Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 45 (2005) 3050–3059
Vision Research
www.elsevier.com/locate/visres

Explicit estimation of visual uncertainty in human motion processing

Erich W. Graf [a,*], Paul A. Warren [b], Laurence T. Maloney [c]

[a] School of Psychology, University of Southampton, UK
[b] School of Psychology, Cardiff University, UK
[c] Department of Psychology, Center for Neural Science, New York University, USA

Received 22 September 2004; received in revised form 4 August 2005

Abstract

We examine whether human observers have explicit access to an estimate of their own uncertainty in extrapolating the motion trajectories of moving objects. Objects moved across a display area at constant speed changing direction at short time intervals. Each new direction was obtained by adding a random perturbation to the previous direction. The perturbation distribution was always symmetric with mean zero (no change in direction) but could differ in variability: objects with low directional variability tended to travel in straight lines while objects with high directional variability moved more erratically. Objects eventually disappeared behind the near edge of an occluder. Observers marked a 'capture region' along the far edge of the occluder that they estimated would contain the object when it re-emerged. We varied both occluder width and directional variability across trials and found that observers correctly compensated for these changes. We present a two-stage model of observer performance in which the visual system first estimates the directional variability of the object and then uses this estimate to set a capture region.

Keywords: Cue combination; Statistical approaches; Motion; Extrapolation

1. Introduction

Real objects can move with very different path characteristics. A hummingbird traveling from one point to another is unlikely to follow the same trajectory as a badminton shuttlecock. The movement of each is effectively stochastic, varying unpredictably from instant to instant, but the stochastic component of motion of the shuttlecock is slight; the hummingbird, in contrast, varies its velocity frequently and its future position is less predictable.

Consider an observer trying to predict exactly where either the hummingbird or the shuttlecock will re-emerge immediately after it passes behind an occluder (Fig. 1). The data available are estimates of position and location sampled at points along the previous, visible part of the trajectory. This sort of statistical task is an example of a *point estimation problem* (Mood, Graybill, & Boes, 1974).

In tasks of this sort, human observers are generally quite adept. Trajectory estimation has been demonstrated early in development, with six-month-old infants able to perceive the continuity of object trajectories, even after brief periods of occlusion (Johnson et al., 2003). In addition, adult observers have been shown to accurately track objects through visual noise (Verghese & McKee, 2002; Watamaniuk & McKee, 1995), as well as multiple objects passing in and out of occlusion (Scholl & Pylyshyn, 1999). Furthermore, the smooth pursuit eye movement system has been shown to integrate occlusion information. If observers direct their attention to 'moving' an imagined target (Pola & Wyatt, 1997), or if they expect the target to reappear (Becker & Fuchs, 1985), smooth pursuit continues, but with

* Corresponding author. Tel.: +44 02380594587; fax: +44 02380594597.
E-mail address: erich@soton.ac.uk (E.W. Graf).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.08.007

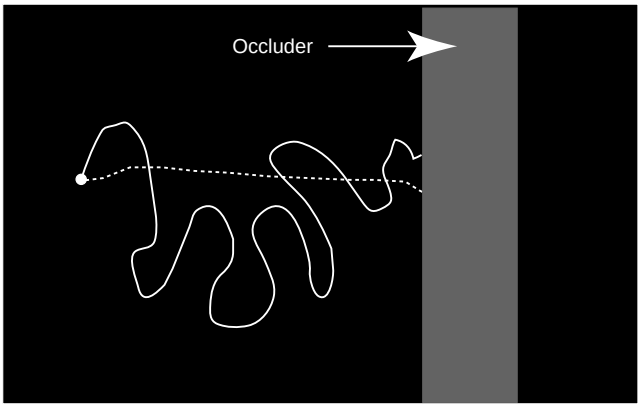

Fig. 1. The random walk stimulus. Two sample paths are shown: one that is 'erratic' (large directional variability—solid line) and one more or less straight path (small directional variability—dashed line).

reduced gain. Eye velocity then increases in expectation of target reappearance (Bennett & Barnes, 2003).

