

# Auxiliary object knowledge influences visually-guided interception behavior

Peter W. Battaglia\*  
University of Minnesota

Paul R. Schrater†  
University of Minnesota

Daniel J. Kersten‡  
University of Minnesota

## Abstract

This work investigated how humans integrate visual information with object knowledge for interception behavior. When attempting to intercept a moving object using only monocular visual information, the optimal interception position may be ambiguous—the observer may be viewing a small object that is near or a large object that is far away. Regardless, humans are quite adept at monocular interception so it is likely that additional information is incorporated to disambiguate the visual information. We hypothesize that object size information is integrated to accomplish this disambiguation. This sort of auxiliary information integration is well-defined by a Bayesian model of information propagation. We derived a Bayesian model that represents scene attributes relevant to intercepting an object and relations among these attributes. Our model combines sensory measurements with prior scene knowledge to infer an object's position. To test our model we asked participants to intercept a moving ball in virtual reality. In some trials participants were able to see and touch the ball before intercepting it, in others they were only able to see it. When allowed to touch the ball, participants showed improved interception performance. Effectively, they discounted the variation in image size that was caused by variation in object size to obtain more accurate knowledge of object distance. This discounting is consistent with Bayesian information propagation and confirms our hypothesis that human participants use Bayesian inference to estimate an object's distance for interception.

**CR Categories:** I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Modeling and recovery of physical attributes; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; J.4 [Social and Behavioral Sciences]—Psychology; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Sensor Fusion;

**Keywords:** interception, perceptual inference, Bayes, virtual reality

## 1 Introduction

### 1.1 Perception for interceptive behavior

Intercepting a moving object requires accurate perceptual decision-making and precise motor control. The observer must determine the

time and position at which the object will reach him, and coordinate a movement to that position at that time. Timing an interception is a well-studied behavior; several authors have reported that information contained in the visual display of an object is sufficient to determine an appropriate interception time [Lee and Reddish 1981; Lee and Young 1985; Peper et al. 1994; Caljouw et al. 2004].

Selecting an interception position can be quite difficult because the available image data is less reliable, or even ambiguous under monocular viewing. A simple example is the ambiguity that occurs when relying on an object's image size to make a distance judgment. The object may be small and close to the observer, or large and far from the observer, yet project to the same image size in both situations. Despite this adversity, monocular interception is still quite accurate [Servos and Goodale 1998].

An observer needs information about the object's size to disambiguate the distance indicated by image size. Unfortunately, object size information is not directly measurable as the object approaches in the moments before an interception. But by gathering object size information from previous experience with the object, an observer may incorporate that experience with the measured image size to arrive at an unambiguous prediction of object distance. A Bayesian model for inferring object distance provides a principled framework for this information combination problem. The purpose of this paper is to formulate the problem of using object size information for interceptive behavior as a process of Bayesian inference and test our model's predictions by comparing them to actual human interceptive performance.

### 1.2 Bayesian modeling of perception

Perception is a process of probabilistic inference of scene properties from sensory data. In the Bayesian approach to modeling perception, scene attributes are treated as random variables; knowledge about these attributes is represented as probability distributions over the possible values [Kersten et al. 2004; Knill and Richards 1996]. Dependencies among scene attributes are represented as conditional probability distributions. The observer's cumulative knowledge about the scene can be represented as a joint probability distribution over all of the scene attributes that can be factored according to the conditional dependencies.

There is growing evidence that biological perception is a process of Bayesian inference [Knill 1998; Ernst and Banks 2002; Weiss et al. 2002; Mamassian and Landy 2001; Körding and Wolpert 2004; Battaglia et al. 2003]. In many of these studies, human participants achieved near-optimal performance given the information they were provided. This supports the theory that for some scenes, the brain has well-defined representations of the relationships among the scene's attributes and is able to integrate sensory measurements with prior knowledge to make estimates of a desired scene property.

Sensory measurements, prior knowledge, and the dependencies among scene attributes are the fundamental components of a Bayesian model. Appendix A provides some details of their roles and a simple example. For a more complete discussion of perception as Bayesian inference consult Kersten et al. [2004].

\*e-mail: batt0086@umn.edu

†e-mail: schrater@umn.edu

‡e-mail: kersten@umn.edu

<sup>0</sup>This work was supported by NIH grant: R01 EY015261 and the NSF Graduate Research Fellowship Program.

## 2 Theory and experiment

### 2.1 Experimental task

We measured human performance on intercepting a moving ball to test our model’s predictions. In our experiment we asked participants to intercept a ball moving in virtual reality that they viewed monocularly. We constrained their hand to move on a fixed line that intersected their viewpoint. Because of this constraining line, the only component of the interception position that was under the participant’s control was the interception distance. The ball moved through the scene and crossed this line at some position (called the *crossing distance*) within the participant’s reach. For a single interception, we recorded a participant’s interception distance as their judgment of ball distance.

As mentioned in the Introduction (section 1.1), monocular visual information for estimating the optimal interception distance is ambiguous. For a particular image size, if the ball is small then it must be near, and will require a near interception distance. If the ball is large then it must be far, and will require a far interception distance.

