. If the trajectory had total occlusion time $>$ $40msec$, it was discarded from the data-set. Only about five percent of the trajectories were discarded from the data due to occlusions that tended to occur at the beginning (i.e., launch off starting block) and return path of the reach. Moreover, one block of

data was excluded due to a computer error. For the remaining 89 blocks, trajectories were sorted by eccentricity and averaged across trials, blocks, and subjects. Simple averaging was justified because the starting location, timing, and end-point location were nearly constant, and provided a good measure of the change in trajectories induced by visual uncertainty. However, to assure that differences in trajectories were due to the visual uncertainty in the task, the mean trajectory for each block was subtracted off the individual trajectories (for the same block) to eliminate any systematic biases that were not due to visual information (e.g. slight variations in starting location between blocks). To achieve the plot in Figure 2.3A, trajectories for each eccentricity were averaged across blocks and subjects, with the global mean (across subjects, blocks and eccentricities) added back.

2.2.4.1 Visual Uncertainty Predictions

A simple model for the effect of uncertainty on MGA was developed by assuming that reach trajectories are based on the best (current) estimates of the object's location and diameter, and that grasps attempt to avoid object-finger collision. Avoiding object-finger collision during transport requires grip aperture to exceed the both the size of the object and the range of possible locations of the object. Previous results show that MGA is linear in target size with a slope around .8, and offsets that vary over several centimeters between experiments (for review see [102]). If it is assumed the effect of target location uncertainty (σ_y) on grip aperture is to increase the effective size of the target, then MGA should be linear in σ_y. Thus, for a fixed target diameter (D), MGA is expected to be linear in target location uncertainty:

$$MGA = \alpha\sigma_y + C \tag{2.1}$$

where α is the slope of the linear relationship and C is a constant that captures the effects of target diameter (D). Values of α represent different collision avoidance criteria in terms of standard deviations. A linear relationship between σ_y and MGA is tested.

Location uncertainty in the task (σ_y) is manipulated by peripheral viewing of the target and hand (Figure 2.2). The azimuthal angle between the target's retinal pro-

jection and the eye's central axis provide an imperfect signal v for target location in eye-centered coordinates. The error in visual sensing is modeled as zero-mean, with a signal dependent variance that models the effects of decreased spatial resolution on visual sensing of peripherally viewed targets. Results from two-point discrimination studies [17, 120, 69] demonstrate that visual uncertainty linearly increases with the eccentricity of the visual information (in mm):

$$\sigma_y = \sqrt{(.15 * v)^2} \tag{2.2}$$

This value was derived from the threshold $(v/30)^\circ$ reported in [69], and converted to mm. Note that similar models for visual uncertainty are used in both [84, 94]. Therefore, if people are estimating their visual uncertainty, it is expected that changes in their MGA should be proportional to changes in σ_y.

2.2.4.2 Maximum Grip Aperture

To assess if people are incorporating knowledge of their visual uncertainty into their grasping movements, the distance between people's thumb and forefinger (i.e., grip aperture) was measured throughout the reach. Maximum grip aperture (MGA) is the maximum distance between the thumb and forefinger during a reach, and it typically occurs 75-80% of the distance to the target object [101]. Its importance is that it serves as a measure of target uncertainty [122] and scales linearly with actual object size (e.g., [88, 87]). Grip apertures and maximum grip apertures were averaged across subjects and blocks to produce Figure 2.4. Change in MGA is defined to be the mean MGA for each eccentricity subtracting off the global mean MGA (across all eccentricities), for that block. Change in MGA was used to adjust for differential absolute levels of MGA across blocks (e.g., fatigue) and days (e.g., sensor placement). This helped assure that any changes in MGA were due to visual information and also allowed us to directly compare and average across subjects. If people are estimating their visual uncertainty and using it during a grasp, MGA should increase with increases in visual uncertainty (Figure 1).

2.2.4.3 PCA

To reveal the principal changes in reaching behavior due to visual uncertainty, principal components analysis (PCA) was performed on finger and thumb approach trajectories. The goal this analysis is to quantitatively describe global trajectory changes due to uncertainty. PCA views sampled trajectories as vectors in a high dimensional space, and finds directions in this space that account for the most scatter. To view trajectories as vectors, the trajectories were resampled every 10 msec between 250-500 msec, which results in x, y, and z measurements on 31 time points. Stacking the time samples from x ,y and z produces a single 93-dimensional vector. For each uncertainty condition i, wthe mean trajectory vector $\vec{m}_i$ was computed separately for the finger and thumb. Principal directions for the mean trajectory vectors across all uncertainty conditions were sought. Performing PCA on this set of vectors finds directions that maximize between-condition variation, or equivalently, directions along which mean trajectories differ the most with respect to uncertainty. The top principal component returned by the analysis vector $\vec{v}$ can be used to approximate the trajectory for each uncertainty condition $\vec{m}_i$:

$$\vec{m}_i \approx a_i\vec{v} + \vec{m}$$

where $\vec{m}$ is the global mean trajectory, and a_i is a coefficient that produces the best approximation. Stated in words, scaling along the principal direction $\vec{v}$ can be used to approximately morph the mean trajectory from one uncertainty condition to another. Thus, PCA can be used to determine how well the effects of visual uncertainty on approach trajectories can be captured by a scaling along common axes. It also provides a principled way to determine how many axes (dimensions) are required to represent differences between trajectories.

PCA was computed on the mean trajectory data for each block (89 total). The first PC accounted for 78% of the finger variance, and 50% of the thumb variance, while both the first and second PC accounted for 94% of the overall variance of the finger and 90% of the thumb variance. This procedure was validated using cross-validation with random subsets of the data. The training matrix used to generate PCA components was compared against the remaining ten percent of the trajectory data to assess quality of reconstruction, and to provide resampling-based estimates of the coefficients. Only the first principal component for the finger and thumb were used to produce the plots for Figure 2.7A,B. Coefficients represent projections of the

excluded data onto the average PC, sorted and averaged over eccentricity.

2.2.4.4 Displacement Vectors

The differences between trajectories in the 80 and 0 degree uncertainty conditions were also computed. These differences can be visualized as a set of *displacement vectors* that demonstrate the positional changes in trajectories that result from visual uncertainty at each time point (Figure 2.7C). Displacement vectors are displayed to provide an intuitive comparison of the mean and PCA reconstructed trajectories. In addition, displacement vectors allow coordination between finger and thumb to be visualized.

In Figure 2.7C displacement vectors are represented by arrows. The direction of the arrow represents the direction along which trajectories changed with increased uncertainty, whereas the length of the arrow reflects the magnitude of the change along the axis. Displacement vectors were computed by subtracting the foveal fixation trajectory (i.e., 0 degrees) from the most eccentric fixation (i.e., 80 degrees), for every 10 msec between 200-650 msec. This range and sampling were chosen to cover the most informative section of the reach.

2.2.4.5 Position Variability Near Contact

Another measure of interest is the variability in position near contact (final portion of movement time (85% - 92% of reach)), across levels of uncertainty. It is plausible that increased sensory uncertainty may lead to increased variability in the end-point location. Therefore, it's expected that increased visual uncertainty will lead to increased variability of the hand near contact. The range between 50 to 100 msec before the estimated "lift period" was chosen to be contact time. The lift period was defined by the "flat-spot" in the x-velocity (region of zero velocity), accompanied by a positive z-velocity. The intuition is that the hand is no longer moving forward (x direction) and also moving in the direction of the required lift direction (z direction). Therefore, contact must precede this lift phase by some brief period of time. The period preceding lift is attractive because almost all systematic differences in mean trajectories across eccentricity have disappeared by this time, which means that differences in variance in this time range reflect a less accurate approach to a common goal state. Mean change in contact variance is defined to be the global average con-

tact variance (across all eccentricities) subtracted from the average contact location for each eccentricity. Change in contact variance was used to adjust for any sources of finger variability across blocks, days, and subjects that were not due to changes in visual information. Contact variance was averaged across fingers, blocks and subjects to produce Figure 2.8A. However, dimensions were not averaged across in order to assess if visual uncertainty differentially impacted contact location variance across dimensions. Linear fits to the values were accomplished using a robust fitting method called Least Trimmed Squares (e.g., [47]), as implemented in the robust statistics software package LIBRA [117]. Least trimmed procedures were used in cases when the inclusion of a point in the fit would lead to an inaccurate description of the data (e.g., cases when a point has high leverage (e.g., Figure 2.8 (x-variance, 80° point))).

2.3 Results

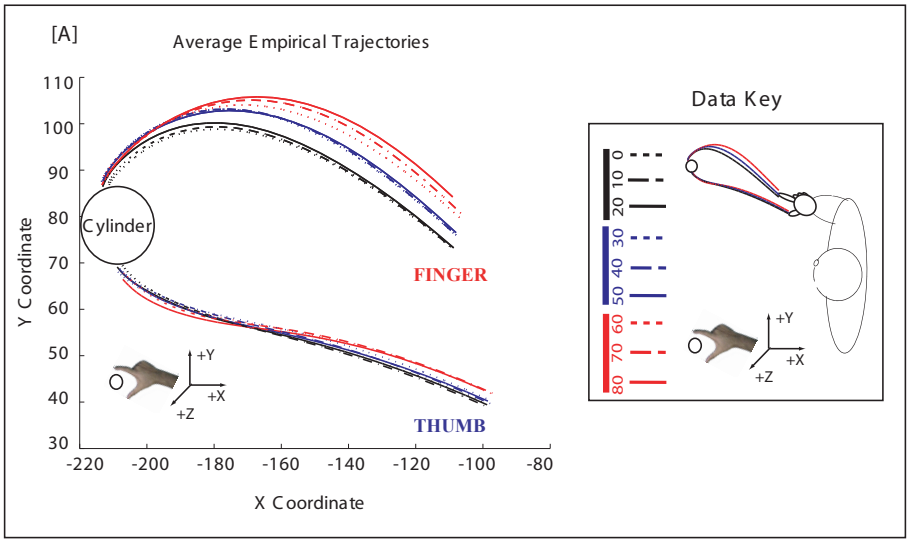

Figure 2.3: Trajectory data across different levels of visual uncertainty. Data are color coded into the viewing eccentricities closest to the target (Black Lines), those mid-distance from the target (Blue Lines) and those furthest from the target (Red Lines). Each eccentricity within a particular category is coded by a unique line type (See Data Key). **[A]** Mean empirical trajectories for the finger (top traces) and thumb (bottom traces) obtained from experiment. It is clear that the finger trajectories became more "hooked" as the visual uncertainty increased, whereas the thumb trajectories varied less across eccentricity.

The results demonstrate that the finger trajectories people took to the object varied as a function of visual uncertainty (Figure 2.3A). It was observed that the approach trajectory of the finger scaled with visual uncertainty, whereas the thumb approach trajectory scaled less across viewing eccentricity. Therefore, visual uncertainty seems

to predominantly impact the approach trajectory taken by the finger. Several more detailed aspects of this variation were examined, including changes in MGA, grip-width over time, the velocity of the fingers during transport, and the timing of the fingers.

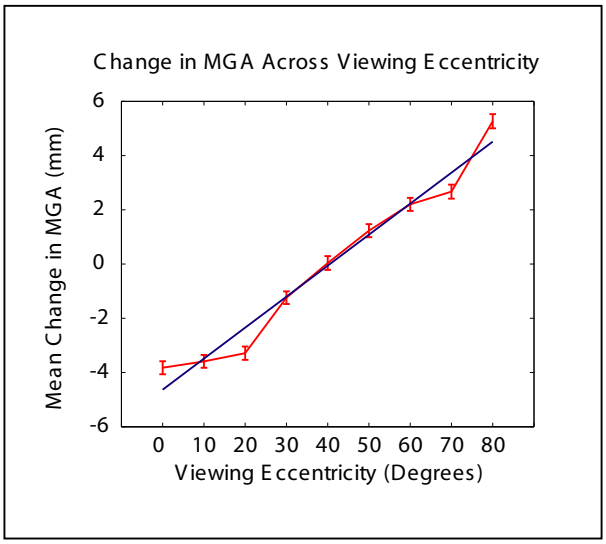

Figure 2.4: Diagram depicting maximum grip aperture from the experiment. [**A**] Average change in MGA across viewing eccentricity. Red line depicts the mean change in MGA, while the error bars represent ±1 SEM. There was a significant linear increase (blue line) in MGA as visual uncertainty increased (See Results Section). This result is also consistent with the model for MGA presented in the paper.

It is clear from the trajectory data that the fingers are changing their spread (i.e., grip aperture) across levels of uncertainty. As predicted by the visual uncertainty model for MGA, the average change in MGA increased linearly with viewing eccentricity (ecc) (Figure 2.4). Least Squares Regression ($MGA = .11(ecc) - 4.63, r^2 = .97$) on the means proved this trend to be significant, $F(1,7) = 207.71, p < .001$. Therefore, MGA appears to scale with the level of visual uncertainty in this reaching task.

In addition to MGA, this project investigated how subjects adjusted their grip aperture over time. Figure 2.5A shows the average grip aperture trends for the experiment. Grip aperture tended to scale with levels of visual uncertainty. However, grip aperture profiles appeared to occur over the same time-course across eccentricity (Figure 2.5A), suggesting that visual uncertainty impacts the magnitude of grip apertures, but not the time at which MGA occurs. In fact, the average time at which MGA occurred ($M = 36.45, SD = .29$) was almost identical across all of the reaches (Figure 2.5B). Least Squares fits to the mean data did not show a significant lin-

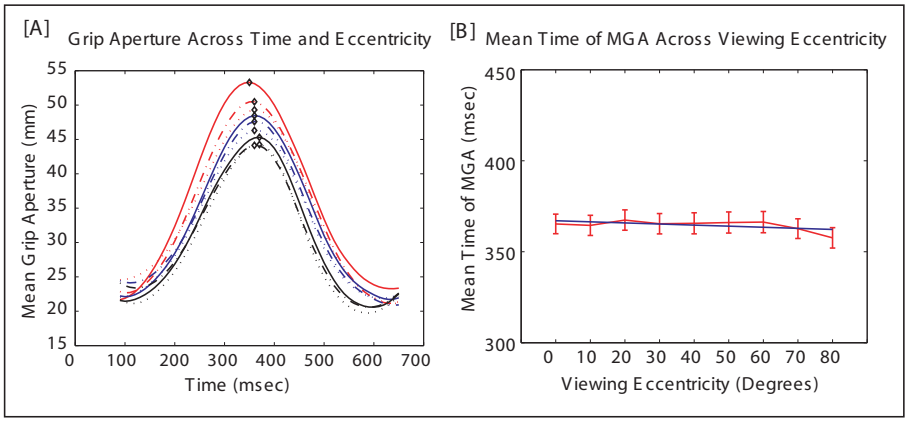

Figure 2.5: Diagram depicting grip aperture results from the experiment. See data key in Figure 2.3 for an explanation of the color codes. **[A]** Average grip aperture over time for each viewing eccentricity. Visual uncertainty appears to scale the amplitude of the grip aperture profile with high uncertainty conditions (Red Lines) having a larger amplitude than medium uncertainty conditions (Blue Lines), which have a larger amplitude than low uncertainty conditions (Black Lines). Black Diamonds show the time at which maximum grip aperture occurred. **[B]** Average times at which MGA occurred across viewing eccentricity. Red line depicts the mean, error bars represent ±1 SEM. As suggested from the grip aperture data ([**A**]), the time at which MGA occurs does not change as a function of visual uncertainty. This is verified by the blue linear regression line (See Results for Details) which is flat across eccentricity.

ear trend of MGA times across eccentricity ($T_{mga} = -.006(ecc) + 36.71, r^2 = .35$), $F(1, 7) = 3.77, p > .05$.

The fact that MGA varied with visual uncertainty, and that grip apertures tended occur over the same general time-course, suggests that there could be differences in the velocities the fingers took to the object. Previous research has also demonstrated changes in transport velocity with increased uncertainty (e.g., [101, 20, 12]). Therefore, it was investigated if velocity systematically changed with visual uncertainty for this grasping task. The velocity profiles for the transport component of the trajectories (X-velocity) show similar systematic changes to MGA across the levels of visual uncertainty (Figure 2.6A). However, it appears that there is only a small change in velocity that occurs around the time of MGA (350 msec) where high levels of uncertainty lead to slightly lower peak velocities (Figure 2.6B). Therefore, in this experiment, it appears that visual uncertainty has a small impact on the transport velocity people take to the object.

To quantitatively investigate how visual uncertainty influences approach trajectories, PCA analysis was employed. Figure 2.7A shows the PCA reconstructed trajectories

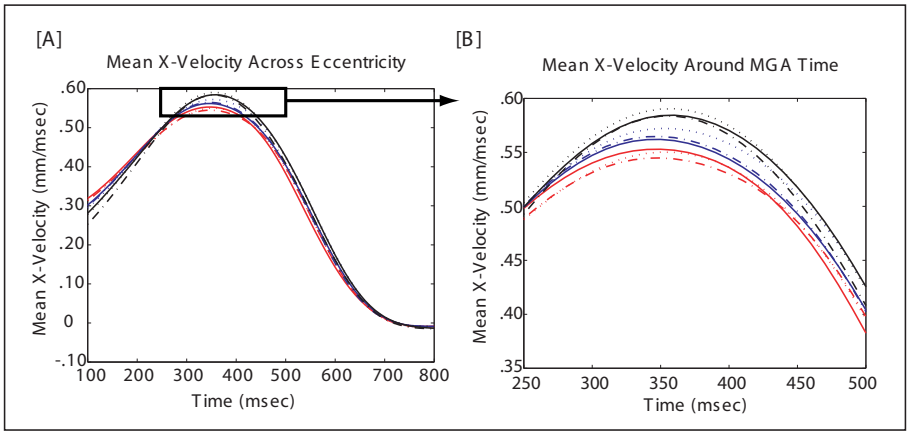

Figure 2.6: Diagram depicting velocity results from the experiment. See data key in Figure 2.3 for an explanation of the color codes. [**A**] Average finger velocities in the x-dimension (i.e., transport dimension) across viewing eccentricity. It is apparent that the velocity profiles are similar across viewing eccentricity. However, around the time of MGA (350 msec), there are small differences. [**B**] Shows the velocity profiles of the finger along the x-dimension around the time of MGA. There is an ordering of velocities at this time where high visual uncertainty conditions (Red Lines) have a lower velocity than medium uncertainty conditions (Blue Lines), which have a lower velocity than conditions with low visual uncertainty (Black Lines).

using only the first principal component. The PCA reconstructed trajectories are very similar to the empirical trajectories (Figure 2.3A), suggesting that increased visual uncertainty results in the fingers scaling along a principal axis of change.

To illustrate the magnitude of scaling along the primary axis (i.e., the first principal component), the coefficient values across eccentricity were plotted. Figure 2.7B shows the coefficient values for the first principal component of the finger and thumb. Least Squares fits to the mean coefficients results in a significant linear trend for the finger ($C_f = .55(ecc) - 21.99, F(1,7) = 203.64, p < .001$) and thumb ($C_t = .30(ecc) - 12.11, F(1,7) = 28.52, p < .01$). It is important to note the linearity of the PCA coefficients. This implies that the observed changes in MGA across levels of visual uncertainty (Figure 2.4) are the result of a scaling both the finger and thumb along a primary axis of change.