For a Bayesian observer, auxiliary ball size information helps disambiguate the distance indicated by image size. Therefore, we tested the impact that auxiliary information of the ball’s size had on human interception performance. Before some interceptions, we allowed our participants to touch the ball (haptic condition). For other interceptions participants were not allowed to touch the ball and had to rely on their prior knowledge of ball sizes (no-haptic condition).

Figures 1 and 2 show diagrams of the experimental scene. The letters denote relevant scene attributes for interception. The participant needs to estimate the crossing distance,  $R$ , to successfully intercept the ball.  $\theta$  indicates the ball’s image size and  $S$  indicates the ball’s actual size (diameter).

### 2.2 Bayes’ net

Bayesian influence is elegantly represented by a class of graphical models called a “Bayes’ network” [Pearl 1988]. It is possible to model perceptual inference of desired scene attributes in Bayes’ nets [Schrater and Kersten 1999; Schrater and Kersten 2000]. One advantage of this representation is that it is easy to visualize the relations among scene attributes. Another advantage is that subgroups of random variables may be easily identified as instances of “cue combination” or “perceptual explaining away” [Kersten et al. 2004]. Recognizing a familiar subgroup may lend insight into what brain mechanisms are responsible for a particular component of the perceptual inference process.

A Bayes’ net that represents the scene attributes and their conditional dependencies for our interception task is shown in Figure 3. This Bayesian model uses monocular visual information ( $\hat{\theta}$ ) and haptic information ( $\hat{S}$ ) to compute a posterior probability distribution over crossing distances,  $p(R | \hat{\theta}, \hat{S})$ .

We assumed observers had no prior assumptions about the cartesian coordinates of the crossing distance,  $X$  and  $Z$ . This is equivalent to assuming a uniform probability distribution on possible values of  $X$  and  $Z$ . Consequently no information from  $\psi$  can propagate through the  $X$  and  $Z$  nodes to influence  $R$ , rendering  $\psi$  and  $R$  independent. This meant that the remaining relevant scene attributes for the Bayesian observer trying to intercept the ball were: the crossing distance  $R$ , image size  $\theta$ , image size measurement  $\hat{\theta}$ , ball size  $S$ , and ball size measurement  $\hat{S}$ .

Appendix B gives the derivation of  $p(R | \hat{\theta}, \hat{S})$ . This derivation does not need to be understood to continue to the model predictions, but an interested reader may find it instructive. The important point of the derivation is that an observer can compute the posterior probability by combining sensory measurements with prior knowledge according to known conditional probability distributions.

### 2.3 Predictions

#### 2.3.1 Bayesian observer crossing distance estimation

The Bayesian observer computed an optimal interception distance. The optimal interception distance was defined as the crossing distance with the highest posterior probability given the image data and prior knowledge (called “maximum posteriori” estimation). The optimal interception distance reflects any relevant scene knowledge, including a haptic ball size measurement if it is available.

According to our Bayesian model (Figure 3), image size depends on the ball’s size and crossing distance. When image size is the only measurement available to the observer (no-haptic condition), the ball’s crossing distance must be inferred using the image size measurement and an assumption of ball size drawn from prior knowledge. In our experiment the actual ball size was randomly selected from a finite uniform distribution so the participant’s assumption of ball size was not necessarily accurate. If the actual ball was smaller than the assumed ball size, the observer would judge it to be farther away than it actually is. If the actual ball was larger than the assumed ball size, the observer would judge it to be nearer than it actually is. These systematic misjudgments will appear as biases in judged distances according to Bayesian inference.

When participants have a measurement of the ball size (haptic condition), that measurement may combine with the prior ball size knowledge and create an assumption about the ball’s size that is generally more accurate than in the no-haptic condition. When attempting to disambiguate the crossing distance as indicated by image size, the more accurate ball size information of the haptic observer will result in more accurate predicted crossing distance estimates. The first prediction of a Bayesian observer is that *auxiliary size information will improve interception performance*.

The way auxiliary ball size information improves the crossing distance estimates is it allows the observer to discount the effect of ball size on image size. Since the image size naturally reflects the ball’s distance and the ball’s size, an observer who is able to “explain away” the effect of ball size on image size will have a better estimate of the ball’s distance. In this way the more accurate ball size knowledge of the haptic observer will tend to eliminate the small and large ball biases discussed above. The second prediction of a

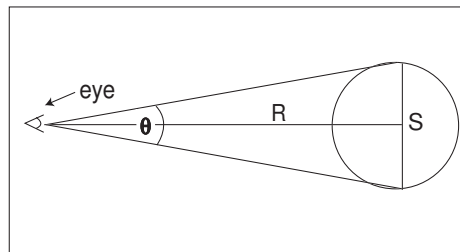


Figure 1: OBSERVER AND BALL DIAGRAM: *The relationship between object size,  $S$ , radial distance between the object and the observer,  $R$ , and image size (expressed as visual angle),  $\theta$ .*

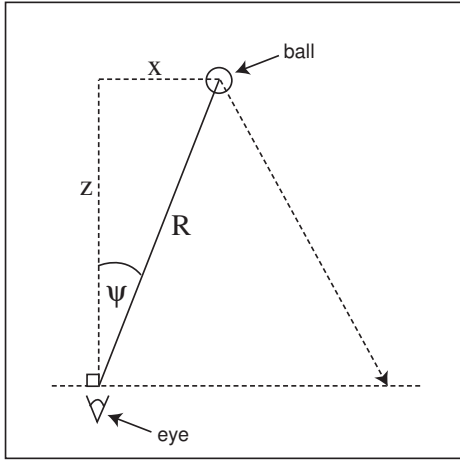


Figure 2: OBSERVER AND BALL TRAJECTORY DIAGRAM: The relationship between the observer, the ball's cartesian coordinates ( $X, Z$ ), and the ball's polar coordinates ( $R, \psi$ ) as it moves in the  $X$ - $Z$  plane.