The pattern of displacement vectors in Figure 2.7C graphically illustrate changes in finger coordination due to visual uncertainty. Uncertainty has an asymmetric impact on both the path and relative timing of index finger and thumb trajectories. To understand the relative timing effects, note that the time intervals between vectors were constant. Thus, horizontal shifts between vectors for the index finger and thumb

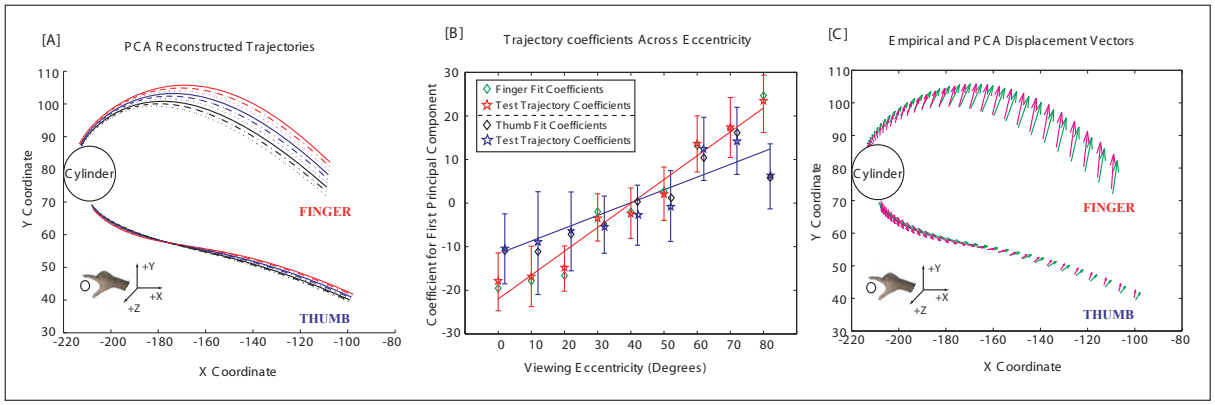

Figure 2.7: PCA reconstructed trajectory data across different levels of visual uncertainty. See data key in Figure 2.3 for an explanation of the color codes. **[A]** Principal Components Analysis (PCA) reconstructed trajectories using the first principal component (PC). There is a high degree of similarity between the empirical (See Figure 2.3) and PCA reconstructed trajectories, using only the first component. **[B]** First PC coefficients used to reconstruct the trajectories in [A]. Error bars represent ± 1 bootstrapped SEM. Least Squares linear fits to the finger (red line) and thumb (blue line) coefficients are also provided (See Results for Details). Larger coefficient values correspond to greater trajectory scaling along the primary dimension. The steeper slope for the finger coefficients show that the finger trajectories scaled more than the thumb trajectories as the amount of visual uncertainty increased. We used a cross-validation procedure to verify the top PCA component had predictive value. PCA was performed on random subsets in the data (diamonds), and the excluded data (stars) were projected onto the top component. Projection coefficients for both the included and excluded data had similar values, demonstrating that the analysis has good predictive value. **[C]** Displacement vectors (See Methods Section) for both the empirical (magenta) and PCA reconstructed (green) trajectories. The magnitude of the vectors for the empirical and PCA trajectories are similar, whereas the direction of the vectors are slightly rotated. Therefore, it appears that the first PC appropriately adjusts for the magnitude of change, whereas higher order PCs control the direction of change.

at matching times can be interpreted as one finger moving more rapidly to the target, and which finger first contacted the object. From the shifts between displacement vectors in Figure 2.7C, it can be seen that the finger preceded the thumb throughout the reach, and that the finger contacted the object before the thumb. In terms of trajectory paths, the index finger changes are much larger than the thumb in the y-dimension, accounting for the majority of the MGA changes (Figure 2.4). The effects of uncertainty on the thumb trajectories are more subtle, involving an early increase in the y-dimension towards the finger, followed by a change in the transport direction that reflects a reduction in thumb velocity relative to the finger. It is proposed that these changes may result from a more conservative approach plan when there is increased visual uncertainty. Specifically, the changes can be interpreted as a strategy that tries to maintain a parallel finger-thumb relationship during the period when both fingers contact the object and grip forces would be first applied.

Displacement vectors also allow the PCA reconstructed trajectories to be compared with the empirical trajectories. If the empirical and PCA reconstructed trajectories have displacement vectors of similar length and orientation, it suggests that people are adjusting their finger approach trajectory for visual uncertainty along a primary axis of change. As stated in the Analysis Section, the direction of the arrow represents the direction along which trajectories change with increased uncertainty, whereas the length of the arrow reflects the magnitude of the change along the axis. It is apparent that the magnitude of the arrows is similar for the empirical and PCA reconstructed trajectories. Whereas, the orientation of the arrow is slightly different between the empirical and PCA trajectories, suggesting that the first PC is sufficient for describing the magnitude of changes in trajectories due to uncertainty, but does not capture more subtle differences.

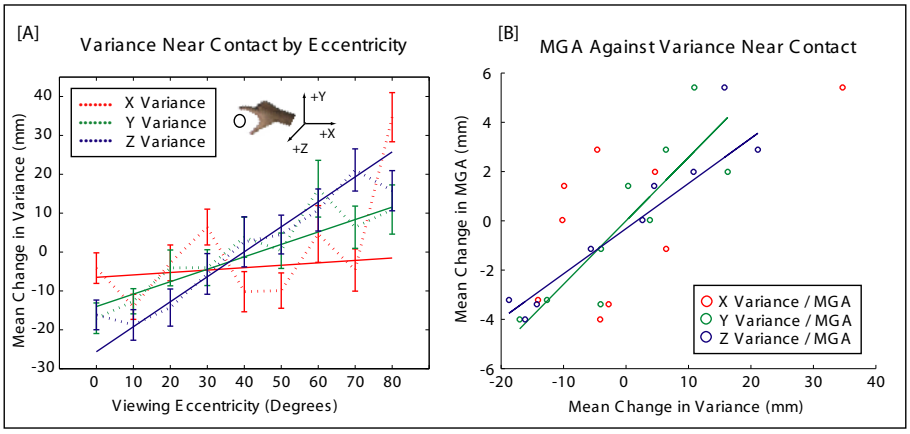

Figure 2.8: Contact variance as a function of visual uncertainty. [**A**] Dashed lines represent the mean change in contact variance across viewing eccentricity, and error bars represent ±1 SEM. Solid lines represent the Least Squares fit to the data. A linear increase in contact variance across viewing eccentricity was observed for the Y- (green line) and Z-dimensions (blue line), but not for the X-dimension (red line; See Results Section). [**B**] Scatter plot of MGA against contact variance with Least Squares regression lines shown for only those correlations that are significant. There were significant correlations between the contact variance in the Y- (green dot) and Z-dimensions (blue dot) and MGA, but not between contact variance along the X-dimension (red dot) and MGA (See Results Section).

It is reasonable to assume when there is more uncertainty about an object's location, there should be more variability in the object-finger contact locations. To test this idea, it was explored how the variability of finger positions near object-finger contact changed with uncertainty. As expected, increases in uncertainty lead to increased variance near the time of contact (Figure 2.8A). Least Trimmed Squares fits to the data displayed no linear increase in variance along the x-dimension ($V_x = .06(ecc) -$

$6.51, r^2 = .05$), but strong linear increases in the y- ($V_y = .31(ecc) - 14.02, r^2 = .91$) and z-dimensions ($V_z = .64(ecc) - 25.64, r^2 = .98$). Please note that Least Trimmed Squares regression is a robust fitting tool that does not allow outliers (e.g., x-dimension, 80°) to influence the fit (See Analysis Section). From this analysis, it is evident that increased visual uncertainty leads to increased contact location variance for dimensions in the picture plane, but not in the depth plane.

Finally, to verify that MGA and variance near contact are both encoding visual uncertainty, the relationship between the two measures was explored (Figure 2.8B). Pearson product moment correlation showed a significant linear relationship between y-variance and MGA ($r_{mga,y} = .86, p < .01$), and z-variance and MGA ($r_{mga,z} = .95, p < .01$), but no significant relationship between x-variance and MGA ($r_{mga,x} = .62, p > .05$). From this result, it appears that changes in MGA predict changes in contact location variance for dimensions in the picture plane, but not in depth.

2.4 Discussion

In accordance with the predictions, it was observed that people change their grasping behavior as a function of the visual uncertainty in the task. More specifically, it was demonstrated that approach trajectories of the finger and thumb scaled as the amount of visual uncertainty increased (Figure 2.3A). It was shown that most trajectory adjustment for visual uncertainty occurred in the index finger, while the thumb showed much smaller changes. One simple explanation for this result is the fact that the index finger is relatively easier to move than the thumb. It is typically longer, lighter, and has more degrees of freedom than the thumb and, as a result, may be easier to adjust along the grip-width (i.e., y-) dimension. Another possibility is that participant's index fingers typically moved to contact an area on the object that was occluded from view, while the thumb was potentially visible throughout the reach. Thus, there may be more uncertainty about the finger's contact location than the thumb's, causing more compensation in the finger's trajectory. Further research would be required to distinguish between these and other possibilities.

More detailed analysis of the trajectory data showed that the PCA reconstructed trajectories using only the first principal component (Figure 2.7A) were very similar to the empirical trajectories, suggesting that the effect of visual uncertainty on trajectories is predominantly a scaling along a principal axis. In fact, the coefficients

(Figure 2.7B) show that this scaling is linear and that the first principal component can accurately account for the magnitude of this effect (Figure 2.7C). This suggests a control strategy that involves using the same force signals to drive the reaches, but adjusting the overall amount of force to compensate for uncertainty.

It was also demonstrated that MGA increased linearly with viewing eccentricity (Figure 2.4). This result is consistent with how the reliability of visual information changes with eccentricity [17, 18, 69, 120]. This also agrees with previous research that has reported increased MGA under conditions of no visual information [122] and when subjects are viewing an object in their visual periphery [101]. However, [101] fit subjects with a contact lens that occluded foveal vision and allowed subjects to freely move their eye when making a reach. They reported that subjects typically viewed the object between 25-30 degrees, but did not control subjects' eye position. Therefore, eye position (i.e., the amount of visual uncertainty) was not systematically varied in their study making it difficult to assess the impact of visual uncertainty on reaching behavior. By systematically varying visual uncertainty, this project was able to demonstrate the functional (linear) effect of visual uncertainty on MGA. In addition, these empirical results were predicted by providing a simple model for MGA. The model assumes that MGA scales with object diameter and the overall uncertainty in the task. In this experiment, the object's diameter was held constant and visual uncertainty was changed through peripheral viewing. Therefore, the observed changes in MGA corresponded to the changes in visual uncertainty, as predicted by the model.

In addition to changes in MGA, it was demonstrated that grip aperture profiles scaled with increased levels of visual uncertainty (Figure 2.5A), and the time of MGA was relatively consistent across levels of visual uncertainty (Figure 2.5B). This fact is somewhat surprising as previous research by [101] reported that MGA occurred sooner when people were forced to use their peripheral vision to make a reach. The discrepancy in the times at which MGA occurred could be due to the fact that subjects in this study were required to complete their reach in less than 1200 msec, whereas [101] allowed subjects to choose their own reach time. This fact may have allowed subjects in their [101] study more opportunity to adjust their grip aperture over the course of a reach, leading to the observed timing differences. Another possibility is that the differences between the studies may result from the properties of the object. In the current study, the object was resting in a "cradle" allowing subjects to collide with the cylinder with little adverse consequence. Conversely, [101] placed the dowel

on the workspace without a cradle, increasing the cost of colliding with the object. Therefore, it could be that subjects in their study altered their grasping behavior to account for the high cost of colliding with the object.

Similar to [122], velocity profiles for the transport component in this study varied only slightly as a function of the visual uncertainty (Figure 2.6). This is in contradiction with other studies (e.g., [101, 20, 12]) that have found large changes in peak velocity as a function of peripheral viewing, no visual information, or glowing objects, respectively. Explanations for the inconsistencies follow similar reasoning as the ones provided for the observed differences in grip aperture timing. One possible reason for this discrepancy is that this project required movements to occur within 1200 msec, whereas, in the studies cited above reach time was unconstrained resulting in a 30% - 33% increase in movement time. In addition, [22] demonstrated that MGA is delayed with increased movement time, suggesting that MGA is better described in terms of distance to the target. In the current study, both the time to complete the reach and the distance to the target were fixed, which may explain why little variation in transport velocity was observed. Another possible reason is that the object in this study was resting in a cradle. This fact may encourage a uniform velocity profile as it is less costly to contact the object with non-zero velocity.

It was also demonstrated that position variance near contact increased with greater levels of visual uncertainty (Figure 2.8A). However, not all directions were equally affected as variance in the image-plane (y- & z-dimensions) increased linearly, whereas depth-plane variance (x-dimension) did not systematically change with visual uncertainty. This result could be due to the fact that the precision of visual information is better for the horizontal and vertical dimensions, but less precise for the depth plane information [113]. This finding reconfirms that finger variability near contact is changing with visual uncertainty. Since MGA is significantly correlated with contact location variance (Figure 2.8B) along the picture plane dimensions, it also suggests that MGA is encoding visual uncertainty.

This chapter has demonstrated that people adjust their reach and grasp behavior as a function of the visual uncertainty in the task. It appears that visual information scales with many reach parameters in a linear manner, which is interesting because it is known that MGA and finger trajectories also scale linearly with object size (e.g., [51, 52, 102]). The effect of visual uncertainty and object size on grasping is

qualitatively similar, although it appears that the magnitude of the trajectory and MGA changes due to visual uncertainty are considerably less. In fact, an estimate based the slopes reported in Figure 2.4 and in [102] (Figure 6A) suggests that the changes in MGA due to visual uncertainty and object size are such that a 1.5 mm change in the object size will result in a similar change in MGA as a 10 degree change in viewing eccentricity. Therefore, although the trajectory and MGA scaling are qualitatively similar, the magnitude of the changes appear to be quite different.

A limitation of this study is that it does not distinguish uncertainty of the hand from uncertainty of the target. Since the target and hand are *both* visible throughout the reach in this experiment, any observed changes in MGA cannot be uniquely attributed to feedback or feed-forward control. Although the model for MGA only considers target location uncertainty [17, 18, 69, 120], it could be that people are also using information about their hand position uncertainty to adjust their MGA. Future work could address how grasping behavior changes when *only* visual uncertainty about the target (or hand) is manipulated. Indeed, previous work has demonstrated that both uncertainty about the hand [63, 94] and target [101, 20, 12] impact reaching behavior. Whether these sources have a differential impact is unknown, and further work is necessary to tease-apart the relative contributions of these sources of uncertainty on the observed changes in grasping behavior reported in this study.

The next chapter will address some of the above concerns by exclusively manipulating target location uncertainty. Moreover, it will explore how another type of uncertainty (i.e., coordinate transformation uncertainty) influences both reach behavior and the reliability of target location estimates obtained by integrating visual and haptic information.

Chapter 3

Multimodal Cue Combination with Coordinate Transformation Uncertainty

3.1 Introduction

When humans reach out and grasp an object, information about the object's location arrives at different times through multiple sensory modalities, each in its own frame of reference. Maintaining an accurate representation of the object's location requires both integrating these sources of information, and updating the stored (remembered) location after changes in body configuration (e.g. eye, head, or hand movements). Recent physiological data supports the idea that visual, auditory and touch information are remapped and combined in multiple coordinate frames. In particular, the parietal reach region uses eye-centered coordinates [9, 16], while pre-motor cortex uses body-centered coordinates [41, 42].

There is also evidence in sensorimotor control that the brain stores object locations for reaching in eye centered coordinates [48, 9, 90]. Eye-centered storage requires remapping locations after every eye-movement [39, 28, 118, 48], while using this representation for reaching involves remapping to body-centered coordinates [24]. Although remapping eye-centered locations after saccades is quite accurate [44, 105, 76, 50, 89, 95, 86, 130, 15, 49], other types of eye movements may introduce substantial error [7]. Moreover, the precision of eye-centered coordinates for reaching

depends on accurate knowledge of the relative positions of body-segments between the eye and the hand. However, the brain's encoding of body articulation is imprecise, and the quality of the encoding is position dependent (e.g., [116, 84]). Uncertainty about the relationship between body segments due to imperfect sensory knowledge is termed *coordinate transformation uncertainty* (CTU - See Figure 3.1). When there is CTU in a system, estimates are degraded by each transformation made between coordinate frames.

Recent work [103] suggests that people select coordinate representations that minimize the impact of errors due to remapping. Using a virtual display, the authors introduced a systematic discrepancy between visual and proprioceptive coordinates of target locations. They showed that visual information is weighed more than proprioceptive information for reach plans to visual targets, but is weighed less for proprioceptive targets where it would introduce larger errors. They interpret these findings as evidence that the coordinate frame used to compute target location is flexible and selected to minimize the impact of coordinate transformation errors.

This work [103] demonstrated that people use knowledge about CTU when integrating *current* sensory information. The goal of this chapter is to investigate if the brain represents and compensates for CTU when making grasping movements to *remembered* (i.e., stored) target locations. To experimentally test for CTU compensation, head-fixed participants repeatedly reached to an occluded cylindrical target while fixating targets that spanned an 80 degree range of eye positions. To manipulate CTU between the eyes and head, the fact that error in eye position encoding varies both with saccade magnitude [116, 84] and the eye's eccentricity away from forward view [54] was exploited. In addition, it is known that grasping movements to visual targets compensate for object location uncertainty by increasing maximum grip aperture (MGA - See Chapter 2). Therefore, if the brain stores targets in eye-centered coordinates and compensates for the effects of CTU, it predicts that MGA should also vary as a function of eye position, even when the target is occluded.

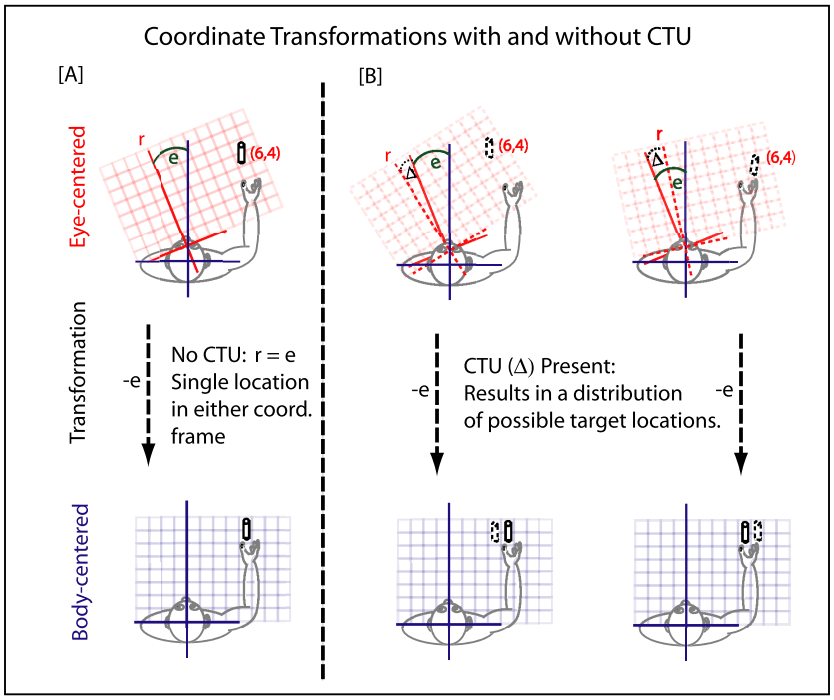

Figure 3.1: Figure illustrates the effect of eye-position CTU on the representation of object location in eye and head coordinates. An object's location in eye-centered coordinates (red grid) and head-centered coordinates (blue grid) can be different – it depends on the eye's position in the head. **[A]** In the case when the measured eye position (e) perfectly corresponds to the actual eye position (r), these representations can be uniquely converted. **[B]** Without this perfect correspondence, there are errors (Δ) between the measured (e) and actual (r) eye positions. The result of errors is that the same position in eye-centered coordinates (e.g.,(6,4)) corresponds to a range of possible locations (i.e., target location uncertainty) in head-centered coordinates, illustrated by two example dashed-line objects.

3.2 Materials and Methods

3.2.1 Subjects

Four subjects (2 males, 2 females) participated in the *visual condition* of this experiment, whereas five subjects (3 males, 2 females) participated in the *target occluded condition* of the experiment. The subjects were all right-handed students from the University of Minnesota and were given monetary compensation for their participation in the study. The subjects ranged from 19 - 32 years of age, and all had normal or corrected to normal vision.

3.2.2 Apparatus

Trajectory data were acquired by attaching three infrared emitting devices (IREDs) to the fingernails of both the forefinger and the thumb that were tracked via an Optotrak 3020 sampling at 100Hz (Figure 3.2E). Cylinder viewing distance was 52 cm, and forward viewing distance to the middle fixation point ("E") was 62 cm. Fixation letters (A-I) corresponded to degrees away from the cylinder (80-0, in 10 degree increments). Subjects began each reach from a starting block located 35 cm from a 2.2 cm diameter cylinder, 11.5 cm in length. The starting block was located 95 cm off the ground plane and approximately 10 cm below the cylinder plane. Once subjects initiated movement (> 2 mm) from the starting block, the trial was initiated. Subjects had 1200 msec to successfully lift the cylinder. The timer was stopped when a switch was tripped on the bottom of the target resting block, requiring subjects to lift the cylinder approximately 5 mm. The cylinder's position was maintained using a 2 cm tall clear plastic tube that was just large enough to allow smooth cylinder movement.

Occluders were used in the *visual* and *target occluded* conditions to remove any information about the hand and target. Moreover, liquid crystal shutter glasses [78] were used to remove visual information at different moments during the reach, depending on the condition (See Procedure Section for details).

3.2.3 Procedure

Head-fixed subjects made repeated reaches to the same spatially fixed target located 40 degrees to the right of the subject (0 degrees in cylinder coordinates - see Figure 3.2). The cylinder was not moved across the workspace to assure that any changes in reaching behavior resulted from changes in eye position/viewing eccentricity and not kinematic demands. Moreover, the cylinder's location was selected such that there was the maximal range of eye positions away from the target (0-80 degrees) to allow for the greatest range of uncertainty.

At the beginning of each trial, the fixation point was announced (e.g., "E"). After subjects fixated, they were allowed to make their reach to the target. They were instructed to reach as quickly (less than 1200 msec) and accurately as possible. Once they successfully lifted and replaced the target, their hand had to be returned within .66 mm of their original starting position before the next fixation point was

announced. If subjects took too long, or didn't lift the cylinder correctly, they heard an error message and the trial was repeated. Fixation points were randomly assigned throughout the trial. In the full range condition, each of eight possible fixations A-H, corresponding to {80,70,60,50,40,30,20,10} degrees target eccentricity were repeated seven times per block. The fixation point corresponding to 0 degree target eccentricity ("I") was excluded to keep similar fixations for both the visual and target occluded participants (since the occluder blocks both the target and the 0 degree fixation letter). Full range subjects ran in six blocks per day over three days, for a total of 1176 (7 repetitions x 8 Fixations x 7 Blocks x 3 Days) trials. Conversely, partial range subjects were cued to fixate one of three possible letters {H,E,B} corresponding to $\{10°, 40°, 70°\}$ away from the target. Each fixation was repeated 21 times per block for seven blocks, resulting in a total of 441 trials (21 repetitions x 3 fixations x 7 blocks x 1 day) . Note that the partial and full range conditions both have the same number of repetitions (147) per fixation point. However, the full range condition allowed us to capture the full functional form of the MGA profile, whereas the partial range condition allowed us to gain statistical power over a shorter period of time. Therefore, in the target occluded condition, 2 subjects were run in the full range condition and 3 subjects were run in the partial range condition. In the visual condition, all subjects were run in the partial range condition.

Subjects in the target occluded condition were never allowed to see their hand or the target as they were blocked by an occluder (Figure 3.2C,D). Moreover, all of the remaining visual information was removed after reach onset by shutter glasses. Subjects were instructed to maintain their eye position after the shutter lenses were closed. As a result, the only information subjects had available to them during a reach was stored haptic information from previous reaches to the target. If target location is stored in eye-centered coordinates, an estimate of their eye position is also necessary to remap the stored target location to body-centered coordinates. This manipulation allowed us to vary the amount of CTU in the task (i.e., the eye position uncertainty - see CTU discussion in Modeling Section) while keeping the reliability of the haptic information constant. Therefore, if subjects are estimating their CTU, maximum grip aperture is expected to vary with eye position.

Subjects in the visual condition were allowed to see the target throughout the duration of their reach. However, at no point were they allowed to see their hand. An occluder was used that allowed subjects to view the target but not see their hand until within

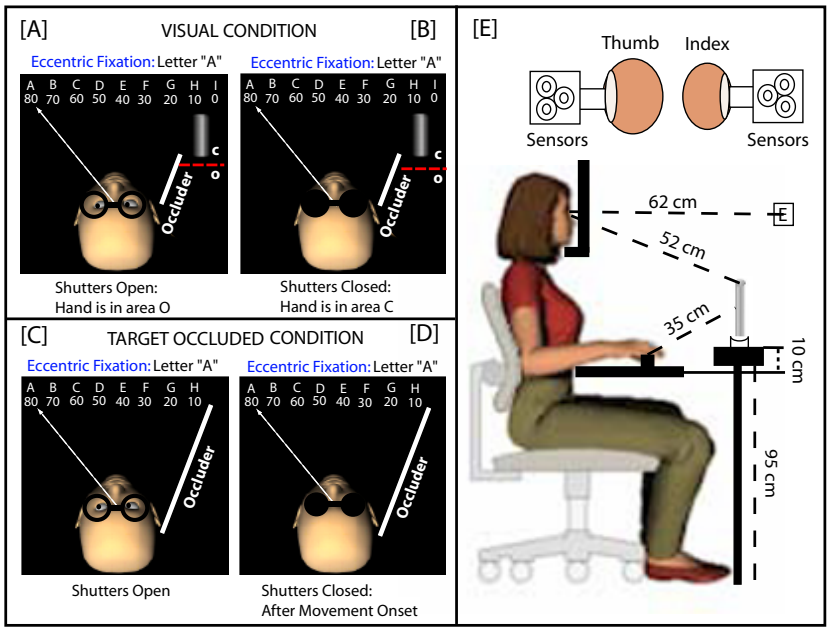

Figure 3.2: Experimental set-up used in psychophysical reaching task. See Methods Section for details. **Visual Condition: (A-B)** Subjects were allowed to see the target, but not allowed to see their hand. Visual information was manipulated by requiring subjects to fixate different points while making their reach. [**A**] When the subject's hand is behind the occluder (Area O), the target and fixation mark were visible. [**B**] However, once the subject's hand reached a point 1 cm from being visible (Area C), the liquid crystal lenses are closed to ensure the subject never viewed their hand before, during, or after the reach. This manipulation guarantees that any changes in MGA are due to changes in the visual uncertainty about target location. **Target Occluded Condition (C-D)** Subjects were not allowed to see their hand or the target at any point. [**C**] However, they were still required to fixate a point before (and while) making their reach. [**D**] After movement onset, vision was occluded using liquid crystal goggles for the duration of their reach, but subjects were still required to maintain their eye position. This manipulation allowed us to vary the CTU while keeping the haptic information specifying object location constant. [**E**] This panel displays the workspace used in the experiment, in addition to the sensor arrangement.

1cm of the target. In addition, liquid crystal shutter glasses were triggered to block vision of both the hand and target during the last cm, where the fingers would be visible from the occluder (Figure 3.2A,B). Subjects were instructed to maintain gaze fixation once the shutters were closed. This manipulation assured that changes in viewing eccentricity only influenced target uncertainty, opposed to uncertainty about both the hand and target. The visual condition allowed us to verify that MGA is encoding target uncertainty for the task.

3.2.4 Analysis

Cubic interpolating splines were fit to trajectories to allow for an analytic description of the trajectory and to compensate for occlusions. If the trajectory had total occlusion time $> 40msec$, it was discarded from the data-set. Only about five percent of the trajectories were discarded from the data due to occlusions that tended to occur at the beginning (i.e., launch off starting block) and return path of the reach. Grip aperture was computed as the distance between the center-points of the sensors on the fingernails. Maximum grip apertures (MGA) were averaged across subjects and blocks to produce Figure 3.5. MGA is the maximum distance between the thumb and forefinger during a reach, and it typically occurs 75-80% of the distance to the target object [101]. Its importance is that it serves as a measure of target uncertainty (e.g., [122]) and scales linearly with actual object size (e.g., [88]).

Change in MGA was computed by subtracting off the global mean of each block (of each subject, across eccentricities) from the mean MGA for each eccentricity. This was done to reduce the effects of inter-subject differences, remove drift in grip width across blocks due to fatigue, and to remove the effects of differences in IRED placement across sessions. Data were pooled across subjects and mean MGA was computed for each eccentricity.

3.2.5 Modeling

A probabilistic model of target location inference was developed with two distinct goals in mind. First, to provide concrete predictions for the effect of target location uncertainty on maximum grip aperture, relying on previous results to provide values for the parameters in the model. Second, to extend these predictions to include the possible effects of coordinate transformation uncertainty on MGA. Separate predictions for eye-centered and head-centered storage of target information are presented. Other possible storage schemes like body-centered, or storage in multiple coordinate frames [6] make the same predictions as head-centered storage. The equations and assumptions used for data modeling are presented here, while derivations are presented in Appendix A.

In the model, reach plans depend on an inference of target location using information remapped to body-centered coordinates. Target information consists of recent visual and haptic information combined with information stored in memory. Memory is

represented by a probability distribution on target coordinates, and operations on this distribution are used to represent the effects of sensory information, and coordinate transformation uncertainty. Two sets of expressions for target inference are presented - one for eye-centered and one for head-centered coordinates - to compare the effects of combining and storing information in different coordinate frames in the presence of CTU.

The principal variables in the model are target location in eye and head-centered coordinates, and visual, haptic and eye position signals. For simplicity, target location is represented in spherical coordinates with origin at the center of the head or at the midpoint between the two eyes, for head or eye-centered coordinates respectively. Modeling is restricted to the azimuthal angular component of target position (angle in the plane containing both eyes and the origin), because it is sufficient to account for changes in grasping and pointing behavior and to discuss optimality. Azimuthal target coordinates are represented by x in eye- and y in head-centered coordinates. For mathematical simplicity, the transform between x and y is approximated by $y = x + r$, where r is the azimuthal angle of eye position with respect to forward view. The approximation results from ignoring the offset between the origins in eye and head centered coordinates and only affects the model of pointing data, as discussed below.

Visual, haptic and eye position signals provide information specifying target location and appropriate coordinate transformations. The azimuthal angle of the retinal projection from eye position provides a noisy visual signal v for target location in eye-centered coordinates. The noise in v is modeled as zero-mean (for convenience and because the predictions do not depend on this quantity), with a signal dependent variance that models the effects of decreased spatial resolution on visual sensing of peripherally viewed targets. Using results from two-point discrimination studies [17, 18, 120, 69], visual uncertainty σ_v linearly increases with the eccentricity of the visual information (in mm):

$$\sigma_v = \sqrt{(.15 * v)^2} \tag{3.1}$$

This value was derived from the threshold $(v/30)^\circ$ reported in [69], and converted to mm. Note that similar models for visual uncertainty are used in both [84, 94].

Touching the object provides a haptic signal h to target location in body-centered coordinates available at the end of each reach . Haptic information from previous trials was included in the model because recent studies have shown haptic and visual information are optimally combined [114, 31] and that haptic experience affects visual judgments [5]. The haptic signal h provides information about y because the relationship between shoulder and head was held constant. As a result, any effect of the CTU between head and body-centered coordinates is similar for all data and can safely be ignored. The noise in h is modeled as zero-mean and constant variance σ_h^2. Based on haptic noise estimates derived from data in [31], the haptic uncertainty was set to be $\sigma_h = 15mm$. This is slightly larger than the $10 - 12mm$ standard deviations found in [114], however, the precise value of this variable did not affect the predictions.

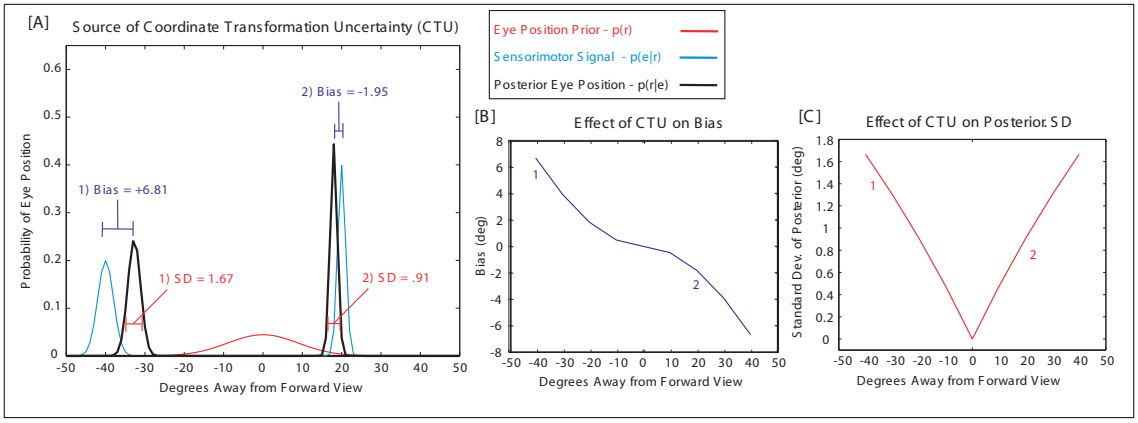

Figure 3.3: Figure depicting the source of coordinate transformation uncertainty (CTU) in the reaching task (See Modeling Section for details). **[A]** Information about eye position conveyed by sensorimotor signals (i.e., efference copy; proprioception; and retinal position of the fixation mark) is modeled by a likelihood function $p(e|r)$: the probability of sensorimotor signals e given the actual rotation of the eye r (blue line). This likelihood is combined with a prior on eye position $p(r)$ (red line) centered on forward view ($0°$) to form the posterior distribution on eye position $p(r|e)$ (solid black line). This distribution provides the information required for transformations between eye- and head-centered coordinate frames. Examples of likelihoods and posteriors are shown for eye positions $-40°$ (1) and $20°$ (2) away from forward view. The difference in widths of the example likelihoods illustrates how errors in sensorimotor signals vary with eye position. Panels [B-C] show consequences of the decreased reliability of sensorimotor signals away from forward view. **[B]** Eye position estimates derived from the mean of $p(r|e)$ are biased away from forward view. Biases result because the decreased reliability of sensorimotor signals e away from forward view produces increased dependence on the prior. **[C]** Shows how uncertainty in eye position estimates increase away from forward view, as measured by the standard deviation of the posterior distribution. This uncertainty propagates to information remapped between eye and head-centered coordinates, an effect called CTU. Numbered locations in Panels B and C illustrate where the examples from Panel A fall on the bias and standard deviation graphs.

Noisy eye position signals e summarize available information about the eye's position in the head, including efference copy of motor commands [72], proprioception [107], and the retinal location of the fixation point [84]. Uncertainty on e is modeled as zero-mean with a signal-dependent noise $\sigma_e = \sqrt{(.05 * |e| + 0.05)^{2\circ}}$ [116, 84]. While the noise is assumed unbiased, it is shown that Bayesian inference of eye position r from signals e produces estimates biased towards forward view (Figure 3.3). The bias results from the combination of eye position signals with a prior on eye position $p(r)$ that encodes the assumption that the angle between eye position and head direction is maintained around zero [106]. For the simulations, the prior was modeled as Gaussian with zero mean and constant uncertainty ($\sigma_r = 10.0°$), which biases eye position estimates by a gain factor $w_e = \sqrt{\frac{\sigma_r^2}{\sigma_e^2+\sigma_r^2}} = \frac{10}{10.05+.05|e|}$. The resulting posterior distribution is used to estimate eye position, and takes on the form $\mathcal{N}(r; \mu_{pe}, \sigma_{pe})$, where $\mu_{pe} = w_e\, e$ and $\sigma_{pe} = w_e\, \sigma_e$.

The decision to put signal dependent noise on eye-position rather than eye-movements deserves comment. Although remapping data between eye- and head-centered coordinates requires specifying absolute eye position, not just eye displacement, it is not completely clear how to characterize this error. Because subjects saccade to the fixation points from forward view for the experimental data presented here, uncertainty due to saccade magnitude cannot be distinguished from uncertainty due to eye position. However, it is maintained that both introduce uncertainty for the following reasons. For eye displacement tasks, eye position uncertainty is well predicted by saccade scatter [116, 84], which varies with saccade magnitude. Moreover, the force required to maintain eye eccentricity is roughly linear in eccentricity, due to signal dependent noise in force generation that is also linear [54]. Therefore, signal dependent noise should also result from eye position per se, independent of the saccade that brought the eye to that location.

3.2.5.1 Modeling Grasping Data:

It is assumed that reach trajectories in this experiment are planned to grasp the object based on the best current estimates of object location and diameter while avoiding object-finger collision. By assuming MGA scales with location uncertainty to avoid object-finger collision, MGA is modeled as proportional to the object's diameter D [88] plus the uncertainty in target's location in head-centered coordinates:

$$|MGA - D| \propto \sigma_y \tag{3.2}$$

Because the hand is occluded throughout the reach, the observed changes in MGA cannot be attributed to feedback control. Changes in MGA are attributed to the value of σ_y during movement planning. Note the assumption of proportionality – the model does not specify that people use a particular collision avoidance criteria. If the brain represents location uncertainty and uses it for reach planning, MGA should vary proportional to σ_y. Below it is tested whether MGA follows the functional trends predicted by y under different storage and sensory conditions.

Reach planning with collision avoidance requires estimates both of object location and uncertainty. These estimates are assumed to result from a Bayesian computation that combines the information available at the time of movement planning, which is specified by the following sequence of events. At the end of $t - 1^{th}$ trial, new haptic data from grasping the object is appropriately remapped (depending on whether storage is in eye or head-centered coordinates) and combined with memory. At the beginning of t^{th} trial, a fixation saccade changes eye position, and new visual information (when available) is combined with memory. The updated memory distribution is transformed to body centered coordinates and a reach plan generated. After reach execution new haptic information is acquired, completing the cycle. Figure 3.4 shows the differences in storage strategies when combining haptic information (from a previous trial) with newly acquired visual information. Note that this diagram omits remapping targets to hand-centered coordinates, which introduces additional coordinate transformation uncertainty. Uncertainty due to head-to-hand remapping is assumed constant because the target, hand and head were fixed during reach planning. However, targets stored in eye-centered coordinates acquire additional eye position CTU when target position is remapped to make a reach.

In Appendix A, expressions for σ_y for four different conditions are derived: target visible with eye-centered storage ($\sigma_{y,vis,eye}$), target occluded with eye-centered storage ($\sigma_{y,occ,eye}$), target visible with head-centered storage ($\sigma_{y,vis,head}$), and target occluded with head-centered storage ($\sigma_{y,occ,head}$), and show that the reliability of Bayesian inference depends on storage strategy due to CTU. We present expressions for σ_y for each condition below. To quantitatively predict MGA, degrees were converted to mm for eye-position information.

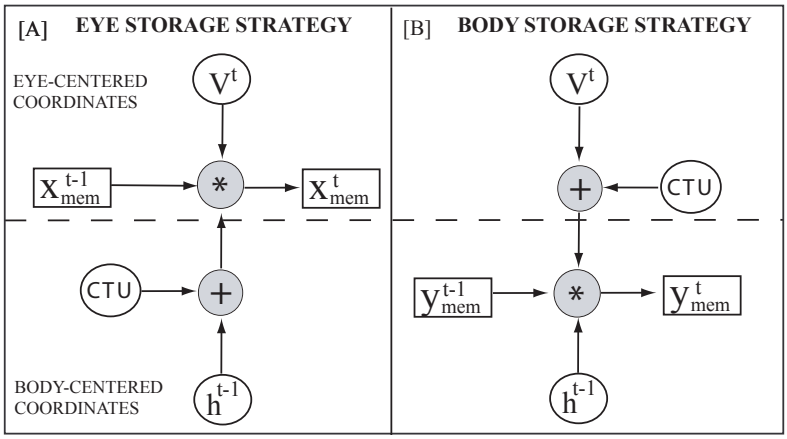

Figure 3.4: The diagram is a state-space representation of the computations, that treats the memory representations of target location as a random variables. Pluses represent effects that shift the mean and/or variance of a memory distribution, while stars represent probabilistic updates of both means and variances that result from multiplying likelihoods. Figure illustrates the effect of CTU on the flow of information between eye- and body-centered coordinates. Because the experimental set-up maintains a constant head-body relationship, this diagram will treat head- and body-centered coordinates synonymously. For the visual condition in this task, there was visual (v), haptic (h) and eye position (CTU) information available during the reach cycle. Moreover, this information is accumulated over time, forming a memory distribution (x_{mem} - eye centered; y_{mem} - body centered) on target location (See Appendix). **[A]** Storing target representations in eye-centered coordinates: The memory distribution from the previous trials (x_{mem}^{t-1}) is maintained in an eye-centered coordinate frame. This distribution is combined with haptic information that was acquired from the previous trial (h^{t-1}) by remapping the haptic information into eye-centered coordinates through a noisy transformation (CTU). At the beginning of the t^{th} trial, the newly acquired visual information (v^t) is used to update the target location representation, forming the target representation that will be used to plan for the t^{th} reach (x_{mem}^t). **[B]** Storing target representations in body-centered coordinates: The memory distribution (y_{mem}^{t-1}) is now maintained in a body-centered reference frame. Haptic information that was obtained from the previous reach (h^{t-1}) is used to update the memory distribution. At the beginning of the t^{th} trial, visual information (v^t) is acquired and remapped into body-centered coordinates via a noisy transformation (CTU). The visual information is then used to combined with the memory distribution to form the target estimate used to make a reach on the current trial.