Bayesian observer is that the improvement in interception performance derives from discounting the ball size's influence on image size.

### 2.3.2 Information propagation

Figure 3 depicts inference in the no-haptic and haptic cases. Three example experiment trials are demonstrated in which three balls with different sizes are presented at different distances. The distances are selected so that  $\theta$  for each ball is the same. The boxes contain no-haptic and haptic observers' probability distributions over scene attributes for the different trials. Figure 3(A.) shows three different measurements of  $S$  indicated by lines at three different size values for the haptic case. In the no-haptic case there are no  $\hat{S}$ s. In Figure 3(B.) the red curve is  $p(S)$  and is all the information a no-haptic observer has about  $S$ . The blue curves are the probabilities on  $S$  for the balls in the haptic condition, inferred from  $\hat{S}$ s. Figure 3(C.)'s green curve is the probability of  $\theta$  inferred from  $\hat{\theta}$ . As mentioned  $\theta$  is the same for all three balls hence there is one green curve for all three balls.

In order to perform an interception, the  $R$  must be inferred. When the no-haptic observer computes the posterior probability of  $R$  for each of the three balls, the red solid line distribution from 3(A.) combines with the green solid line distribution from 3(C.) to form the red solid line distribution in Figure 3(D.). The red curve in 3(D.) represents the posterior probability of crossing distance,  $p(R | \hat{\theta})$ , for the no-haptic observer. Because  $\theta$  is the same for all three balls, and the no-haptic observer relies on the same  $p(S)$ , his  $p(R | \hat{\theta})$  distributions are the same for the three different balls. The blue dotted curves in 3(D.) are the haptic observer's  $p(R | \hat{\theta}, \hat{S})$  distributions. The haptic observer combines the blue dotted line distributions from 3(A.) with the green solid line distribution from 3(C.) to form the blue dotted line distributions in Figure 3(D.).

Since the no-haptic observer's assumption of ball size must be the same for each ball, the differences among ball sizes do not affect the posterior distribution on  $R$ . For the haptic observer the different ball sizes cause different posterior probabilities on  $R$ . In actuality, for different sized balls to form the same image size, they must be at different distances, so the haptic observer's information is more

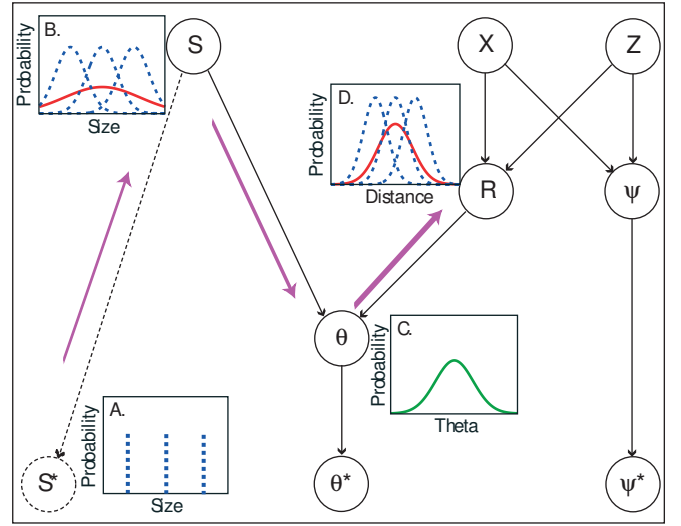


Figure 3: INFORMATION FLOW THROUGH BAYES' NET IN NO-HAPTIC AND HAPTIC CONDITIONS: The nodes correspond to scene attributes identified in Figures 1 and 2 as well as measurements of those attributes ( $S^*, \theta^*, \psi^*$ ) in the text the measurements are denoted  $\hat{S}, \hat{\theta},$  and  $\hat{\psi}$ , respectively. The black dotted lines indicate direct ball size information that is available to participants only in the haptic condition. The purple arrows represent information flow among the nodes through Bayesian inference.

accurate.

## 3 Methods

### 3.1 Stimuli

Experiments were conducted in a custom visual haptic workbench that provides calibrated visual and haptic information about objects in a virtual scene (Figure 4). Participants viewed the virtual scenes, head-fixed (using a headrest) and monocularly (eye-patch on left eye), displayed on a 21" CRT (pixel resolution of 1600 x 1200, 85 Hz). The display was positioned above and at a 90 deg angle from the participant's main direction of gaze and was viewed through a mirror to create a virtual image of the scene in front of the participant. Below the mirror was a reachable space in which participants held a PHANTOM force-feedback stylus (Sensible Technologies) that tracked their hand position throughout the experiment. The PHANTOM was also used to generate force feedback to simulate stylus-object contact when objects were touchable (haptic condition only), and to generate contact feedback during the interception task. A common coordinate frame was established for this virtual space that put the origin at the participant's right eye, the x and y axes lined up with the imaging plane, and gaze direction was along the -z axis.