Bayesian inference with *eye-centered storage* maintains a distribution on target location x that captures both a target estimate μ_{eye} and its uncertainty σ_{eye}^2. The distribution is corrected by sensory information remapped to eye-centered coordinates (when available), and adjusted for the effects of each eye movement. When the target is visible, visual and haptic information are combined with memory in eye-centered coordinates and passed to head-centered coordinates for reach planning. The uncertainty in target location resulting from this computation is given by:

$$\sigma_{y,vis,eye} = \sqrt{\frac{\sigma_{eye}^2(\sigma_h^2+\sigma_{pe,k-1}^2)\sigma_v^2}{\sigma_{eye}^2(\sigma_v^2+\sigma_h^2+\sigma_{pe,k-1}^2)+\sigma_v^2(\sigma_h^2+\sigma_{pe,k-1}^2)}+\sigma_{pe,k}^2} \tag{3.3}$$

where haptic information acquires additional uncertainty $\sigma_{pe,k-1}^2$ when brought into eye-centered coordinates from CTU due to errors in eye position sensing on the previous trial, and the whole expression acquires additional uncertainty $\sigma_{pe,k}^2$ from the transformation from eye to head-centered coordinates, necessary for making a reach. Target occlusion removes any visual information, resulting in the following expression:

$$\sigma_{y,occ,eye} = \sqrt{\frac{\sigma_{eye}^2(\sigma_h^2+\sigma_{pe,k-1}^2)}{\sigma_{eye}^2+\sigma_h^2+\sigma_{pe,k-1}^2}+\sigma_{pe,k}^2} \tag{3.4}$$

Bayesian inference in head-centered coordinates is similar, except a memory distribution is maintained on y with mean μ_{head} and variance σ_{head}^2. Visual information is remapped to head-centered coordinates (acquiring additional uncertainty $\sigma_{pe,k}^2$) and combined with memory before reach planning resulting in the following uncertainty expressions when the target is visible:

$$\sigma_{y,vis,head} = \sqrt{\frac{\sigma_{head}^2(\sigma_v^2+\sigma_{pe,k}^2)\sigma_h^2}{\sigma_{head}^2(\sigma_v^2+\sigma_h^2+\sigma_{pe,k-1}^2)+\sigma_h^2(\sigma_v^2+\sigma_{pe,k-1}^2)}} \tag{3.5}$$

and when the target is occluded:

$$\sigma_{y,occ,head} = \sqrt{\frac{\sigma_{head}^2\sigma_h^2}{\sigma_{head}^2+\sigma_h^2}} \tag{3.6}$$

Although the Equations 3.3 - 3.6 appear complicated, they simplify dramatically when the memory variance is much larger than the current information ($\sigma_{eye-head} \gg \sigma_v$). By assuming large memory uncertainty, the predictions only depend on models for visual, haptic, and eye position uncertainty. More specifically, Equation 3.3 reduces

to:

$$\sigma_{y,vis,eye} = \sqrt{\frac{\sigma_v^2(\sigma_h^2 + \sigma_{pe,k-1}^2)}{\sigma_v^2 + \sigma_h^2 + \sigma_{pe,k-1}^2} + \sigma_{pe,k}^2} \tag{3.7}$$

While Equation 3.5 simplifies to:

$$\sigma_{y,vis,head} = \sqrt{\frac{\sigma_h^2(\sigma_v^2 + \sigma_{pe,k-1}^2)}{\sigma_v^2 + \sigma_h^2 + \sigma_{pe,k-1}^2}} \tag{3.8}$$

And Equations 3.4 and 3.6 reduce to $\sigma_{y,occ,eye} = \sigma_h + \sigma_{pe,k}$ and $\sigma_{y,occ,head} = \sigma_h$, respectively. The simplified equations were used to make the predictions shown in Figure 3.5. For large memory variance, there is no learning across trials and the integration of information is a more general form of cue combination that incorporates coordinate transformation uncertainty and its impact on the reliability of information resulting from different storage strategies. It should be noted that inaccurate location memory is not unreasonable: memory standard deviations greater than 4 deg occur in tasks with similar delays involving humans [29, 57, 99] and monkeys [121, 7]. In addition, a limited memory model on the basis of cross-trial analyses could not be rejected – any influences of fixations from previous trials were too small to be reliably measured.

3.2.5.2 Modeling pointing data

To illustrate the effect of CTU on pointing behavior, data from a previously published experiment [71] was modeled. Similar to how MGA adjusts for uncertainty associated with the target location, pointing behavior is expected to change with the mean of the target location distribution. As previously mentioned in the Modeling Section, CTU biases target location estimates toward forward view as a result of the prior distribution on eye position (Figure 3.3). Therefore, if people are estimating their CTU, their pointing behavior is expected to change with these biases.

Participants in the [71] experiment were required to fixate an eccentrically presented

target. After the target was extinguished for a period of time, subjects were asked to point to the remembered location of the target. In this task, pointing to target required the subjects to point to the mean location of their estimated eye position, as eye position and target location were coupled at presentation. Pointing biases were recorded as the difference between subject settings and true target direction. Data were pooled across subjects and medians (with 95% confidence intervals) were reported. We extracted this data from Figure 2D of [71] using a computer program, replotting it with model predictions superimposed in Figure 3.5C. Full descriptions of data collection and methods are found in the original paper.

Since visual information is constant ($v = 0$) for all targets, pointing biases must result from biases in eye position estimates (like those shown in Figure 3.3B). The model predicts biased eye position estimates as a consequence of the use of priors in Bayesian inference. Previous work suggests the brain employs a prior belief that saccade magnitudes are small [84]. Here a similar assumption is made, that the brain has a prior belief that the highest probability eye position is forward view, with eccentric eye positions increasingly less probable. This prior has the effect of biasing eye position estimates toward forward view. When remembered visual targets are remapped to body-centered coordinates, the presence of eye position CTU is expected to bias target location estimates toward forward view as a result of the prior distribution on eye position (Figure 3). If the brain incorporates knowledge of the effects of CTU on remapped object location estimates, pointing behavior is expected to mirror biases in eye position estimates.

We use the eye-centered target storage model (Equation A.17) and the same parameter values to generate predictions, except haptic data are excluded. Neglecting the influence any systematic motor biases, pointing direction is assumed match the target estimate $\mu_{y,vis,eye}$. Based on the experimental design, we assume $v = 0$ and $\mu_{eye} = 0$, which reduces $\mu_{y,vis,eye}$ to $w_e e$. Predicted bias in pointing direction is computed as the difference between the observed pointing directions (modeled as $w_e e$) and the actual target direction (given by e), resulting in:

$$pointingbias = w_e e - e = \frac{10e}{10.05 + .05|e|} - e$$

which produces the curve shown in Figure 3.5C.

3.3 Results

Previous work shows that maximum grip aperture scales with object size (e.g., [87, 88]) and also increases when visual information is degraded (e.g., [122, 101]). It was tested whether reaching behavior similarly adjusts for changes in target location uncertainty introduced by varying CTU. Specifically, participants reached to the remembered locations of occluded targets while eye position was varied. The experimental logic is that if it is found that maximum grip aperture changes with eye position when the target is occluded, it suggests the brain uses knowledge of CTU for reach planning.

To make the predictions more precise, the amount of target uncertainty that should be introduced by varying eye position was modeled. A Bayesian model of eye position sensing was developed, using published data to provide realistic values for model parameters. Figure 3.3 shows the behavior of Bayesian inference of eye position. Because the biases and uncertainty that arise in the inference of eye position propagate to all information passed between eye and head-centered coordinates, the figure also illustrates the predicted consequences of CTU on estimates of target location. Examples of Bayesian inference of eye position are shown in Figure 3.3A. Due to signal dependent noise, eye position uncertainty (measured by the standard deviation of the posterior distribution) increases away from forward view, shown in Figure 3.3C. The use of the prior shown in 3.3A biases estimates of eye position towards forward view as shown in Figure 3.3B. Because remapping from eye to head-centered coordinates involves adding an estimate of eye position, biases in target estimates should mirror those for eye position. In addition, the uncertainty in remapped target information should vary with eye position like the curve shown in Figure 3.3C. In essence, these profiles are fingerprints for identifying whether information has been transformed between eye and head-centered coordinates. That is, if the brain is estimating CTU and using that information for reach plans, MGA should change similar to Figure 3.3C, and pointing to change with Figure 3.3B.

Experimental data are shown in Figure 3.5, with model predictions superimposed. At reach onset in the target occluded condition, all vision was extinguished (Figure 3.2C,D). Although both the fixation point and the target were not visible during a reach, subjects' maximum grip aperture increased for eye positions away from forward view (Figure 3.5A). Moreover, the smallest MGA occurred near forward view, rather

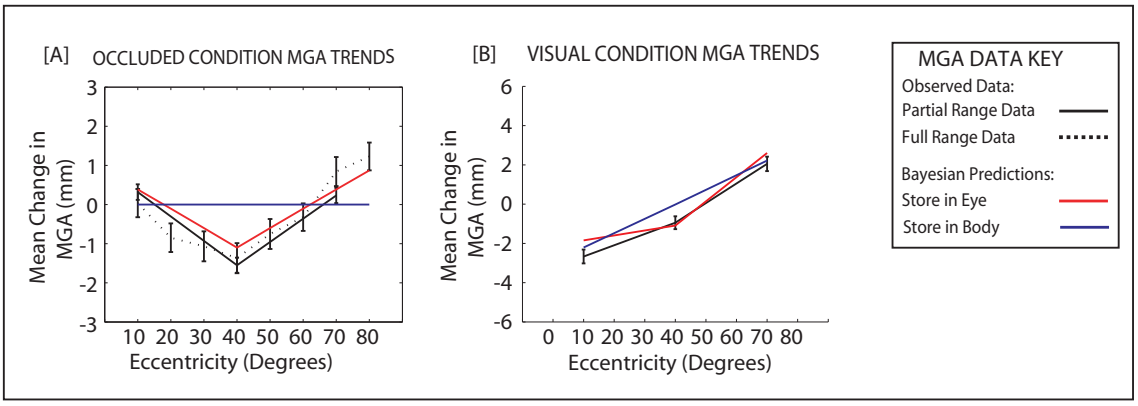

Figure 3.5: Figure showing the predicted and actual MGA trends for the experiment. [**A**] Target Occluded Condition: Mean change in MGA averaged across all subjects as a function of fixation eccentricity (the angle between fixation and target). Mean change in MGA ± 1 SEM is depicted by the solid black circles for subjects in the partial range condition (See Methods Section), and a dotted gray line for subjects in the full range condition. Bayesian model predictions are superimposed on the data for body-centered storage (solid blue line) and eye-centered storage (red lines) for two kinds of data pooling across trials. Predictions for the grasping data are rescaled to match the range of observed MGA changes. The data are consonant with an eye-centered storage strategy, and also show awareness and behavioral compensation for CTU resulting from eye-position uncertainty. [**B**] Visual Condition: Mean MGA values ± 1 SEM are depicted by the solid black line. MGA varied close to linearly with eccentricity, verifying MGA captures position uncertainty. Superimposed model predictions are similar but differ in convexity: approximately linear for body-centered (blue line) and concave for eye-centered (red line), with data similar in shape to the eye-centered predictions.

than at the target location, showing that the effect is a consequence of eye position and not target location. Although subjects made hundreds of identical reaches in this experiment over the course of many hours across several days, uncertainty about object location varies with eye position as predicted by storing information in an eye-centered reference frame (Figure 3.5A). In contrast, if object location were stored in a body-centered coordinate frame (e.g. head or hand), there should be no effect of eye position (Figure 3.5A).

To verify that MGA is a measure of object location uncertainty, the same experiment was run on a different group of subjects with a visual occluder that allowed view of the target, but not the hand (Figure 3.2A,B). Visual location uncertainty increases linearly with the eccentricity of the target, due to changes in retinal acuity in the periphery [17, 18, 120, 69]. The data show an almost linear change in MGA with retinal acuity, verifying MGA as a measure of target location uncertainty. Moreover the deviation from linearity is in the direction predicted by eye-centered storage, but not head-centered (Figure 3.5B).

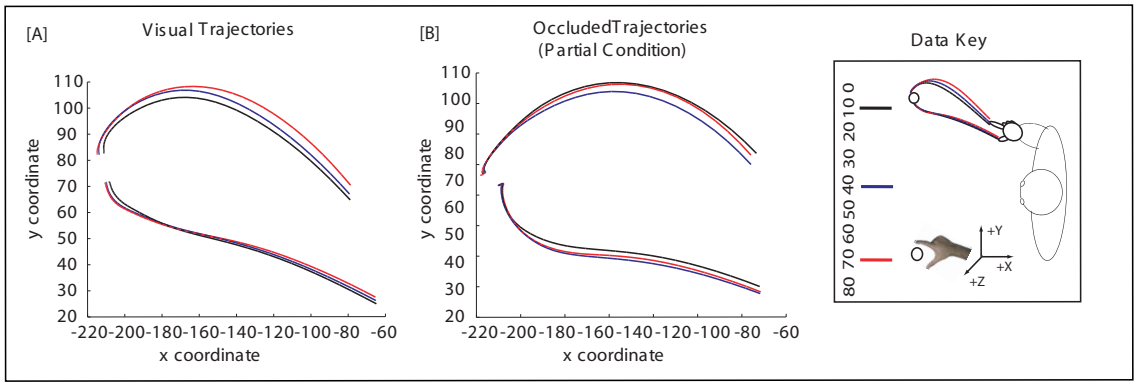

Figure 3.6: Figure demonstrates the effect of uncertainty on finger (top traces) and thumb (bottom traces) approach trajectories. Each trace represents the mean trajectory for that fixation point, across all subjects. The color of the trace represents the degree that fixation deviates from the target location ($10° = black, 40° = blue, 70° = red$). **[A]** Visual Condition: As the amount of visual uncertainty increases, finger trajectories become more hooked, whereas thumb trajectories do not vary substantially. **[B]** Occluded Condition: Although the magnitude of change is smaller, there is a reordering of the finger approach trajectories such that the trajectory that corresponds to the forward view location (40°) now has the least amount of hook. Both of the other fixation points (10°&70°) appear to have similar amounts of hook. This result is predicted by the increase in eye position uncertainty away from forward view, supporting the notion that people are estimating their CTU when making a grasp. Note that the finger trajectory in the target occluded condition also appears to "wrap-around" the object. Therefore, it also appears that subjects contacted the object with a different part of their finger when no visual information was available.

The MGA results summarize systematic changes in subjects' approach trajectories that occur when the amount of target uncertainty increases, for both occluded and visual data (Figure 3.6). Trajectories from the visual condition show that the effect of increased uncertainty is to widen the excursion between finger and thumb and make the finger trajectories became more "hooked" (Figure 3.6A). Trajectory changes in the target occluded condition are similar, but the widening pattern is reordered such that the trajectory associated with forward view has the least amount of hook (Figure 3.6B). The similarity in the trajectory changes in the visual and occluded conditions suggest that both are the result of target uncertainty.

It was also investigated whether the model could account for biases in pointing data that have been used to infer target storage in eye-centered coordinates (see [24] for review). In these studies, target direction is briefly presented using auditory [70, 90], visual [48, 71, 77, 1, 2], or proprioceptive [90] information while eye position is varied, after which subjects point to remembered target direction. These studies show systematic changes in pointing error with eye position as qualitatively predicted by eye-centered storage. Most of these studies have complicated designs (e.g. multiple

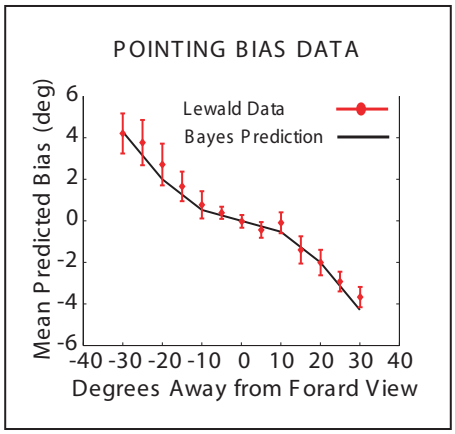

Figure 3.7: Figure showing CTU predicted and actual pointing bias trends. Pointing bias data extracted from Figure 2D of [70] (black dot) ± 95% confidence intervals with Bayesian model predictions superimposed (red solid line). The good agreement between predictions and pointing data suggests biases may be the result of CTU.

fixations per trial [48] and/or mix biases due to coordinate transformation effects with those due to visual sensing and motor production. However, the model generates simple predictions for the data presented in [71] because eye position forms the only available information to target direction. There is good agreement between data and model predictions, shown in Figure 3.7.

3.4 Discussion

The experimental data show that participants adjust their grasping behavior to compensate for target location uncertainty introduced by errors both in visual and eye position sensing (CTU). These results suggest that the brain represents CTU and incorporates its effects on estimates of target location. Previous studies have shown that the brain represents the reliability of sensory data both within (e.g., [66, 94]) and between modalities [5, 31, 10], in addition to representing the end-point variance of motor movements [46, 110, 111]. By representing CTU, the brain has all the components required to predict performance of complex sensorimotor behaviors involving perception-action cycles.

Although it is maintained that the change in grasping behavior in the occluded condition to target uncertainty, it may be that holding eye position away from forward view creates an attentional load that affects grasping. This possibility is believed to be implausible for two reasons. First, it is unclear why decreased attention would create

the same kind of trajectory changes as visual uncertainty, however, it was shown that eye position affects finger trajectories in the same way as changes in the amount of visual information. Second, reaching normally requires focal attention to be shifted to the target [104]. However, it is argued that allocating focal attention to the target should be easier when fixating the target than when looking straight ahead.

In addition, the effect of CTU on grasping in the target occluded condition is consistent with an eye-centered memory representation for object location. The only sensory information in this condition is the brief haptic contact with the object at the end of each reach. Despite the lack of visual information, grasping varies with eye position consistent with bringing haptic information into eye-centered coordinates using a noisy transformation. These results extend previous neural (e.g., [9, 91] and psychophysical evidence (from pointing) (e.g., [48, 71, 77, 24]) for eye-centered target storage by showing that eye-centered storage persists in a task for which there is no visual information specifying target location. Note that these results do not preclude the possibility that target information is stored in multiple coordinate frames [6]. In this case reach planning may not be exclusively based on an eye-centered target representation, however, the results suggest that an eye-centered representation is both updated without vision and is incorporated in reach plans.

Although it seems intuitive that an eye-centered coordinate frame is employed when visual information is present, it is less clear why non-visual information is also being stored in eye-centered coordinates (i.e., target occluded condition). One possibility is that the brain uses a robust strategy for information storage. If the brain assumes that the loss of visual information is temporary, then eye-centered storage allows rapid prediction of the target's location which is useful for error correction if the target reappears. It may be the case that people who have had extended periods of poor visual information (i.e., low-vision or blind) will not use eye-centered target representations. Another possibility is that visual fixation marks promoted an eye-centered storage strategy. In particular, fixation marks may have been used to maintain an accurate representation of the body's configuration with respect to the apparatus. Computing hand and target location relative to a fixation point would be equivalent to computing hand and target locations in eye-centered coordinates. In future work it could be investigated if MGA changes across eye position remain when a tactile fixation point is used in the target occluded condition.

It was demonstrated that grasping behavior adjusts for coordinate transformation uncertainty introduced by errors in eye position sensing, suggesting the brain has an internal model capable of predicting the consequences of CTU. In addition, a Bayesian model was provided that quantitatively describes the impact of CTU on both the reliability and bias of the posterior distribution on target location. The model was used to offer an explanation for previously reported biases in pointing, in addition to predicting the observed psychophysical data. Together, these results suggest that CTU may affect behavior in tasks where multimodal cue combination is being performed and body articulation is not fixed. Because previous research on multimodal cue combination has primarily focused on tasks with constant body articulation (e.g., [5, 31, 10]), CTU could safely be ignored. However, in tasks with variable body articulation the coordinate transformation uncertainty introduced by errors in joint sensing may cause behavioral changes that will be inexplicable if CTU is not taken into account.

3.5 Uncertainty Conclusions

It is clear that people are incorporating estimates of their uncertainty both within (Chapter 2) and across (Chapter 3) perceptual modalities when making a grasp. Moreover, it appears that people use knowledge of their CTU when performing a grasp. Together, these results imply that people have an understanding of their uncertainty. Therefore, if the brain is going to select actions minimize the risk associated with a task, the only thing that is lacking is evidence that it understands and acts according to a natural loss function (i.e., one that results from the task demands). The next chapter will reverse-engineer a loss function for grasping tasks, in addition to testing the model by requiring people to lift objects at different orientations.

Chapter 4

Natural Loss Function for Grasping Behaviors

4.1 Introduction

Natural tasks are usually specified in terms of high-level goals like lifting an object. However, to execute a movement, high-level goals must be converted into a detailed control strategy at the level of forces and torques (or muscle contractions). This problem is difficult because the conversion is underconstrained – there are an infinite number of ways to reach to and lift an object. As a result, understanding how the brain accomplishes this conversion is a major goal in motor control research.

Optimal control theory provides an elegant solution to this dilemma, explaining why people reach in a very predictable and stereotypical manner, even with ambiguity (e.g. [46]). The basic idea is that a specific movement strategy is selected by minimizing an expected loss function defined on movement parameters and task goals. For example, researchers have differentially paid subjects based on their performance (e.g., [110, 111]) in a reaching task. Essentially, researchers are experimentally imposing the reward function onto the task. They found that people adjust their mean endpoint location such that it maximizes the expected gain for the task. However, a limitation of this approach is that the loss function is not natural - people do not typically get paid when reaching to and grasping objects. Therefore, it is desirable to attempt to reverse-engineer a loss function that can account for natural movements.

There has been a considerable amount of effort over the past thirty years to reverse engineer a loss function for reaching (e.g. [34, 46, 109]). One such effort maintains that the motor system attempts to perform the smoothest (i.e., minimum jerk) movement towards a goal state [34]. This optimization principal was able to predict the trajectories people took when reaching from Point A to Point B, through an intermediate point (i.e., via-point task). The limitation to this approach is that the goal state (i.e., Point B) is experimentally imposed. In natural grasping behavior people are not told where to place their fingers on an object. Moreover, specifying a goal state also imposes the terminal velocity of the finger (i.e., zero). However, when moving to grasp an object, it may be beneficial to contact the object with a non-zero velocity in order impart the momentum of the hand to the object, allowing for smooth object movement.

Another attempt at reverse engineering a loss function maintains that the motor system is concerned with making accurate movements, in the presence of signal-dependent motor noise [46]. Such a minimum error optimization principal was also able to accurately predict the trajectories people took when performing a via-point task. However, it suffers from the same limitations of imposing the contact conditions (i.e., location and velocity) people must use as a result of the task constraints.

Finally, recent research has suggested that the sensorimotor system attempts to minimize the number of controls it makes during a reach trajectory [109]. This minimum control framework maintains that the system only controls for variability that will adversely impact performance and was able to accurately predict the trajectories used in a via-point task. However, this approach also suffers from the limitation that the contact conditions were imposed on the task.

Due to the above limitations, there are still many gaps in our knowledge of human sensorimotor control. For example, one of the key planning problems for lifting an object is where to place finger and hand contacts. Without this information, trajectory plans cannot be formulated. Nevertheless, previous proposals do not address this problem. Instead they focus on understanding trajectory planning after the contact conditions (location, velocity and acceleration) have been experimentally imposed. This approach has yielded a series of loss functions that can successfully predict trajectory data (e.g. minimum jerk [34]; minimum end-point error [46]; minimum cost-to-go [109]; etc.). However, specifying good contact conditions is important be-

cause they can vary drastically across tasks. For example, in a task where the subject has to lift an object, finger locations determine the relative controllability of the object. Moreover, the velocity and acceleration at contact could be planned to allow the hand's momentum to lift the object.

Indeed, previous research has demonstrated that peoples' hands shape differently during a reach [3] and that the hand contacts the object at different locations [21] depending on the manipulation being performed on the object. Moreover, recent research [74] has shown that people select different contact locations based on the predictability of the object's center of mass. Although these empirical finding allude to the impact of the task being performed on finger contact locations, they do not provide a cogent theory that is able to predict these changes across tasks.

Therefore, the goal of this chapter is to propose a theory for how the brain solves the contact selection problem. The theory maintains that the brain selects contact conditions through an internal model of the physics of object manipulation – how and where forces must be applied to effectively move objects. It will be demonstrated that this knowledge, coupled with a strategy of moving objects with a minimum control strategy, can be used to derive a natural loss function for optimal contact point selection. The empirical validity of these ideas are investigated through human reaching experiments that require subjects to lift or touch objects arranged at different orientations.

4.2 Reach Hierarchy

The core idea of this proposal is that the strategy for lifting an object begins by planning for the intended motion of the object. It is further assumed that object trajectory planning follows the same minimum control principles as hand and finger trajectory planning. Intuitively, when an object is manipulated, it is functionally an extension of the hand and is (presumably) controlled by the same system that controls an empty hand. Specifically, this theory proposes that the object's motion plan is controlled by encoding the desired object via-points and goal-state (locations, velocities, accelerations) in a loss function on object motion. Furthermore, it is assumed that an optimal control strategy is used to find the set of forces required to generate an optimal object trajectory. However, the required object control forces are a function of finger placement and external forces (like gravity; Figure 4.3C). Using

equations of rigid body motion, contact dynamics and kinematics, it will be demonstrated that the object motion loss function can be rewritten in terms of a loss on contact conditions. Optimizing this loss function results in contact conditions that can achieve the desired object motion with minimum force. Finally, these optimal contact conditions induce a loss function on approach trajectories.

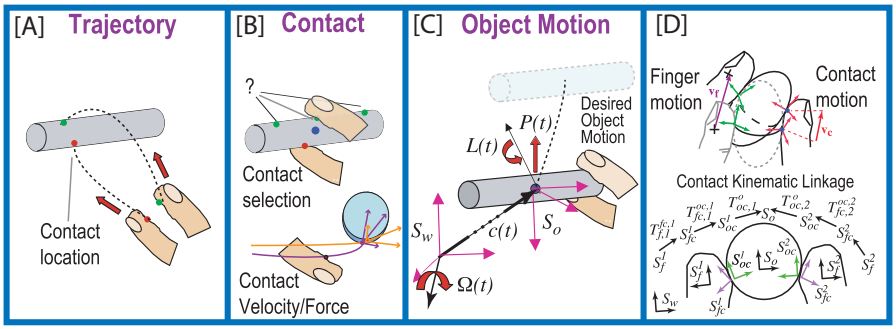

Figure 4.1: Demonstrates the three components of a reach task. Red dots signify the thumb's position. Green dots represent the index finger's position, and blue dots show the object's center of mass. [**A**] *Approach trajectory.* Optimal control trajectory to contact conditions. [**B**] *Selecting contact conditions.* Select contact conditions that can produce the desired object motion with minimum force and torques. [**C**] *Planning for object's motion.* Optimal control of object motion. Notice that this separation has a hierarchial arrangement (from right to left). [**D**] Illustrates relationships between object motion and finger motions. Given frictional contact without slippage, the object and fingers are connected by a kinematic chain. From this view, contact selection amounts to choosing a particular object-hand linkage.

This proposal for converting a goal for object movement into a motor plan can be thought of as forming a planning hierarchy (Figure 4.1). At the first level, the task goal is converted into object motion goals. At the next level the intended object motion is used to determine good contact conditions for control of the object. Finally, the contact conditions form the goals for approach trajectory generation. For a given desired object motion, only a subset of finger contact configurations will minimize the risk. Thus the proposed framework can be used to predict what contact conditions people should use to achieve a desired object motion.

4.3 Methods

4.3.1 Experimental Details

To test if people were reaching in a manner predicted by the natural loss function, subjects were required to touch and lift cylindrical objects at varying orientations

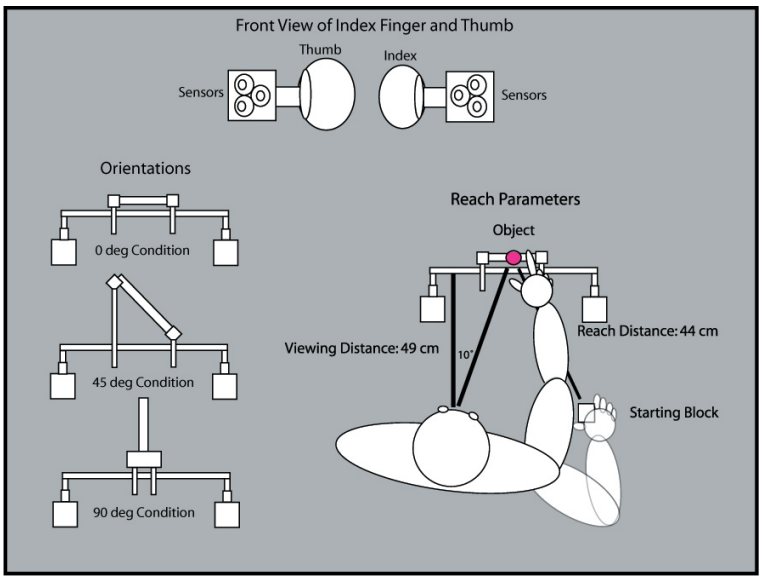

Figure 4.2: Reach set-up employed in this experiment. Subjects were required to reach to objects at varying orientations (horizontal, 45 degree, vertical) while their fingers were tracked. See text for details.

(Figure 4.2: horizontal, 45 degrees, and vertical, with respect to gravity). When the fingers support an object, gravity induces torques on the object that varies with the object's orientation and finger locations. Notice the effects of gravity must be taken into consideration when lifting the object, but not when touching the cylinder. Therefore, it is predicted that subjects in the lift condition will place their fingers in locations that achieve the desired object motion while canceling gravity-induced torques. Whereas, in the touch condition, subjects will successfully contact the experimentally imposed location.

Four head-fixed subjects were required to reach to a spatially fixed object located approximately 44 cm away, with a viewing distance of 49 cm. Subjects were instructed to reach as quickly (less than 1200 ms) and accurately as possible. Once the reach was completed, their hand was returned to the starting block for the next trial. Infrared emitting devices were attached to the fingers and their movements were recorded via an Optotrak 3020 sampling at 100Hz. Each subject ran in one reach condition (i.e., touch or lift) at one object orientation (horizontal, 45 degrees or vertical) per day. For each condition, the starting block was rotated to make the relative finger paths in space nearly identical for the three orientations. The session terminated after they completed 120 reaches per condition. Every subject ran in each of the six possible conditions.

A calibration procedure was used to determine a probabilistic relationship between a fingertip model and the finger sensors and the position of the object relative to the Optotrak. Fingertips were modeled as elliptical generalized cylinders (15mm peak width and 10mm peak thickness) whose radius followed a profile roughly matched to the author's finger. Finger-object impact events were determined from the trajectory data the large magnitude jerk events with the highest probability of first collision (based on the calibration data). Contact positions were computed from the highest probability contact location at the end of the impact event, while contact velocities were referred to time immediately preceding impact.

The heart of this proposal is that the brain understands object manipulation and can bring object motion within its control framework. Hence, it is necessary to explain how object motion can be rewritten in terms of the motions of finger contacts and forces. It will then be shown how to rewrite the cost function for the object movement task in terms of contact conditions, which allows the selection of optimal contact conditions. Finally the predictions of this approach will be compared with contact data from the experiment.

4.4 Controlling Object Motion with the Fingers

A fundamental idea in robotic manipulation is that a stably grasped object forms a closed kinematic linkage with the hand through frictional contact points ([80, 13]). What this means is that (with force closure) the finger contacts act as virtual links that attach the object to the hand (Figure 4.1D). Roughly speaking, human fingertip contact (without sliding) is instantaneously equivalent to a (virtual) spherical joint with two degrees of freedom. Given a description of the object-finger contacts, the motion of the object is determined by the motions of the fingers, as long as contact is maintained.

The motion of a rigid object can be described using generalized coordinates and generalized momentum. Generalized coordinates $X(t) = [\vec{c}(t),\ \Omega(t)]^T$ combine a vector to the object's center of mass $\vec{c}(t)$ with a vector that parameterizes the three degrees of freedom in the rotation $\Omega(t)$ between the reference and object's coordinate frames. Specifically, the direction of $\Omega(t)$ is the rotation axis, and the length is the angle. Then the rotation is given by the matrix exponential $\mathbf{R}(t) = \exp(\Omega^{\times}(t))$, where $\Omega^{\times}(t)$ is the skew symmetric matrix that would accomplish a cross product by

$\Omega(t)$ via a matrix multiply. Generalized momentum is similar, combining the linear momentum vector $\vec{P}(t) = m_o \frac{d\vec{c}}{dt}$ with the angular momentum $L(t) = I(\Omega(t))\omega(t)$ into one vector, where m_o is the object's mass, and $I(\Omega(t)) = \mathbf{R}(t)\, I_{body}\mathbf{R}^T(t)$ is the inertia tensor in the world frame and I_{body} is the fixed inertia tensor in the object's frame.

Using generalized coordinates and momenta, the equations for the dynamics of an object's motion are easy to write down. The special case of object manipulation using two finger contact in stable force closure (without slippage) is considered. The dynamics of the object's motion as a function of the generalized finger contact forces $\vec{u}_i$ for two fingers is given by the Newton-Euler equation:

$$M(\Omega(t))\ddot{X}_o(t) = \vec{u}_1(\vec{r}_1) + \vec{u}_2(\vec{r}_2) + \vec{G}(\vec{r}_1, \vec{r}_2) \tag{4.1}$$

where $X_o(t) = [\vec{c}(t)\theta(t)]^T$ is a Cartesian vector representing the position and orientation of the object, $M(\Omega(t))$ is generalized mass matrix specified by the object mass and inertia tensors, $\vec{r}_i$ are the contact locations in object coordinates, and $\vec{G}$ the force and torque from gravity. These forces and torques need to be supplied by the fingertips while preserving finger-object contact.

Preserving object contact constrains both finger forces and finger motions. To maintain static frictional contact, finger forces $\vec{u}_i$ must have tangent plane components less than the static coefficient of friction μ times the surface normal component. Finger motions required to maintain contact can be written in terms of the object's motion, and the location and local geometry of contact points. Let S_o,$S_{oc,i}$, $S_{fc,i}$, and $S_{f,i}$ denote coordinate frames for the object (at the center of mass), the Gauss frame defined by the object's surface normal and tangent plane at the i^{th} contact point, finger i's Gauss frame at it's contact point, and finger i's reference frames, respectively. Then a chain of transforms can be defined $T_{f,i}^{fc,i}$, $T_{fc,i}^{oc,i}$, $T_{oc,i}^{o}$ between the coordinate frames above, where subscripts denote the origin frame and superscripts the destination. Moreover, the velocity of the object can be written in terms of finger velocities using Jacobians derived from the transforms above and the kinematic contact constraints (see [80] for details):

$$\dot{X}_o(t) = \mathbf{J}_{f,i}\left(X_{f,i}(t), \vec{r}_i\right) \dot{X}_{f,i}(t) \tag{4.2}$$

Given non-sliding contact and force closure, these Jacobians are sufficient to determine the finger motions given an initial contact point. Thus the object motion can be rewritten in terms of desired finger motions and contact forces, which are straightforward to convert into joint torques given inverse dynamics models for finger motion. Abstractly, there are different system models for the fingers before: $\dot{X}_f = h_{free}(t, X_f, \vec{u})$, during: $\dot{X}_f = h_{contact}(t, X_f, \vec{u}, \vec{v}_c, \vec{r}_c)$ and after contact $\dot{X}_f = h_{f+obj}(t, X_f, \vec{u}, \vec{r}_c)$, which leads to different control laws $\vec{u} = \pi(t, X_f, \vec{v}_c, \vec{r}_c)$ applying in each period, where $\vec{u}$ is the vector of required finger forces. Note the dependence of the contact and object systems on contact location $\vec{r}_c$ and contact velocity $\vec{v}_c$.

4.5 Loss Function for Object Motion Planning

This loss function combines two elements, task constraint costs penalizing deviations from desired object via states $Y^*_{o,j}(t_j)$ and finger control costs that are always present:

$$\begin{aligned} C_{obj} &= \sum_{j=1}^{Nvia} Q_j\left(X_o(t_j), Y^*_{o,j}\right) + \sum_{i=1}^{Nfing} \int_0^{t_{fin}} R_i(\vec{u}_i(t))dt \\ &= \sum_{i=1}^{Nfing}\sum_{j=1}^{Nvia} Q_j\left(T^o_f(t_j)X_{f,i}(t_j), Y^*_{o,j}\right) + \int_{t^+_{con}}^{t_{fin}} R_i(\pi^i_{f+obj}(X_{f,i}, \vec{r}_c))dt + \\ &\quad \int_0^{t^-_{con}} R_i(\pi^i_{free}(t, X_{f,i}))dt + \int_{t^-_{con}}^{t^+_{con}} R_i(\pi^i_{contact}(X_{f,i}, \vec{v}_c, \vec{r}_c))dt \end{aligned}$$

Intuitively, minimizing this loss function will result in a set of finger positions that achieves accurate object motion, with minimum pinch-force. Theoretically, the optimal control strategy will minimize the expectation of this loss function over control laws $\pi(.)$. Because the control law after and during contact depends on contact conditions, the optimal overall control law will indirectly choose optimal contact conditions. In this formulation, optimal contact points could be selected online as a part of an expected cost-to-go minimization, or off-line to serve as intermediate task goals.

To make testable predictions from this theory, the following simplifying assumptions

were made: $Q_j\left(X_o(t_j), Y^*_{o,j}\right) = \alpha||X_o(t_j) - Y^*_{o,j}(t_j)||^2$, and $R_i(\vec{u}_i(t)) = \beta||\dot{u}_i(t)||^2/dt^2$, where α and β adjust the relative importance of the two costs. In addition, because the cost on controls is related to the finger jerk, it is expected that the hand and object trajectories produced by an optimal control will be close to minimum jerk predictions on average. These simplifications were used to generate loss functions on contact location as described in the methods section.

4.5.1 Loss Function Computation

To simplify risk function evaluation, the rigid body manipulation equations above were simulated over a dense grid of contact finger locations. Because the movements involved no intended rotation, an overall net force impulse could be precomputed that would drive the object along the desired minimum jerk trajectory to ($[0, 0, 50]$) in 500 ms. At each time step finger contact forces and finger movements of the above model finger were computed that would produce the value of the net force impulse while maintaining contact with least effort. Simulation parameters were: friction=1.1, hand mass of .0474 kg, object's mass 0.185 kg, and inertia was computed from the standard formula for a cylinder. Different measures of cost were then computed from the computed control signals and finger motions. Object motion cost is a sum of the deviation from desired object end state and the amount of relative motion between object and fingers introduced by rolling of the contacts. Object jerk and control cost measure the object jerk (linear and rotational) and the change in required controls, respectively. These cost measures differed very little qualitatively so only object cost is shown. Finally, the risk associated with a contact location was computed by convolving the evaluated cost function with the contact motor error distribution estimated from each subject's touch data.