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

Figure 5 shows a screenshot of the virtual workspace. The visual stimuli included a 3mm virtual sphere that spatially coincides with

the participant's actual hand position in the virtual workspace. The bottom of the virtual workspace was a visible and touchable ground plane with a checkerboard visual texture. The hand sphere and ground plane were always present throughout every trial.



Figure 4: PARTICIPANT SEATED AT WORKSPACE: A participant performing the experiment. He views a mirror that reflects the display's image into his eyes. Below the mirror he holds a force-feedback stylus that provides a haptic interface to the virtual workspace. All visual and haptic stimuli are spatially coincident and rendered accurately.

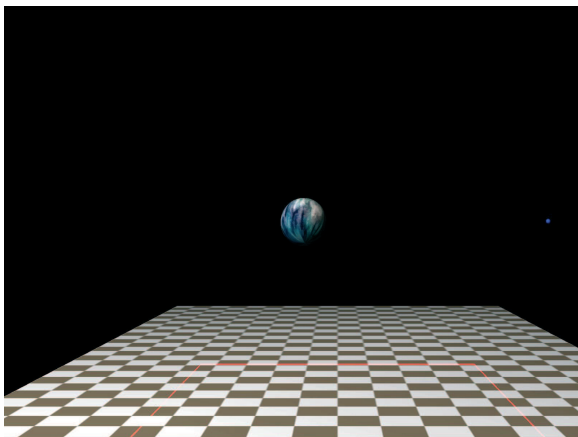


Figure 5: VIRTUAL WORKSPACE: This is a screenshot of the virtual workspace that the participants experienced. The small blue sphere represents the participant's hand position. The larger ball is the visual and haptic target. The ground plane is the checkerboard surface along the bottom of the scene.

The object that the participants interacted with was a virtual ball whose diameter was randomly selected from between 1.4 and 4.2 cm. Each ball had a random texture mapped onto it that did not contain any cues to distance. In some trials, participants were allowed to touch the ball and the visual and haptic stimuli accurately reflected the ball's geometry.

## 3.2 Task

The participant's task was to intercept a moving ball as it crossed an invisible line through the origin, +15 deg (positive angles were

rightward) from the z-axis in the x-z plane (the *constraining line*). The interception involved moving the stylus to the position along the line where the ball was predicted to cross. To simplify the task, the stylus was constrained to stay on the line by applying a cylindrical force field directed toward the line whose magnitude was determined by the stylus end-point's perpendicular distance from the line. Preceding the interception phase was an exploration phase that provided visual and haptic feedback about the ball's size for half the trials. Participants were instructed to use the exploration phase to observe the ball so that they could best perform a subsequent interception of it.

## 3.3 Trials

### 3.3.1 Exploration phase

Each trial began with one of two types of exploration phases: haptic and no-haptic. Each block contained an equal number of randomly interleaved haptic and no-haptic trials. During the exploration phase, the ball was always located in the x-z plane and remained stationary. In the haptic exploration condition, participants were able to see and feel a virtual ball for as long as they wanted. Before they were allowed to continue to the interception phase of the trial, they had to touch the ball at least once. In the no-haptic exploration condition, the participants' hands were constrained to move along the constraining line used in the interception phase. Participants were allowed to view the ball for a minimum of one second and no maximum time. The ball's distance was randomly selected from between 30 and 64 cm. The ball's eccentricity was randomly selected from between -8.5 and 8.5 degrees, where 0 degrees is in front of the participant along his line of sight.

### 3.3.2 Interception phase

Once the participant was satisfied with the exploration, he depressed the mouse button to begin the interception phase of the trial. Regardless of whether a trial's exploration phase was haptic or no-haptic, the interception phases were the same. Figure 6 shows a diagram of the interception phase. The ball was always located in the x-z plane during the interception phase. Once the mouse button was depressed, the stylus was constrained to move along the constraining line described above. At the same time, the ball rapidly travelled to a random starting position in the left side of the virtual workspace and remained stationary once it arrived. The ball's starting position was determined in polar coordinates: the distance from the eye was randomly selected from a uniform distribution between 100 and 150 cm; while the ball's starting angle (the visual eccentricity) from the z axis was uniformly selected from between -17 and -5 degrees. Once the participant was ready to begin the actual interception, he depressed the mouse button and the ball began to move toward the constraining line. The ball's velocity was randomly selected from between 25 and 37.5 cm/s. In each trial the ball crossed the constraining line within the participant's reach and continued off the screen. The crossing distance on the line was randomly selected from between 34 and 60 cm. The ball's total travel time was between 1.3 and 4.8 seconds.