4.6 Observed Contact Positions and Velocities Depend on the Task

The theory makes some simple qualitative predictions. Contact locations that make it easy to move and stabilize the object should be selected. Moreover, the direction and magnitude of contact velocities should produce a smooth, low control transition between the approach and manipulation components of the reach. External forces that change the location where stabilizing forces should be applied can be used to

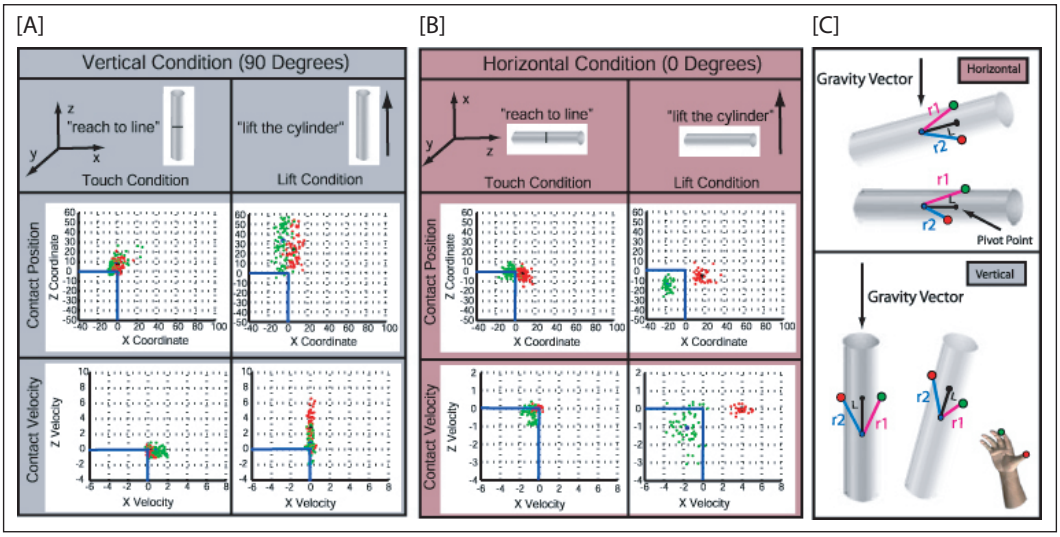

Figure 4.3: Empirical contact condition data for different tasks. End-point locations and velocities of the thumb (red dots) and index finger (green dots) are given. Blue lines on graph demonstrate the location of the center of mass $(0, 0, 0)$, and the zero velocity (mm/csec^2) location. **[A]** Contact location and velocity data for a vertically oriented cylinder. Notice in the touch condition people were reaching to the center of mass and terminating with zero velocity. However, in the lift condition, people were contacting the object above the center of mass and with positive z-velocity. **[B]** Contact location and velocity data for a horizontally oriented cylinder. Notice in the touch condition people were reaching to the center of mass and terminating with zero velocity. However, in the lift condition, people were contacting the object with a staggered finger arrangement (thumb slightly to right of the center of mass, but finger more right) and with positive x-velocity. **[C]** Diagram illustrating the important concepts for the empirical predictions. r_1 & r_2 are the distance of the finger's and thumb's contact location from the center of mass, respectively. L is a point that bisects an imaginary line between the contact locations.

test the contact location prediction. A simple way to accomplish this is to vary the orientation of an object with respect to gravity. As shown in Figure 4.3C, stably holding a cylinder requires canceling torques due to gravity. Gravity induced torques depend on the distance between the cylinder's center of mass and the pivot point (L) and the orientation of the cylinder. The amount of torque that can be applied by the fingers is proportional to the distance between the fingers. Thus, planning for object motion would predict that subjects should adopt a staggered finger arrangement at contact when lifting the cylinder horizontally, parallel finger locations when lifting the object vertically, and intermediate finger offsets for orientations between the two extremes. A plausible alternative to planning for object motion is to choose a generic strategy which selects contact conditions that work in many situations. For lifting the cylinder in this task, such a strategy exists: reach for the center of mass with a parallel finger arrangement.

Assuming that people are planning for the object's motion also leads to contact velocity predictions. More specifically, it predicts that contact velocities will be selected that achieve the desired object motion for a given object orientation. To experimentally test these velocity predictions, subjects were additionally required to touch (i.e., no object motion) the same cylinders. It should be noted that there were two separate touch conditions where subjects were asked to touch imaginary lines and real lines at the object's center of mass for each orientation.

4.7 Empirical Results

Figure 4.3A,B shows the results from a typical subject in the horizontal and vertical lift condition. It is apparent that the subject is following the object motion planning prediction: lifting with a staggered finger arrangement in the horizontal condition (where the thumb is slightly to the right of the center of mass, and the finger is further right of the center of mass), and lifting with a parallel configuration in the vertical condition. All subjects showed an identical trend, with the finger offset in the 45 deg condition falling between the 0 and 90 deg conditions. These results are not simply due to motor limitations–subjects were quite good at touching the imaginary line in all orientations. In fact, contact locations for the touch condition did not vary significantly across orientations.

In addition, contact velocities vary with object motion and orientation in ways qualitatively predictable from planning for object motion. Figure 4.3 shows that a typical subject in the touch condition contacted the object with zero velocity for both orientations in agreement with task demands. In the lift condition, it may not be desirable to contact the target with zero velocity. Instead, we can use the momentum of our hand to help carry the object to the desired location. Therefore, it is predicted that subjects will contact the object with a positive thumb x-velocity for the horizontal condition, and a positive thumb and finger z-velocity for the vertical condition. All subjects showed finger velocities that match these predictions, and the horizontal and vertical data for one subject is shown in figure 4.3.

From this section it is clear that contact conditions changed in the expected directions across tasks and orientations. The next section will outline evidence from the horizontal condition that suggests the empirical lift data are in agreement with the loss function detailed above.

4.8 Are Empirical Contact Locations Optimal?

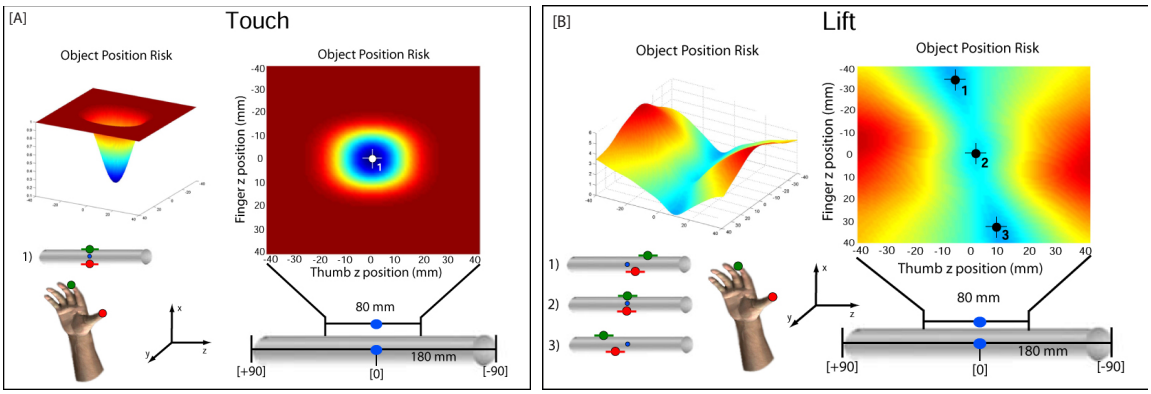

Figure 4.4: Risk landscapes for the horizontal touch and lift conditions. Index finger (green dots) z-positions are plotted against thumb (red dots) z-positions. The blue dot represents the object's center of mass. Only the middle 80 cm of the cylinder are shown. **[A]** *Touch risk landscape.* Heaviside function was used as loss function as it was assumed that the task required subjects to place any part of their finger and thumb (both assumed to be 1 cm) on the line. **[B]** *Lift risk landscape.* Loss function detailed above was used to develop landscape. Notice that there are a set of finger arrangements that produce accurate object motion.

Figure 4.4 illustrates the object position risk landscape for a horizontally oriented object as a function of the task. Blue areas show the low cost finger arrangements that overlap the contact line in the touch condition and produce good object motion in the lift condition. Figure 4.4 shows that there are essentially three distinct finger arrangements that produce good object motion. Conceptually, the top and bottom arrangements (Figure 4.4B(1,3)) are such that if the thumb falls to the right of the object's center of gravity, then the finger should be more right, and vice versa. Given the setup (i.e. right-handed reaching and start point right of center of mass), the (Figure 4.4B(1)) finger arrangements are more likely than those in (Figure 4.4B(3)). The middle arrangement (Figure 4.4B(2)) takes-on a more central approach where the fingers on parallel with each other. However, in order to achieve accurate object motion, it requires the fingers to be very near the center of mass. The area of low risk object motion for a central strategy is much smaller than those for a staggered strategy. Therefore, when making a right handed reach, the first finger arrangement may be the best of the possible low-risk arrangements. Finally, it should be noted that this risk function results solely from the physics of the task and motor noise estimates and has no free parameters.

Figure 4.5 shows the object position risk landscape with the four subject's data superimposed on the landscape. In the touch condition, it's apparent that subjects were

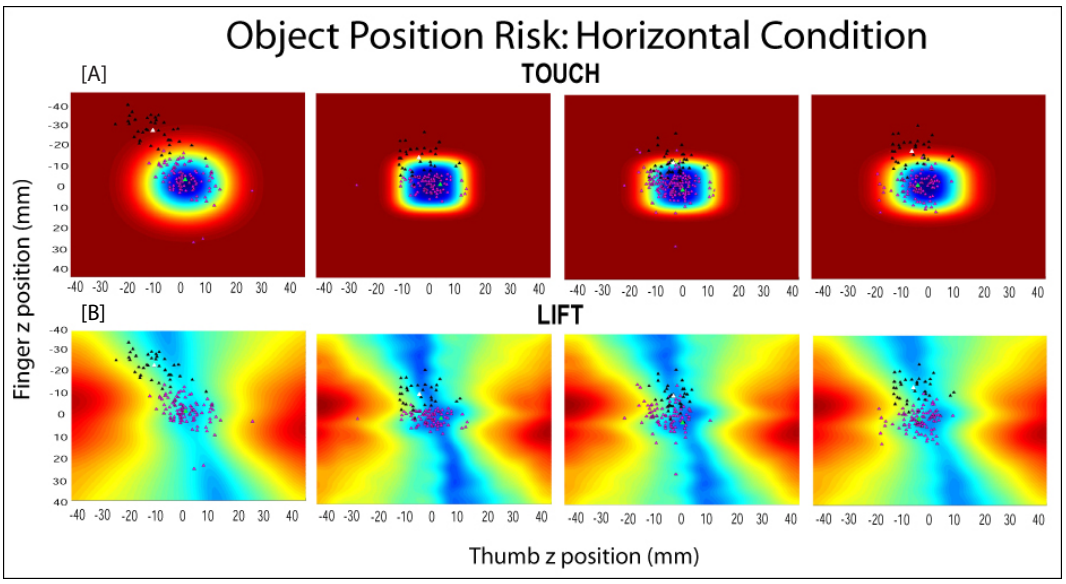

Figure 4.5: Risk landscapes plotted under data from four subjects. Purple triangles represent finger positions from an individual touch trial, and the green triangle represents the mean finger position across all of the touch trials. Black triangles are the finger locations from an individual lift trial, and the white triangle is the mean finger position across the lift trials. [**A**] Risk landscapes for touch data. It is apparent that the touch data falls within low risk areas for the touch landscape, but the lift data does not fall within the low risk areas for this landscape. [**B**] Risk landscape for the lift data. From the lift data, it seems that people are placing their fingers in positions that minimize the object position risk. This suggests that people are planning for the object's motion and placing their fingers in locations that produce accurate object motion.

doing as the task demanded - reach to the line located at the center of gravity. There is little variability between subjects because there was only a small range of positions that satisfied the task (Figure 4.5A). Figure 4.5A also demonstrates that the lift data does not fall within areas of low risk for the touch condition. This is important because it is plausible that subjects are adhering to a generic control strategy that works in several different situations (e.g., reach to the center of mass with a parallel finger arrangement). However, this data suggests that people are not employing a generic control strategy.

In the lift condition (Figure 4.5B), finger arrangements varied slightly across subjects. However, they were not arbitrarily placing their fingers on the object as every subject contacted the object with a staggered finger arrangement located on an area with low object-motion risk. Notice that the areas of low object motion risk are extremely small - for the average finger position, there are $\leq 5mm$ of low cost thumb positions. Since the landscape puts a cost on the object's motion, it can be assumed that people

were planning for the object's motion and placing their fingers in positions to achieve the desired motion, with minimum force and torque.

4.9 Discussion

The empirical results show that people use a parallel finger arrangement and contact the cylinder with zero velocity in the touch condition, whereas they use a staggered finger arrangement and contact the cylinder with positive x-velocity, in the horizontal lift condition (Figure 4.3). Moreover, it was demonstrated that people use different finger arrangements across cylinder orientations in the lift condition. These differences are not observed in the touch condition, suggesting that changes in finger contact locations and velocities are the result of the differential effect of gravity on the dynamics of lifting the cylinder. These results support the idea that contact points are selected to optimize a loss function on object motion. More generally, these results suggest that people understand the physics involved with lifting an object and that reach planning is hierarchical with object motion determining contact conditions and trajectory controls.

This theoretical framework provides an explanation for previous empirical findings demonstrating that hand shaping [3] and contact locations vary according to the task demands [21] and the predictability of the object's center of mass [74]. More specifically, previous research [3] found that the hand displayed gradual preshaping to the object when the task required lifting, but not for other conditions. This agrees with the proposed loss function since the effects of gravity must be accounted for when performing a lift, but not when simply reaching to the object. Unlike previous research, the proposed loss function can *predict* the changes in contact locations (and velocities) people display when lifting an object.

Moreover, previous research has demonstrated that the contact locations people select on a vertically oriented cylinder are inversely proportional to the height in which they are required to place the cylinder on a shelf [21]. The authors propose an explanation for this finding by stating that contact locations are selected that allow for kinematic comfort when transporting the object. Although the current model does not account for kinematic terms, this is a likely reason why finger arrangement 1 (Figure 4.4B) was preferred to finger arrangement 3 (i.e., since subjects reached with their right hand, finger arrangement 3 would require the hand to "roll-over" itself in an unusual

position). Although taking this term into consideration would result in a reduced risk landscape, it does not explain the results produced in this and other (e.g., [74, 3]) experiments. Indeed, it could be that subjects are using a loss function similar to the one proposed in this chapter, and pruning-off possible finger positions by other (i.e., kinematic) loss terms.

Other research [74] has shown that peoples' contact locations change when the center of mass was predictable, but did not change when it could not be predicted. These results are consonant with the current model since people must be able to predict the physical dynamics of object manipulation in order to select contact conditions that achieve an object motion with minimum force. Such a forward model is only accurate if the important physical properties of the object can be accurately estimated. The major benefit of the current model is that it *directly predicts* the changes in contact locations people will display across tasks and objects. Moreover, it is also capable of predicting the changes in velocity displayed when people lifted objects at varying orientations (Figure 4.3).

Overall, this chapter proposed a natural loss function that was able to predict changes in contact locations observed in this study, in addition to data from previous research. The loss function is based on the physics of object manipulation and assumes that people choose contact locations that produce accurate object motion with minimum force. Future studies could investigate how these forward models are learned and the role of sensory information in this process as many of the physical properties needed cannot be directly observed. The next chapter will proved a summary of this thesis, in addition to providing domains of research that may benefit from a decision theoretic approach.

Chapter 5

General Discussion

This paper argued that SDT has utility for understanding natural human behaviors. However, in order for the brain to actively estimate the expected risk for a particular movement, it necessitates that it understands both its uncertainty and the loss function that is associated with the task. Therefore, evidence that people appear to be grasping objects in a manner consistent with these ideas is summarized below.

5.1 Summary of Compensation for Visual Uncertainty

Chapter 2 demonstrated that people are estimating and adjusting for their visual uncertainty when making grasping movements. People adjusted their MGA according to the amount of visual uncertainty in the task. They also appeared to adjust their grasp by extending their finger along a principal axis of change. These results suggest that the brain compensates for target location uncertainty by scaling MGA along a primary axis. Moreover, it appears that people are using a more conservative grasping strategy in the presence of uncertainty to minimize the probability of colliding with or missing the target.

Although these results demonstrate that increases in uncertainty led to more conservative grasping behavior, in some cases it may beneficial to perform additional "information seeking" behaviors to reduce uncertainty before acting. For example, if someone were viewing a television program and wanted to grasp their beverage, they could either reach to the drink without removing fixation from the screen, or saccade to the target location before making the reach. In this example, the first scenario is

similar to the task being performed in our experiment (i.e., reaching to peripherally viewed objects) and should lead to more conservative reaching behavior. However, in the second scenario the observer makes a saccade to gain additional information before performing their reach. As a result, the observed grasping behavior should be similar to those of a foveal reach (i.e., smaller MGA).

This raises the issue of when it's optimal to act under high levels of uncertainty and when is it optimal to seek additional information to reduce uncertainty. Indeed, it may be that the cost of making particular errors plays a crucial role in this process. For example, when driving a vehicle in heavy traffic, it is less costly to collide with or miss your beverage than to collide with another vehicle. Therefore, this would predict that people should maintain fixation on the road while reaching to their peripherally viewed drink. Conversely, if the driver performed the same task when traffic is light, they may perform information seeking behavior by saccading to the beverage before making their reach. Future research could explicitly test these ideas to assess the trade-off between reaching under uncertainty and information seeking strategies.

5.2 Summary of Compensation for CTU

In addition to MGA scaling with visual uncertainty, Chapter 3 showed that people account for their coordinate transformation uncertainty (CTU) when performing a grasping task. CTU results from inaccurate knowledge about the relative location of body-segments, and leads to the degradation of estimates as they are passed between coordinate frames. The results of this effort suggest that the brain is aware of CTU and that it incorporates this knowledge into reach plans. Moreover, these results suggest that CTU should be considered in cases when multimodal cue combination is being performed and body-position is not held fixed.

Visual uncertainty (Chapter 2) and CTU (Chapter 3) are both examples of uncertainty that is the result of *internal* limitations of the system (i.e., visual uncertainty results from the layout of photoreceptors on the retina; and CTU results from inaccurate sensory knowledge about the relationship between body segments). However, it is also possible that our brain should adjust for *external* sources of uncertainty when making a grasp. For example, when seated at a desk the arrangement of items on the desk (e.g., coffee cup, phone) is relatively consistent, although there is usually some variability between where an item is located one day compared to the next. If

someone wished to grasp an object, the brain could use knowledge about the distribution of possible object locations to plan a saccade that maximizes the probability of acquiring reliable visual information to guide the reach. Future research of mine will investigate this idea by explicitly manipulating the uncertainty in the target's location and assessing its impact on subsequent eye and hand movements. Other sources of external uncertainty that could be further investigated include uncertainty in the target's identity [104], uncertainty in the target's shape, and uncertainty in the object's friction. These efforts would provide a comprehensive picture for how both internal and external sources of uncertainty are used by the brain when performing actions.

The experiments described in this thesis have demonstrated that uncertainty resulting from visual information and CTU are estimated when making a grasp. However, in order to compute the expected risk, it also necessitates that the loss function for the task be known. An effort to reverse-engineer a loss function for grasping was proposed in Chapter 4 and is summarized below.

5.3 Summary of *Natural* Loss Functions

Chapter 4 of this paper outlined and formalized a *natural* loss function for human reach and grasp. The chapter emphasized the importance of this loss function for predicting contact conditions across tasks. An experimental test of the model showed that people are placing their fingers in locations that produce accurate object motion with minimum force. These results suggest that the brain understands the physics involved with object manipulation and that people may be planning for the intended motion of the object.

In addition to understanding the loss function that predicts natural grasping behavior, it is also desirable to understand what strategy the brain uses to guide the coordination between the eyes and hands. When performing a natural task, such as dialing an unfamiliar phone number, our eyes and hands work in concert to both gain the information needed about the phone number and to guide the hand to the appropriate location on the phone pad. Despite the fact that natural tasks require both eye and hand movements, most of the existing research has studied eye and hand movements in isolation. Future research of mine will investigate the principals that drive this naturally occurring synergy under differential conditions of reward/loss.

Overall, this paper demonstrated that the brain adjusts for visual uncertainty and CTU when performing grasping movements. Moreover, a natural loss function was proposed that can account for variations in the contact conditions people use when making a grasp. Together, these results imply that SDT can be used to understand natural movements, such as reach and grasp. However, SDT could also be useful for modeling many different perception-action cycles, illustrated by the next section.

5.4 Applications to Other Perception-Action Cycles

5.4.1 Human Navigation

One example of a perception-action cycle that could benefit from a SDT approach is human navigation. Advances in robotics have developed useful algorithms that select landmarks based on the error they reduce (e.g, [108]). It is argued here that humans may be doing something similar. Perhaps the role of landmarks is to reduce the uncertainty in position and action space. It would be interesting to develop an Bayesian algorithm that optimally selects landmarks on the basis of the uncertainty they reduce. The algorithm could then be compared to human performance to assess if these same landmarks are being selected by humans. If the algorithm and human behavior match, it would suggest that people are selecting landmarks that optimally reduce position and action uncertainty.

Another useful tool developed in the AI community is called *inverse reinforcement learning* (IRL) [85]. Essentially, IRL derives the loss functions people are using from the actions they made at particular states. The authors suggest that this technique is appropriate when there is no good reason to assume a loss function for the task (i.e., much of psychology). Using this technique in navigation may provide insight as to what loss functions people are using and how they affect subsequent navigation performance. The above two approaches allow for the exploration into the role of uncertainty and loss functions in human navigation. This is one example of a perception-action cycle to which SDT can be applied, while another that investigates the utility of SDT for intelligent vehicle design is provided below.

5.4.2 Intelligent Vehicles

Engineers have been striving to introduce new technologies into vehicles since the inception of the automobile. For example, a recent technology has been proposed that improves the contrast of the environment during nighttime driving. Although the intuition may be that this technology will improve driving performance, "night vision" for vehicles should be approached with caution. There is a phenomenon in human performance that is called *behavioral homeostasis* (e.g., [40]). In this example it would suggest that introducing a piece of technology that improves safety (i.e., night vision), will subsequently lead to more hazardous behavior (i.e., increase speed).

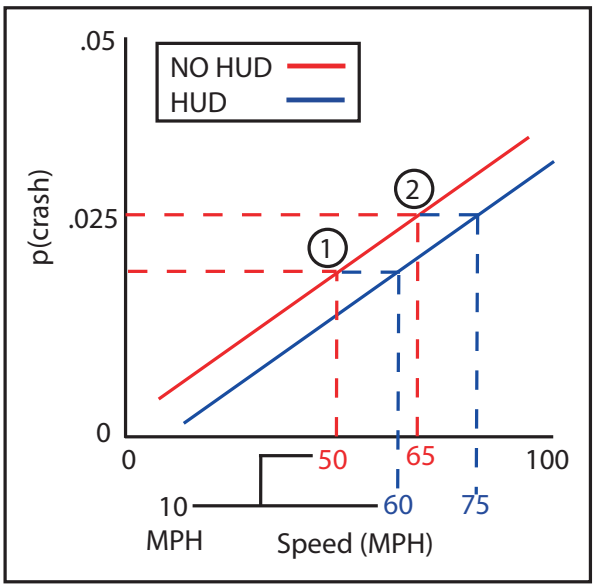

Figure 5.1: Figure illustrates how night vision HUDs may lead to increased speed when introduced into vehicles. The diagram assumes that people are willing to accept some threshold of risk (i.e., probability of crashing) when driving. Please note that some people (2) are willing to accept more of a risk than others (1). Regardless of the level of risk people are willing to accept, introducing night vision into vehicles (blue line) will decrease the probability of crashing by reducing the uncertainty in the task, compared to when there is no night vision (red line). The result is that people can now increase their speed (e.g., 10 mph) to maintain their desired level of uncertainty. This increase in speed could have an impact on several accident factors such as average dollars per crash, average severity of injury or the probability life will be lost. Therefore, considerations should be taken before introducing this kind of technology into vehicles.

In the context of SDT, this implies that the reduction in uncertainty lead to an increase in speed. A loss function attempts to maintain a constant (low) probability of crashing would account for these behavioral results (Figure 5.1). Suppose people choose their speed by accepting some low probability of crashing. Introducing night

vision into vehicles lowers the probability they will crash. Therefore, people can increase their speed to levels above those when night vision is absent.

If the loss function and uncertainty in the task could be derived, it would be extremely useful to many groups. For example, engineers could predict the reduction in uncertainty introducing new technologies would provide. Similarly, government officials could assess the predicted increase in speed that would result from the technology. They could cross-reference that information with accident data to see how the speed increase could influence such things as average damage per crash (in dollars), average injury rate and severity, and average death rate. Government officials could then use that information to determine whether or not the expected monetary gain for introducing the technology is worth the predicted accident risk.

5.5 Conclusion

This paper attempted to demonstrate that SDT lends itself well to human perception and action. From the work described in the paper and the above examples, hopefully the utility of using SDT for human perception-action cycles is apparent. Since the field is only in its infant stage, it will be exciting to see how this approach can advance our understanding of natural human behavior.

Appendix A

Data Modeling

We developed a probabilistic model that simulates how target location information is combined and stored in order to compare to the grasping data presented here, as well as the pointing data presented in [71]. In this section the computations are structured to clearly define the dependence of reaches on the most recent sensory information and to illustrate the effects of coordinate transformation uncertainty.

The basic computations and assumptions of the model are as follows. We assume remembered target location can be represented by a probability density function. For example, the memory distribution for eye-centered coordinates is represented by $p_{mem}(x|m_{k-1})$, where x is the target location in eye-centered coordinates and m_{k-1} summarizes visual and haptic experience up to the $k-1^{th}$ trial. At each trial before the k^{th} reach, perceived target location is computed by combining recent sensory data with target information in memory. The perceived target location is represented by a density function $p_x(x|v_k, h_{k-1}, e_{k-1}, m_{k-1})$, that makes explicit the dependence on the current trial's visual information v_k (when available) and the previous trial's haptic information h_{k-1} (which introduces dependence on eye position information e_{k-1}). Finally, it is assumed reach actions are based on perceived target location transformed to body-centered coordinates, represented by $p_y(y|v_k, h_{k-1}, e_{k-1}, m_{k-1}, e_k)$, and potentially corrupted by CTU due to errors in eye position sensing (for eye-centered storage).

First, expressions for perceived target location and the effects of CTU are derived. To derive explicit formula for perceived object location, Gaussian approximations

are used for all distributions, where $\mathcal{N}(x;\mu,\sigma^2)$ denotes a Gaussian density on x with mean μ and variance σ^2. However, all distributions in the model are realistic insofar as the parameters are known. Specific assumptions about the form of these distributions are described in Chapter 3.

A.1 Perceived Target Location in Eye-Centered Coordinates

To form the distribution on perceived target location, visual information from the current trial $p(v_k|x)$ and haptic information from the previous trial $p(h_{k-1}|x,e_{k-1})$ are combined with memory using probabilistic inference. This results in the following expression, when the target is visible:

$$p_{x,vis}(x|v_k,h_{k-1},e_{k-1},m_{k-1}) = \frac{p_{mem}(x|m_{k-1})p(v_k|x)p(h_{k-1}|x,e_{k-1})}{p(v_k,h_{k-1}|e_{k-1},m_{k-1})} \quad \text{(A.1)}$$

and the following expression when the target is not visible:

$$p_{x,occ}(x|h_{k-1},e_{k-1},m_{k-1}) = \frac{p_{mem}(x|m_{k-1})p(h_{k-1}|x,e_{k-1})}{p(h_{k-1}|e_{k-1},m_{k-1})} \quad \text{(A.2)}$$

where v_k denotes the retinal location of the target, h_{k-1} is the target location information conveyed by touching the object on the previous trial, e_{k-1} are eye position signals used to bring haptic information into eye-centered coordinates (discussed below), and the probabilities in the denominators are normalization constants that do not affect inference.

We model visual and haptic distributions as Gaussian:

$$p(v_k|x) = \mathcal{N}(v_k;x,\sigma_v^2) \quad \text{(A.3)}$$
$$p(h_{k-1}|x,e_{k-1}) = \mathcal{N}(h_{k-1};h_x,\sigma_{h,x}^2) \quad \text{(A.4)}$$
$$p_{mem}(x|m_{k-1}) = \mathcal{N}(x;\mu_{eye,k-1},\ \sigma_{eye,k-1}^2) \quad \text{(A.5)}$$

where h_x and $\sigma_{h,x}^2$ are the mean and variance of the haptic distribution remapped into eye-centered coordinates. Remapping has the effect of introducing bias and additional

uncertainty due to errors in eye position sensing:

$$h_x = 40^\circ - w_e\, e_{k-1} \qquad \sigma^2_{h,x} = \sigma^2_h + \sigma^2_{pe,k-1} \tag{A.6}$$

The terms e_{k-1} and $\sigma^2_{pe,k-1}$ are the mean and variance of a distribution describing eye position inference, and w_e is a gain factor that encodes a bias in eye position estimation toward forward view. Both of these are described in the next section below.

The posterior distributions for target location with and without visual information are Gaussian:

$$p_{x,vis}(x|v_k, h_{k-1}, e_{k-1}, m_{k-1}) = \mathcal{N}(x; \mu_{x,vis}, \sigma^2_{x,vis})$$

$$p_{x,occ}(x|h_{k-1}, e_{k-1}, m_{k-1}) = \mathcal{N}(x; \mu_{x,occ}, \sigma^2_{x,occ})$$

with formulas for mean and variance given by:

$$\mu_{x,vis} = \frac{1}{\sigma^2_{x,vis}} \frac{\mu_{eye,k-1}}{\sigma^2_{eye,k-1}} + \frac{v}{\sigma^2_v} + \frac{h_x}{\sigma^2_{h,x}} \tag{A.7}$$

$$\sigma^2_{x,vis} = \frac{\sigma^2_{eye,k-1}\sigma^2_v\sigma^2_{h,x}}{\sigma^2_{eye,k-1}(\sigma^2_v + \sigma^2_{h,x}) + \sigma^2_v\sigma^2_{h,x}} \tag{A.8}$$

$$\mu_{x,occ} = \frac{1}{\sigma^2_{x,occ}} \frac{\mu_{eye,k-1}}{\sigma^2_{eye,k-1}} + \frac{h_x}{\sigma^2_{h,x}} \tag{A.9}$$

$$\sigma^2_{x,occ} = \frac{\sigma^2_{eye,k-1}\sigma^2_{h,x}}{\sigma^2_{eye,k-1}\sigma^2_{h,x} + \sigma^2_{h,x}} \tag{A.10}$$

Models for visual and haptic information are presented in the Modeling Section.

A.2 Coordinate Transformations and Eye Position Sensing

Modeling coordinate transformations using probability theory involves defining a joint distribution that relates the position of the target object in both coordinate frames using the parameters of the transformation. Although this is complicated in general, target location was represented in angular coordinates so that the transformation

between head and eye coordinates becomes approximately linear[1]: $y = x + r$, where r denotes the azimuthal (angular) coordinate of the eye with respect to forward view. Probabilistically the transformation is a transition kernel expressing the relation between y and x for every value of eye position r: $p(y|x, r)$. A transition kernel for the linear transformation above can be represented by a Dirac delta function: $p(y|x, r) = \delta(y - (x + r))$. Information about eye position r is represented by $p(r|e)$, where e is a vector that summarizes efference copy of motor commands [72], proprioception [107], and the retinal location of the fixation point [84]. Coordinate transformation uncertainty arises from marginalizing the transition

$$p(y|x, e_k) = \int_r p(y|x, r)p(r|e_k)dr \tag{A.11}$$

The effect of CTU is to introduce uncertainty in the relationship between x and y. In particular, assuming $p(r|e_k) = \mathcal{N}(r; w_e\, e_k, \sigma^2_{pe,k})$ is approximately Gaussian (modeled below), the above integral results in:

$$p(y|x, e_k) = \mathcal{N}(y; x + w_e\, e_k, \sigma^2_{pe,k}) \tag{A.12}$$

Using D to symbolize an arbitrary set of data, coordinate transformations from eye to head-centered coordinates (and vice versa) integrate across the transition kernel and incorporate available eye position signals:

$$\begin{aligned} \textbf{Eye} \rightarrow \textbf{Head} \qquad p(y|D, e_k) &= \int_x p(x|D)p(y|x, e_k)dx \\ \text{and} \\ \textbf{Head} \rightarrow \textbf{Eye} \qquad p(x|D, e_k) &= \frac{1}{p(D|e_k)} \int_y p(D|y)p(y|x, e_k)dx \end{aligned} \tag{A.13}$$

where $p(D|e_k)$ in the second expression is a normalization constant. These transformations arise when information in eye-centered coordinates is transformed to body-

[1]The approximation stems from the fact that the eye is offset from the center of the head and that the eye does not rotate around its center. However, the offset is constant and hence does not affect the results, and the effect of off-axis eye rotation is neglige compared to the rotation.

centered before a reach, and when haptic information is brought into eye-centered coordinates. Because Equations A.14 are convolution integrals given the form of the density in Equation A.12, the effect of CTU is to shift the mean of the transformed distribution by $w_e\, e_k$ and to its variance add $\sigma^2_{e,k}$. For haptic information brought into eye-centered coordinates, the effects of CTU previously have been shown in Equation A.6. Using Equation A.14 to modify the expressions in Equations A.8 & A.10 for perceived target location in eye-centered coordinates, the distributions for perceived target location in head-centered coordinates are given by:

$$\begin{aligned} p_{y,vis}(y|v_k, h_{k-1}, e_{k-1}, m_{k-1}, e_k) &= \int_x p_{x,vis}(x|v_k, h_{k-1}, e_{k-1}, m_{k-1}) p(y|x, e_k) dx \\ &= \mathcal{N}(y; \mu_{x,vis} + w_e\, e, \sigma^2_{x,vis} + \sigma^2_{pe,k}) \end{aligned} \tag{A.14}$$

with vision and without vision by:

$$\begin{aligned} p_{y,occ}(y|v_k, h_{k-1}, e_{k-1}, m_{k-1}, e_k) &= \int_x p_{x,occ}(x|h_{k-1}, e_{k-1}, m_{k-1}) p(y|x, e_k) dx \\ &= \mathcal{N}(y; \mu_{x,occ} + w_e\, e, \sigma^2_{x,occ} + \sigma^2_{pe,k}) \end{aligned} \tag{A.15}$$

Next eye position sensing is modeled, after which the model components are pulled into the form used to generate data predictions.

A.3 Modeling Eye Position Sensing

We assume that the distribution encoding eye position $p(r|e)$ is derived from e and prior information $p(r)$ according to Bayes' formula:

$$p(r|e) = \frac{p(r)p(e|r)}{\int_r p(r)p(e|r)dr}, \tag{A.16}$$

where $p(e|r) = \mathcal{N}(e; r, \sigma^2_e)$, and $\sigma^2_e = (.05 * e + 0.05)^{2\circ}$ [116, 84], models the eye position signals as unbiased but with signal-dependent noise that varies as a function

of eye position . The prior on eye position $p(r)$ reflects the fact that the eyes are usually forward with respect to the head [106]. For the simulations, the Gaussian eye position prior was fixed to be $\mathcal{N}(r; 0, \sigma_r^2 = 100^\circ)$. The posterior distribution has the form:

$$p(r|e) = \mathcal{N}(r; \mu_e, \sigma_{pe}^2)$$

where $w_e = \frac{\sigma_r^2}{\sigma_e^2+\sigma_r^2}$, $\sigma_{pe}^2 = w_e\sigma_e^2$ and $\mu_e = w_e e$.

A.4 Perceived Object Location in Head-Centered Coordinates

Expressions for the mean and variance of the perceived object distribution in head-centered coordinates are generated below. Predictions for pointing and grasping based on these results are found in the Results.

Let $\left(\mu_{eye}, \sigma_{eye}^2\right)$ denote the average mean and variance of the memory distribution in eye-centered coordinates.

Combining Equations A.14 & A.15 with Equations A.8, and A.10 results in the following expressions:

Target Visible, Eye-centered coordinates

$$\begin{aligned}
\mu_{y,vis,eye} &= \sigma_{y,vis,eye}^2 \left(\frac{\mu_{eye}}{\sigma_{eye}^2} + \frac{v}{\sigma_v^2} + \frac{h - w_e\, e_{k-1}}{\sigma_h^2 + \sigma_{pe,k-1}^2} \right) + w_e e_k \\
\sigma_{y,vis,eye}^2 &= \frac{\sigma_{eye}^2(\sigma_h^2 + \sigma_{pe,k-1}^2)\sigma_v^2}{\sigma_{eye}^2(\sigma_v^2 + \sigma_h^2 + \sigma_{pe,k-1}^2) + \sigma_v^2(\sigma_h^2 + \sigma_{pe,k-1}^2)} + \sigma_{e,k}^2 \qquad \text{(A.17)}
\end{aligned}$$

Target Occluded, Eye-centered coordinates

$$\begin{aligned}
\mu_{y,occ,eye} &= \sigma^2_{y,occ,eye}\left(\frac{\mu_{eye}}{\sigma^2_{eye}} + \frac{h - w_e\, e_{k-1}}{\sigma^2_h + \sigma^2_{pe,k-1}}\right) + w_e e_k \\
\sigma^2_{y,occ,eye} &= \frac{\sigma^2_{eye}(\sigma^2_h + \sigma^2_{pe,k-1})}{\sigma^2_{eye} + \sigma^2_h + \sigma^2_{pe,k-1}} + \sigma^2_{pe,k}
\end{aligned} \tag{A.18}$$

A.5 Storing Information in Head-Centered Coordinates

To model storage in head-centered coordinates, all the computations above remain the same, except the memory update is performed in head-centered coordinates y, with memory distribution $p(y|m_{k-1})$. When visual information is available the head-centered memory distribution is updated as follows:

Target Visible, Storage in head-centered coordinates

$$\begin{aligned}
\mu_{y,vis,head} &= \sigma^2_{y,vis,head}\left(\frac{\mu_{head}}{\sigma^2_{head}} + \frac{v + w_e\, e_k}{\sigma^2_v + \sigma^2_{pe,k}} + \frac{h}{\sigma^2_h}\right) \\
\sigma^2_{y,vis,head} &= \frac{\sigma^2_{head}(\sigma^2_v + \sigma^2_{pe,k})\sigma^2_h}{\sigma^2_{head}(\sigma^2_v + \sigma^2_h + \sigma^2_{e,k-1}) + \sigma^2_h(\sigma^2_v + \sigma^2_{pe,k-1})}
\end{aligned} \tag{A.19}$$

Target Occluded, Storage in head-centered coordinates

$$\begin{aligned}
\mu_{y,occ,head} &= \sigma^2_{y,occ,head}\left(\frac{\mu_{head}}{\sigma^2_{head}} + \frac{h}{\sigma^2_h}\right) \\
\sigma^2_{y,occ,head} &= \frac{\sigma^2_{head}\sigma^2_h}{\sigma^2_{head} + \sigma^2_h}
\end{aligned} \tag{A.20}$$

Without vision, integrating across v and x removes all dependence between y and r and hence there is no eye position dependence for memory distributions stored in head-centered coordinates.

Bibliography

[1] Admiraal M.A., Keijsers, N.L.W., and Gielen, C.C.A.M. (2003). Interaction between gaze and pointing toward remembered visual targets. *J Neurophysiol* 90, 21362148.

[2] Admiraal, M.A., Keijsers, N.L.W. and Gielen, C.C.A.M. (2004). Gaze Affects Pointing Toward Remembered Visual Targets After a Self-Initiated Step. *J Neurophysiol*, 92, 2380-2393.

[3] Ansuini, C., Santello, M., Massaccesi, S., Castiello, U. (2006) Effects of end-goal on hand shaping. *J Neurophysiol*, 95, 2456 2465.

[4] Anderson, C. H. (1994). Basic elements of biological computational systems. *International Journal of Modern Physics*, 5, pp. 135-137.

[5] Atkins, J.E., Fiser, J., and Jacobs, R.A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, 41, 449-61.