The participant's task was to position the stylus tip at the point that the ball was predicted to cross the constraining line. As the ball crossed the constraining line, the participant received visual and haptic feedback on the accuracy of the interception. If the stylus tip was positioned within 3.2 cm of the ball's center the moment it crossed the constraining line, a virtual collision was felt (a small force pulse was applied to the stylus the participant held) and



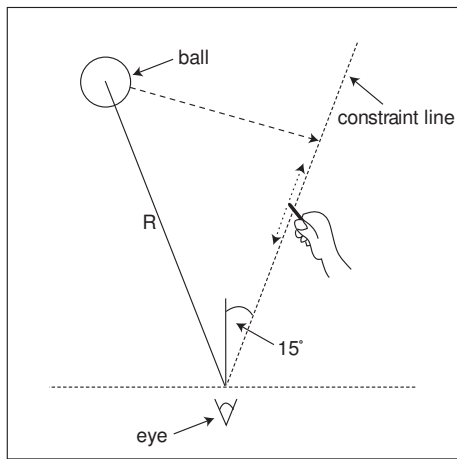


Figure 6: INTERCEPTION PHASE DIAGRAM: The ball starts on the left side of the virtual workspace and moves toward the participant's hand that is located on the constraining line. The participant attempts to intercept the ball as it crosses the constraining line.

the hand position sphere briefly turned green. Otherwise, no force-feedback was generated and the hand sphere briefly turned red. The stylus position at the time the ball crossed the constraining line was recorded and the distance of that position from the participant is referred to as the “judged distance” in the Results (section 4).

### 3.4 Participants

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

### 3.5 Data analysis

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

Two different 2D regression analyses were performed on each participant's data. Figure 7 shows the scatterplots for a typical participant's interception behavior. Each point represents a single trial's interception. The x-axis shows the actual distance that the ball crossed the constraining line (crossing distance) and the y-axis shows the participant's judged distance. In the first 2D regression, the linear dependence between the crossing and judged distances was measured. In the second 2D regression, the linear dependence between the image size and judged distance was measured. Confidence intervals around all regression parameters and significance tests were computed using a bootstrap procedure that iterated 1000 times. The specific results from the different regressions are presented below.

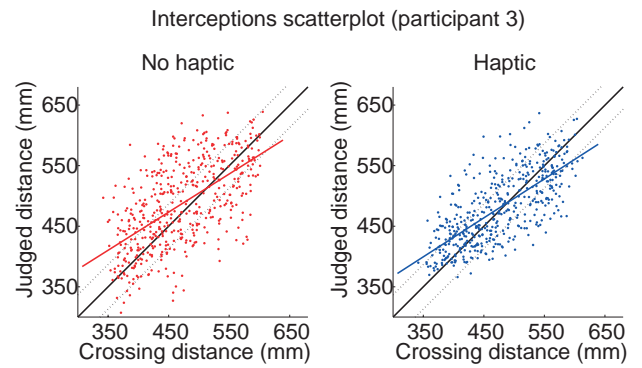


Figure 7: INTERCEPTIONS SCATTERPLOT (PARTICIPANT 3): These scatterplots show a typical participant's interception behavior. Each dot represents a single trial's data point. The x-axis shows the actual distance that the ball crossed the constraining line (crossing distance). The y-axis shows the judged distance where the participant thought the ball would cross. The black diagonal line shows the line that indicates perfect interceptions. Within the flanking dotted lines is the range in which participants received positive feedback for their interception attempt. The colored line is the regression line for the data.

## 4 Results

### 4.1 Does auxiliary ball size information improve interception performance?

From the 2D regression analysis on actual and judged distances,  $r^2$  is the proportion of variance in individual judged distances that is accounted for by the linear coefficient of crossing distance. The greater the value of  $r^2$ , the less the data is scattered from the best-fit line. Figure 8 shows the  $r^2$  values for each participant. Decreased data scatter around the regression line indicates more consistent interception performance, and generally more accurate interceptions. Thus  $r^2$  was used as a measure of participant interception performance; significant differences between haptic and no-haptic condition  $r^2$ -values indicated significant improvements in interception performances. Participants 1-4 all had significantly less variance in interception performance in the haptic condition over the no-haptic condition ( $p < .001$  for participants 1, 3, 4;  $p < .003$  for participant 2). Participant 5 showed insignificantly less variance in interception performance in the haptic condition over the no-haptic condition. Her failure to significantly improve her performance was interpreted as a ceiling effect; her performance in the no-haptic condition was so high that motor noise or other limits on performance may have contributed to her inability to improve her performance above that high level in the haptic condition.

For four of five participants, providing auxiliary haptic information about the ball's size significantly improved their interception performance. A Bayesian view of perceptual inference predicts that auxiliary ball size information cooperates with other information to improve crossing distance estimates. These results qualitatively support the assertion that the brain is able to exploit its knowledge of the relationship between ball size and crossing distance when generating interception behaviors.

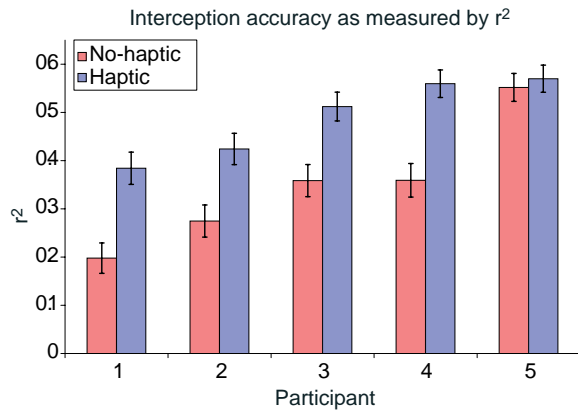


Figure 8: INTERCEPTION ACCURACY AS MEASURED BY  $r^2$ : The bars in the graph represent  $r^2$ -values. The red bars depict accuracy for the no-haptic condition, the blue bars depict accuracy for the haptic condition. Participants 1-4 show significantly improved performance in the haptic condition over the no-haptic condition.