[6] Avillac, M., Denve, S., Olivier, E., Pouget, A. and Duhamel, J.R. (2005). Reference frames for representing the location of visual and tactile stimuli in the parietal cortex. *Nature Neuroscience*. 8(7). 941-949.

[7] Baker, J.T., Harper, T.M., and Snyder, L.H. (2003). Spatial Memory Following Shifts of Gaze. I. Saccades to Memorized World- Fixed and Gaze-Fixed Targets. *J Neurophysiology*, 89, 25642576.

[8] Barlow, H.B. (1962). A method of determining the overall quantum effciency of visual discriminations. *J. Physiol. (Lond.)*, 160, pp. 155-168.

[9] Batista A.P., Buneo C.A., Snyder L.H., and Andersen R.A. (1999). Reach Plans in Eye-Centered Coordinates, *Science*, 285, 257-260.

[10] Battaglia, P.W., Jacobs, R.A, and Aslin, R.N. (2003) Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, 20, pp. 1391-1397.

[11] Berger, J., (1985). Statistical Decision Theory and Bayesian Analysis, 2nd edition, Springer Verlag, New York.

[12] Berthier, N.E., Clifton, R.K., Gullipalli, V., McCall, D., and Robin, D. (1996). Visual information and object size in the control of reaching. *Journal of Motor Behavior*, 28, 187-197.

[13] Bicchi, A. and Kumar, V. (2000). Robotic Grasping and Contact: A Review. *In Proc. IEEE Int. Conf. on Robotics and Automation*, 1, pp. 348-353.

[14] Blakemore, S.J., Wolpert, D.M. and Frith, C.D. (2000). Why can't you tickle yourself? NeuroReport, 11, p. 11-15.

[15] Blouin, J., Amade, N., Vercher, J.L., Teasdale, N., and Gauthier GM. (2002). Visual signals contribute to the coding of gaze direction. *Exp Brain Res*, 144, 281-292.

[16] Buneo, C.A., Jarvis, M.R., Batista, A.P., and Andersen, R.A. (2002). Direct visuomotor transformations for reaching, *Nature*, 416, 632-636.

[17] Burbeck, C.A. (1987) Position and spatial frequency in large-scale localization judgments. *Vision Res*, 27, 417428.

[18] Burbeck, C.A., Yap, Y.L. (1990) Two mechanisms for localization? Evidence for separation-dependent and separation-independent processing of position information. *Vision Res*, 30, 739750.

[19] Cheng, S. and Sabes, P. N. (2006). Modeling Sensorimotor Learning with Dynamical Systems. *Neural Computation*, 18, 760-793.

[20] Chieffi, S., Gentilucci, M. (1993). Coordination between the transport and the grasp components during prehension movements. *Experimental Brain Research*, 94, 471-477.

[21] Cohen R.G., Rosenbaum D.A. (2004) Where grasps are made reveals how grasps are planned: generation and recall of motor plans. *Experimental Brain Research*, 157, 486495.

[22] Connolly J.D., and Goodale, M. A. (1999). The role of visual feedback of hand position in the control of manual
prehension. *Experimental Brain Research*, 125, 281-286.

[23] Cuijpers, R.H., Smeets, J.B., Brenner, E. (2004). On the relation between object shape and grasping kinematics. *J Neurophysiol.*, 91, 2598-2606.

[24] Crawford, J.D., Medendorp, W.P., and Marotta, J.J. (2004). Spatial transformations for eye-hand coordination. *Journal of Neurophysiology*, 92, 10-19.

[25] Daw, N.D., (2003). Reinforcement learning models of the dopamine system and their behavioral implications. *PhD. Thesis, CMU.*

[26] Dickhaut, J., McCabe, K., Nagode, J.C., Rustichini, A., Smith, K., and Pardo, J.V. (2003). The impact of the certainty context on the process of choice. *Proceedings of the National Academy of Science*, 100, pp.3536-3541.

[27] Dreiseitl, S. (2004). Online lecture notes in statistical pattern recognition.

[28] Duhamel, J.R., Colby, C.L. and Goldberg, M.E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255, pp. 9092.

[29] Elliott, D., and Madalena, J. (1987). The influence of premovement visual information on manual aiming. *Q J Exp Psychol A*, 39, 541-559.

[30] Ellsberg, D. (1961). Risk, ambiguity and the savage axioms, *Quarterly Journal of Economics*, 75, pp. 643-669.

[31] Ernst, M.O. and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429-33.

[32] Ernst, M.O., Banks, M.S., and Bulthoff, H.H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3, pp. 69-73.

[33] Fiorillo, C.D., Philippe, N.T., and Shultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299, pp.1898-1901.

[34] Flash, T., and Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5, pp.1688-1703.

[35] Foldiak P. (1993). The ideal homunculus: statistical inference from neural population responses. In Computation and Neural Systems, ed. F Eeckman, J Bower, pp. 5560. Norwell, MA: Kluwer Acad. Publ.

[36] Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., and Massey, J. T. (1989). Mental Rotation of the Neuronal Population Vector. *Science*, 243, pp. 234-236.

[37] Ghahramani, Z., Wolpert, D.M. and Jordan, M.I. (1997). Computational Models of Sensorimotor Integration In P. G. Morasso and V. Sanguineti (eds.), *Self-Organization, Computational Maps and Motor Control*, Elsevier Press, pp. 117-147.

[38] Gold, M. R., Siegel, J. E., Russell, L. B. and Weinstein, M.C.(eds) (1996). Cost-effectiveness in health and medicine, Oxford University Press, Oxford.

[39] Goldberg, M.E., Bruce, C.J. (1990). Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. *J Neurophysiology*, 64, 489-508.

[40] Gonzalez, J.J., and Sawicka, A. (2003). Modeling instrumental conditioning - The behavioral regulation approach. *IEEE. Proceedings of the Hawaii International Conference on System Sciences*, January 6-9, 2003, Big Island, Hawaii.

[41] Graziano, M.S., Yap, G.S., Gross, C.G. (1994). Coding of visual space by premotor neurons, *Science*, 266, 1054-1057.

[42] Graziano, M., Hu, X. and Gross, C. (1997). Coding the locations of objects in the dark. *Science*, 277, pp. 239241.

[43] Graziano, M.S., Hu, X., Gross, C.G (1997). Visuospatial properties of ventral premotor cortex, *J. Neurophysiology*, 77, 2268-2292.

[44] Hallet, P.E., Lightstone, A.D. (1976). Saccadic eye movements to flashed targets. *Vision Res* 16, 107-114.

[45] Hamilton, A.F., and Wolpert, D.M. (2002). Controlling the action of statistics: Obstacle avoidance. *Journal of Neurophysiology*, 87, 2434-2440.

[46] Harris, C.M., and Wolpert, D.M. (1998). Signal dependent noise determines motor planning . *Nature*, 394, pp.780-784.

[47] Hawkins, DM (1994). The Feasible Solution Algorithm for Least Trimmed Squares Regression. *Computational Statistics and Data Analysis*, 17, 185-196.

[48] Henriques, D.Y.P., Klier, M.J., Smith, M.A., Lowy, D., and Crawford, J.D. (1998). Gaze-Centered Remapping of Remembered Visual Space in an Open-Loop Pointing Task. *Journal of Neuroscience*, 18, 1583-1594.

[49] Herter, T.M., Guitton, D. (1998). Human head-free gaze saccades to targets flashed before gaze-pursuit are spatially accurate. *J Neurophysiology*, 80, 2785-2789.

[50] Israel, I., Berthoz, A. (1989). Contribution of the otoliths to the calculation of linear displacement. *J. Neurophysiology*, 62, 247-263.

[51] Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In: Long, J., Baddeley, A. (eds.), Attention and Performance IX (pp: 153-169). Erlbaum, Hillsdale, NJ.

[52] Jeannerod, M. (1984) The timing of natural prehension movements. *Journal of Motor Behavior*, 16, 235-254.

[53] Jensen, N. E. (1967). An introduction to Bernoullian utility theory: I Utility functions. *Swedish Journal of Economics*, 69, pp. 163-183.

[54] Jones, K., Hamilton, A. and Wolpert, D.M. (2002). The sources of signal dependent noise during isometric force production. *Journal of Neurophysiology*, 88, 1533-1544.

[55] Kaelbling, L.P., Cassandra, A.R., and Kurien, J.A. (1996). Acting Under Uncertainty: Discrete Bayesian Models for Mobile-Robot Navigation. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems.*

[56] Kanade, T. (1981). Recovery of the 3-D shape of an object from a single view. *Artificial Intelligence*, 17, pp. 409-460.

[57] Karn, K.S., Moller, P., and Hayhoe, M.M. (1997). Reference frames in saccadic targeting. *Exp Brain Res*, 115, 267-282.

[58] Kersten, D. and Schrater, P. R., Pattern Inference Theory: A Probabilistic Approach to Vision. To appear in R. Mausfeld, and D. Heyer (Ed.), Perception Theory: Conceptual Issues Chichester: John Wiley and Sons, Ltd.

[59] Kesten, H. (1958). Accelerated stochastic approximation. *Annals of Mathematical Statistics*, 29, 41-59.

[60] Knill, D.C. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Res*, 38, pp. 1683-711.

[61] Knill, D.C. (1998). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Res*, 38, pp. 1655-82.

[62] Knill, D.C., and Saunders, J.A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, pp.2539-2558.

[63] Kording, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427, pp.244-247.

[64] Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., and Caminiti, R. (1995). Representing spatial information for limb movement: role of area 5 in the monkey. *Cerebral Cortex*, 5, pp.391-409.

[65] Landy, M. S. and Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, 18, pp. 2307-2320.

[66] Landy, M.S., Maloney, L.T., Johnston, E.B., and Young, M.J. (1995). Measurement and modeling of depth cue combination. *Vision Research*, 35, 389-412.

[67] Legge, G.E., Klitz, T.S., and Tjan, B.S. (1997). Mr. Chips: an ideal-observer model of reading. *Psych. Review*, 104, pp. 524-53.

[68] Legge, G.E. (1999). Class notes on signal detection theory.

[69] Levi, D.M., and Klein, S.A. (1996). Limitations on position coding imposed by undersampling and univariance, *Vision Research*, 36, 2111-2120.

[70] Lewald, J. (1998). The effect of gaze eccentricity on perceived sound direction and its relation to visual localization. *Hearing Research*, 115, 206-216.

[71] Lewald, J., Ehrenstein, W.H. (2000). Visual and proprioceptive shifts in perceived egocentric direction induced by eye-position. *Vision Research*, 40,539-547.

[72] Lewis, R.F., Gaymard, B.M., and Tamargo, R.J. (1998). Efference Copy Provides the Eye Position Information Required for Visually Guided Reaching. *J Neurophysiol*, 80, 1605-1608.

[73] Liu, Z., Knill, D.C., and Kersten, D. (1995). Object Classi.cation for Human and Ideal Observers. *Vision Research*, 35, pp. 549-568.

[74] Lukos, J., Ansuini, C., and Santello, M. (2007). Choice of Contact Points during Multidigit Grasping: Effect of Predictability of Object Center of Mass Location, J Neuroscience, 27, 3894 3903.

[75] McNeil, B., Weichselbaum, R. and Pauker, S. (1981). Speech and survival: Tradeoffs between quality and quantity of life in laryngeal cancer. *The New England Journal of Medicine* 305, pp. 982-987.

[76] Mckenzie, A., Lisberger, S. G. (1986). Properties of signals that determine the amplitude and direction of saccadic eye movements in monkeys. *J. Neurophysiology*, 56, 196-207.

[77] Mergner, T., Nasios, G., Maurer, C., and Becker, W. (2001). Visual object localisation in space: Interaction of retinal, eye position , vestibular and neck proprioceptive information. *Exp Brain Res*, 141, 33-51.

[78] Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope, *Behavior Research Methods,Instruments and Computers*, 19, 449-456.

[79] Mirtich, B., and Canny, J. (1995). Impulse based simulation of rigid bodies. *Proceedings of the 1995 symposium on Interactive 3D graphics*, April

[80] Montana, D. (1992). Contact Stability for Two-Fingered Grasps, *IEEE Transactions on Robotics and Automation*, 8, pp. 421-430.

[81] Mora, F., and Myers, R.D. (1977). Brain self-stimulation: direct evidence for the involvement of dopamine in the prefrontal cortex. *Science*, 197, pp. 1387-1389.

[82] Murphy, K., Torralba, A., and Freeman, W. (2003). Using the forest to see the trees: A graphical model relating features, objects and scenes. *Advances in neural information processing.*

[83] Nash, J. F. (1950). Equilibrium points in N-person games. *Proc. Natl. Acad. Sci.*, USA, 36, pp. 48-49.

[84] Niemeier, M., Crawford, J.D., and Tweed, D.B. (2003). Optimal transsaccadic integration explains distorted spatial perception. *Nature*, 422, 76-79.

[85] Ng, A.Y., and Russell, S. (2000). Algorithms for inverse reinforcement learning. *Proceedings of the Seventeenth International Conference on Machine Learning.*

[86] Ohtsuka, K. (1994). Properties of memory-guided saccades toward targets flashed during smooth pursuit in human subjects. *Invest. Ophthalmol. Vis. Sci.*, 35, 509-514.

[87] Paulignan Y, Jeannerod M, MacKenzie C, Marteniuk (1991). Selective perturbation of visual input during prehension movements. 2. The effects of changing object size. *Exp Brain Res.*, 87, 407-420.

[88] Paulignan, Y., Frak, V.G., Toni, I., Jeannerod, M. (1997). Influence of object position and size on human prehension movements, *Experiental Brain Research*, 114, 226-234.

[89] Pelisson, D., Guitton, D., and Munoz, D. P. (1989). Compensatory eye and head movements generated by the cat following stimulation-induced perturbations in gaze position. *Exp. Brain Res.*, 78, 654-658.

[90] Pouget, A., Doucom, J-C, Torri, J., Bavelier, D., (2002). Multisensory spatial representations in eye-centered coordinates for reaching. *Cognition*, 83, B1-B11.

[91] Pouget, A., Deneve, S and Duhamel, J.R. (2002). A Computational Perspective on the Neural Basis of Multisensory Spatial Representations. *Nature Neuroscience Review*, 3, pp. 741-747.

[92] Sabes, P.N., and Jordan, M.I. (1997). Obstacle Avoidance and a Sensitivity Model of Motor Planning. *The Journal of Neuroscience*, 17, 7119-7128.

[93] Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, 76, pp. 2790-2793.

[94] Saunders, J and Knill, D. C. (2004). Visual feedback control of hand movements. *J. of Neuroscience*, 24, 3223-3234.

[95] Schlag, J., Schlag-Rey, M., and Dassonville, P. (1990). Saccades can be aimed at the spatial location of targets flashed during pursuit. *J Neurophysiology*, 64, 575-581.

[96] Schrater, P. (1998). Local Motion Detection: Comparison of Human and Ideal Model Observers. *Ph.D. thesis, Philadelphia: University of Pennsylvania*.

[97] Schultz, W. Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* 275, pp. 1593-1598.

[98] Shannon, C.E., and Weaver, W. (1949). The mathematical theory of communication, Champaign, IL: U. Illinois Press.

[99] Sheth, B.R., and Shimojo, S. (2001). Compression of space in visual memory. *Vision Research*, 41, 329-341.

[100] Simoncelli, E. P. (1993). Distributed Analysis and Representation of Visual Motion. *Ph.D. Thesis, Cambridge, MA: Massachusetts Institute of Technology.*

[101] Sivak, B. and MacKenzie, C.L. (1990): Integration of visual information and motor output in reaching and grasping: the contributions of peripheral and central vision, *Neuropsychologia*, 28, 1095-1116.

[102] Smeets, J.B.J., Brenner, E. (1999) A new view on grasping. *Motor Control*, 3, 237-271.

[103] Sober, S.J., and Sabes, P.N. (2005). Flexible Strategies for Sensory Integration During Motor Planning.*Nature Neuroscience*, 8,490-497.

[104] Song, J.H., and Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, 6, 982-995.

[105] Sparks, D.L., Mays, L.E. (1983). Spatial localization of saccade targets. I. Compensation for stimulation-induced perturbations in eye position. *J Neurophysiology*, 49, 45-63.

[106] Stahl, J.S. (2001) Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity. *Exp Brain Res*,136, 200–210.

[107] Steinbach, M. J. (1986) Proprioceptive knowledge of eye position. *Vision Res.*, 27, 1737–1744.

[108] Thrun, S. (1997). Bayesian landmark learning for mobile robot localization. *Machine Learning*.

[109] Todorov, E., and Jordan, M. (2003). Optimality feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11), pp. 1226-1235.

[110] Trommershauser, J., Maloney, L.T., and Landy, M.S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. *J. Opt. Soc. Am. A*, 20, pp.1419-1433.

[111] Trommershauser, J., Gepshtein, S., Maloney, L.T., Landy, M.S., Banks, M.S. (2005). Optimal compensation for changes in task-relevant movement variability. *Journal of Neuroscience*, 25, 7169-7178.

[112] van Beers, R.J., Sittig, A.C., and Denier van der Gon, J.J. (1996). How humans combine simultaneous proprioceptive and visual position information. *Exp. Brain Res.*, 111, 253-261.

[113] van Beers, R.J., Sittig, A.C., and Denier van der Gon, J.J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, 122, pp. 367-377.

[114] van Beers, R.J., Sittig, A.C., and Denier van der Gon, J.J. (1999). Integration of proprioceptive and visual position-information: an experimentally supported model. *Journal of Neurophysiology*, 81, pp. 1355-1364.

[115] Van Donkelaar, P. and Staub, J. (2000). Eye-hand coordination to visual versus remembered targets. *Exp Brain Res*, 133, 414-418.

[116] van Opstal, A. J., and van Gisbergen, J. A. M. (1989). Scatter in the metrics of saccades and properties of the collicular motor map, *Vision Research*, 29, 1183-1196.

[117] Verboven, S., and Hubert, M. (2005), LIBRA: a MATLAB Library for Robust Analysis, *Chemometrics and Intelligent Laboratory Systems*, 75, 127-136.

[118] Walker, M.F., Fitzgibbon, J., Goldberg, M.E. (1995). Neurons of the monkey superior colliculus predict the visual result of impeding saccadic eye movements. *J Neurophysiology*, 73, 1988-2003.

[119] Webb, A. (2002). Statistical Pattern Recognition, John Wiley and Sons, New York.

[120] Whitaker K, Latham K (1997) Disentangling the role of spatial scale, separation and eccentricity in Webers law for position. *Vision Res*, 37, 515524.

[121] White, J.M., Sparks, D.L., Stanford, T.R. (1993). Saccades to remembered target locations: an analysis of systematic and variable errors. *Vis Res*, 34, 79-92.

[122] Wing, A.M., Turton, A. and Fraser, C. (1986). Grasp size and accuracy of approach in reaching. *Journal of Motor Behaviour*, 18, 245-260.

[123] Wise RA, and Hoffman DC (1992) Localization of drug reward mechanisms by intracranial injections. *Synapse* 10:247-263.

[124] Wolpert, D.M, Flanagan, J.R. (2001). Motor prediction. Current Biology, 11, R729-732.

[125] Wolpert, D.M., Ghahramani, Z, and Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880-1882.

[126] Wolpert, D.M., Goodbody, S.J. and Husain, M. (1998). Maintaining internal representations: the role of the human superior parietal lobe. *Nature Neuroscience*, 1, pp. 529533.

[127] Yuille, A.L., and Bulthoff, H.H. (1996). Bayesian theory and psychophysics. Perception as Bayesian Inference, (Eds.) D.Knill, W. Richards. Cambridge University Press, pp. 123-161.

[128] Zemel,R. S. (1997). Combining probabilistic population codes. *International Joint Conference on Artifical Intelligence*, Denver, CO: Morgan Kaufmann.

[129] Zemel,R. S. (1998). Probabilistic interpretation of population codes. *Neural Computation*, 10, pp. 403-430.

[130] Zivotofsky, A.Z., Rottach, K.G., Averbuch-Heller, L., Kori, A.A., Thomas, C.W., DellOsso, L.F., Leigh, R.J. (1996). Saccades to remembered targets: the effects of smooth pursuit and illusory stimulus motion. *J Neurophysiology*, 76, 3617-3632.