## 4.2 Does the improvement in interception performance derive from discounting the ball size's influence on image size?

In the no-haptic condition participants had no auxiliary ball size information. In this condition we expected they would rely on a general prototype ball size, perhaps represented as a prior probability distribution over a range of possible ball sizes. Like the example above with a small, near ball and a large, far ball, distance ambiguity existed in the no-haptic condition because ball sizes on different trials varied between 1.4cm and 4.2cm. If the observer assumed a common ball size for all trials (somewhere in between the smallest and largest ball sizes for instance), the smaller balls would be judged further than they actually were and the large balls would be judged nearer than they actually were. Figure 9 shows the same scatterplot as Figure 7 but with each trial point's color indicative of the ball size used in that trial. Trials were assigned to one of three bins according to the trial's ball size with equal numbers of balls in each bin (because ball sizes were uniformly distributed over a finite range, each bin roughly corresponds to ball size ranges with the same width). Best-fit lines for each different bin of trials are also shown. The no-haptic condition's best-fit lines are more separated than the haptic condition's best-fit lines. This means in the haptic condition participants' judged distances did not systematically depend on ball size as much as in the no-haptic condition. The effect supports a prediction of the Bayesian inference model: participants should discount the effect that ball size has on image size and make judgments that are less dependent on the ball size component of image size variation.

To quantify the above effect, a 2D regression analysis on image sizes and judged distances was performed. The linear coefficient on the image size term reflects its correlation with judged distance. According to the above logic, across no-haptic trials with the same image size, participants should judge the crossing distance to be the same regardless of variations in ball size. When the image size varied in the no-haptic condition, participants' distance judgments should be inversely proportional to image size.

In the haptic condition, participants had access to auxiliary ball size information. In this condition we expect ball size information should cooperate with image size to influence distance judgments. Specifically, for trials with the same ball size, if the image size

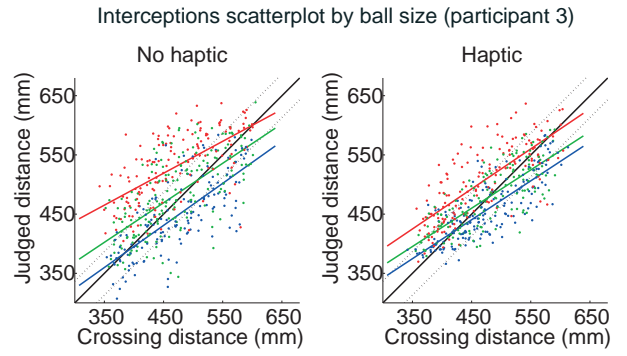


Figure 9: INTERCEPTIONS SCATTERPLOT BY BALL SIZE (PARTICIPANT 3): This figure is the same as Figure 7 but with each trial's ball size indicated by one of three colors. Red dots represented small balls, green dots represented medium sized balls, and blue dots represented large balls. The colored lines are regression lines for each size of ball. In the haptic condition those regression lines were nearer to each other.

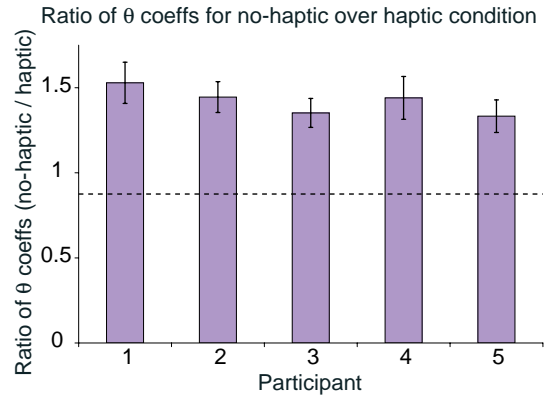


Figure 10: RATIO OF  $\theta$  COEFFICIENTS FOR NO-HAPTIC OVER HAPTIC CONDITION: The bars represent the ratio of the no-haptic linear coefficient of image size and judged distance over the haptic coefficient for each participant.

varies, the participants' distance judgments should vary inversely. Likewise, for trials with the same image size, if the ball size varies, the participants' distance judgments should vary directly. So we expect that in the haptic condition, the linear coefficient of image size should be non-zero to reflect the variance with judged distance. But, the linear coefficient of image size should be less than in the no-haptic condition because the judged distances share dependence on ball size and image size.

We compared the linear coefficient of image size between the no-haptic and haptic conditions by computing their ratio. A ratio of one implies there was no distinction in image size dependence between no-haptic and haptic conditions, a ratio significantly greater than one shows that when auxiliary ball size information was provided to participants, their judged distances discounted the effect of ball size on image size and depended less strongly on image size. Figure 10 illustrates these ratios for each participant. All participants showed ratios significantly higher than one ( $p < .001$ ).

## 5 Conclusion

### 5.1 Interpretation of results

To summarize the results, when haptic information about a ball's size was provided to five participants, four were significantly better at intercepting the ball when it moved through the scene. The participant who did not display this trend appeared to suffer from ceiling effects because her performance was so high even in the no-haptic condition, she had no room to improve in the haptic condition. This result is consistent with the hypothesis that the brain integrates auxiliary ball size information to improve distance judgments for interception behavior.

To determine how the auxiliary ball size information improved distance judgments, we contrasted how participants treated image size information between the no-haptic and haptic conditions. In the no-haptic condition, participants' distance judgments were significantly more linearly related to image size than in the haptic condition. In the haptic condition, participants interpreted the image size in the context of the size of the ball to discount the effect of ball size on image size. This allowed participants to make distance judgments that reflected the ball's crossing distance unbiased by the size of the ball that was being intercepted in the haptic condition. This is consistent with a model of perceptual inference in which auxiliary measurements are used to "explain away" invalid interpretations of a scene and reduce possible perceptual decisions to the remaining valid interpretations [Knill and Kersten 1991; Bloj et al. 1999].

### 5.2 Impact and implications

This study contributes to a growing corpus of evidence supporting the view that the human brain uses probabilistic inference to make perceptual decisions [Kersten et al. 2004; Knill 1998; Ernst and Banks 2002; Weiss et al. 2002; Mamassian and Landy 2001; Körding and Wolpert 2004; Battaglia et al. 2003]. It is clear that biological perceptual systems use prior knowledge in addition to sensory measurements during perception. Still many details of Bayesian models for perception remain to be experimentally tested.

For instance, when a Bayesian observer infers a posterior probability distribution, the posterior reflects the quality of the information that was combined to compute it. An open question is whether the brain measures the quality of the information it collects so as to perform optimal inference. For example, in the interception task presented in this study, if the auxiliary ball size information were diminished in quality, a Bayesian observer would rely more on the prior knowledge of ball size and less on the haptic measurement of ball size when computing the probability of crossing distance. Experimentally human participants could be tested for this prediction by presenting "squishy" balls during haptic exploration whose precise size is difficult to determine. Presumably this diminished ball size information would be less effective in "explaining away" variation in image size, and reduce the difference in interception performance between the haptic and no-haptic conditions.

In general, Bayesian models of scene attributes are a powerful tool for quantifying scene information with respect to an observer. Comparing human performance with Bayesian model predictions allows experimentalists to analyze how scene attributes influence perception. Further, evidence of near-optimal performance by human participants in behavioral tasks supports the idea that the brain exploits knowledge of the dependencies among scene attributes to make perceptual decisions. This evidence give theoreticians a rational basis

on which to propose similar mechanisms being involved with other behavioral, cognitive, and neurophysiological observations.

## A Components of a Bayesian model

*Sensory measurements* are observable scene attributes. Consider an observer trying to determine the curvature of an object. The curved surface generates a visual texture pattern on an observer's retina. This pattern is an example of a sensory measurement.

*Prior knowledge* about scene attributes is represented as a probability distribution. For example, if the above observer is outdoors, the sun is the likely light source. This observer's prior probability distribution will have most probability mass in the overhead direction.

*Dependencies among scene attributes* are represented as conditional probability distributions. In our example, the visual texture measurement is conditionally dependent on the curvature; as the curvature changes, the texture measurement changes accordingly. The conditional likelihood is often called the generative model because it specifies how a measurement is generated by a scene attribute.

Measurable attributes may also be dependent on other scene attributes that need not be explicitly estimated. These other attributes are called auxiliary attributes. For the observer trying to determine the surface curvature of an object, attributes such as lighting direction, spectral reflectance, and the rendering variables are considered auxiliary attributes: they influence the measurements of curvature (or the curvature values), but are not necessary to know the curvature.

The goal of a Bayesian model for perception is to compute a probability distribution over a desired scene attribute. Prior knowledge and sensory measurements may be combined according to the conditional probabilities that represent the relationships among scene attributes to compute the posterior probability of the desired scene property. The posterior probability distribution contains all of the information from the other scene attributes as they relate to the desired attribute. The posterior probability of a desired scene attribute,  $A$ , given some measurements the observer makes,  $M$ , is denoted  $p(A|M)$ . Often this posterior probability function is unknown, but using Bayes' rule it can be expressed as:

$$p(A|M) = \frac{p(M|A)p(A)}{p(M)}$$

$p(M|A)$  is the conditional likelihood of a particular scene attribute generating the observer's measurements. This represents the relationship between the measurements and the attributes that causes them.

$p(A)$  is the prior probability over possible  $A$  values. This term represents what the observer already knows about  $A$  before making any measurements.

$p(M)$  is the total likelihood of the measurement. This represents how likely a measurement is. If the observer knows the possible scenes, this term can be computed as the integral over these possibilities,  $p(M) = \int_A p(M|A)p(A)$ .

## B Derivation of posterior probability

The joint probability over the scene attributes relevant to interception is  $p(R, \theta, S, \hat{\theta}, \hat{S})$ . Factoring this distribution gives Eqn. 1 and

updating the factored joint probability by the conditional relationships defined in Figure 3 gives Eqn. 2:

$$p(R, \theta, S, \hat{\theta}, \hat{S}) = p(R | \theta, S, \hat{\theta}, \hat{S}) p(\hat{\theta} | \theta, \hat{S}, S) \quad (1)$$

$$= \frac{p(\hat{S} | \theta, S) p(\theta | S) p(S)}{p(\theta | S) p(S)} \quad (2)$$

Applying Bayes' rule, we can substitute:

$$p(R | \theta, S) = \frac{p(\theta | R, S) p(R | S)}{p(\theta | S)} \quad (3)$$

into Eqn. 2:

$$p(R, \theta, S, \hat{\theta}, \hat{S}) = \frac{p(\theta | R, S) p(R | S)}{p(\theta | S)} p(\hat{\theta} | \theta) p(\hat{S} | S) p(\theta | S) p(S) \quad (4)$$

$R$  and  $S$  are conditionally independent and we can cancel  $p(\theta | S)$  so the joint reduces to:

$$p(R, \theta, S, \hat{\theta}, \hat{S}) = p(\hat{\theta} | \theta) p(\hat{S} | S) p(\theta | R, S) p(R) p(S) \quad (5)$$

We assume the observer knows the generative models,  $p(\hat{\theta} | \theta)$  and  $p(\hat{S} | S)$ , and the relationship between ball size, crossing distance, and image size shown in Figure 1,  $p(\theta | R, S)$ . By combining prior knowledge about  $R$  and  $S$  ( $p(R)$  and  $p(S)$  respectively) with measurements  $\hat{\theta}$  and  $\hat{S}$ , we have a full representation of the joint distribution over these variables.

The posterior probability distribution that the observer wants to compute is  $p(R | \hat{\theta}, \hat{S})$ . Because  $\theta$  and  $S$  do not appear in this formulation, the observer integrates over those variables to get the joint over the relevant variables:

$$p(R, \hat{\theta}, \hat{S}) = \int_{\theta, S} p(R, \theta, S, \hat{\theta}, \hat{S}) d\theta dS \quad (6)$$

Applying Bayes' rule we can formulate the posterior probability:

$$p(R | \hat{\theta}, \hat{S}) = \frac{p(R, \hat{\theta}, \hat{S})}{\int_R p(R, \hat{\theta}, \hat{S}) dR} \quad (7)$$

Substituting from Eqns. 5 and 6 the observer represents the posterior probability as:

$$p(R | \hat{\theta}, \hat{S}) = \frac{\int_{\theta, S} p(\theta | R, S) p(R) p(\hat{\theta} | \theta) p(\hat{S} | S) p(S) d\theta dS}{\int_R \int_{\theta, S} p(\theta | R, S) p(R) p(\hat{\theta} | \theta) p(\hat{S} | S) p(S) dR d\theta dS} \quad (8)$$

Since we assumed the Bayesian observer knows all of these component distributions, it is capable of explicitly computing the posterior probability distribution. This formulation may not be analytically tractable. Regardless, the posterior can be numerically approximated in a straightforward manner.

## References

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, 7, 1391–1397.

BLOJ, M. G., KERSTEN, D., AND HURLBERT, A. C. 1999. Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature* 402, 877–879.

CALJOUW, S. R., VAN DER KAMP, J., AND SAVELSBERGH, G. J. P. 2004. Catching optical information for the regulation of timing. *Experimental Brain Research* 155, 427–438.

ERNST, M. O., AND BANKS, M. S. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 6870, 429–33.

KERSTEN, D., MAMASSIAN, P., AND YUILLE, A. 2004. Object perception as bayesian inference. *Annual Review of Psychology* 55, 271–304.

KNILL, D., AND KERSTEN, D. 1991. Apparent surface curvature affects lightness perception. *Nature* 351, 228–230.

KNILL, D., AND RICHARDS, W. 1996. *Perception as Bayesian Inference*. Cambridge University Press, Cambridge.

KNILL, D. 1998. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research* 38, 17, 2635–2656.

KÖRDING, K., AND WOLPERT, D. 2004. Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.

LEE, D. N., AND REDDISH, P. E. 1981. Plummeting gannets: a paradigm of ecological optics. *Nature* 293, 293–294.

LEE, D. N., AND YOUNG, D. S. 1985. Visual timing of interceptive action. In *Brain Mechanisms and Spatial Vision*, D. Ingle, M. Jeannerod, and D. Lee, Eds. Martinus Nijhoff, Dordrecht.

MAMASSIAN, P., AND LANDY, M. S. 2001. Interaction of visual prior constraints. *Vision Research* 41, 20, 2653–68.

PEARL, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, San Mateo, CA.

PEPER, L., BOOTSMAN, R. J., MESTRE, D. R., AND BAKKER, F. C. 1994. Catching balls: how to get the hand to the right place at the right time. *Journal of Experimental Psychology* 20, 3, 591–612.

SCHRATER, P. R., AND KERSTEN, D. 1999. Statistical structure and task dependence in visual cue integration. In *Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling*, S.-C. Zhu, Ed.

SCHRATER, P. R., AND KERSTEN, D. 2000. How optimal depth cue integration depends on the task. *International Journal of Computer Vision* 40, 1, 73–91.

SERVOS, P., AND GOODALE, M. A. 1998. Monocular and binocular control of human interceptive movements. *Experimental Brain Research* 119, 1, 92–102.

WEISS, Y., SIMONCELLI, E. P., AND ADELSON, E. H. 2002. Motion illusions as optimal percepts. *Nature Neuroscience* 5, 6, 598–604.