Observers also interpolate static contours (a point estimation task related to path extrapolation) with high accuracy (Warren, Maloney, & Landy, 2002, 2004). Interpolation accuracy for 3D localization of a point in the invisible region of an occluded contour is on the order of seconds of arc. A recent study (Singh & Fulvio, 2005) has demonstrated that observers also reliably extrapolate contours.

In this article, however, we are not concerned with point estimation. Paired with every point estimation problem is an *interval estimation problem* (Mood et al., 1974) whose solution is a measure of the uncertainty of the point estimate. This measure could be as simple as the standard error of the point estimate or as complex as a full specification of its distribution. We may predict the emergence points of the hummingbird and of the shuttlecock optimally by whatever criterion we choose, but we are likely to have much less confidence in our prediction for the hummingbird. Here, we examine not how well human observers extrapolate trajectories, but whether human observers can explicitly and consistently estimate their own extrapolation uncertainty; whether they know precisely how much harder it is to intercept a hummingbird than a shuttlecock.

This is an important issue when viewed within the context of models of visual cue combination, which we now briefly review. Many psychophysical tasks involve estimation of the properties of an object such as its location or its velocity. In complex, everyday scenes, there are often multiple sources of visual information (or *cues*) relevant to estimation of the property. Previous work in the cue combination literature has concentrated on the *homogeneous cue* case where there are n cues represented by independent random variables $V_1, V_1, \ldots, V_n$, each of which is itself an estimate of the unknown object property, v, perturbed by additive Gaussian noise with mean 0 and variance $\sigma_i^2 > 0$ (see Landy, Maloney, Johnston, & Young, 1995). Each of these cues is unbiased (its expected value is v) and, if we had to choose to use only one of these cues, then it would be reasonable to select the cue with smallest variance. However, if we seek to minimize the variance of the resulting estimate, $r_i = \sigma_i^{-2}$, we can do better by combining the cues as follows. We define the *reliability*[1] of a cue as the reciprocal of its variance, $r_i = \sigma_i^{-2}$, and form the weighted estimate,

$$V = \sum_{i=1}^{n} w_i V_i, \tag{1}$$

where

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j}. \tag{2}$$

Then the combined estimate is also unbiased and it has reliability,

$$r = \sum_{j=1}^{n} r_j. \tag{3}$$

It is evident that the reliability of the combined estimate is as great as or greater than the reliability of any single cue V_i (Cochran, 1937; Ernst & Bülthoff, 2004; Oruç, Maloney, & Landy, 2003). Moreover, the reliability of the combined estimate V is the highest possible for any cue combination rule that gives an unbiased estimate (Cochran, 1937; Ernst & Bülthoff, 2004; Oruç et al., 2003). There is considerable experimental evidence showing that, in the homogeneous cue case, observers do combine cues so as to achieve reliabilities greater than they could achieve by using only a single one of the cues (Ernst & Banks, 2002; Landy & Kojima, 2001; Oruç et al., 2003).

However, to combine cues optimally, the visual system requires an estimate of the reliability (i.e., variance) of each cue (Landy et al., 1995; see also Jacobs, 2002). Each cue effectively enters into the computation as an ordered pair (V_i, r_i). The outcome of the cue combination computation above is potentially *two* numbers, (V, r), and it is certainly of interest to ask whether the observer can make use of the final estimate of reliability, r, as readily as he or she makes use of the estimate V, in psychophysical tasks.

There is indirect evidence that reliability estimates affect behavior. Trommershäuser and colleagues examined how subjects plan rapid movements in environments where there are explicit monetary penalties associated with the outcomes of movements. They found that observers correctly take into account their own movement variability in planning their movements (Trommershäuser, Maloney, & Landy, 2003a, 2003b). However, observers were not directly asked about visual

[1] Terminology introduced by Backus and Banks (1999). A more reliable cue has a lower variance and vice versa.

reliability and the conclusion that they had implicit access to an accurate estimate of their own visual reliability was inferred from their overall good performance. In addition, the use of uncertainty information has been investigated in cue integration tasks (Gharamani, Wolpert, & Jordan, 1997; Jacobs, 1999; Knill & Saunders, 2003; Körding & Wolpert, 2004) and motor planning and control (Todorov & Jordan, 2002; van Beers, Baraduc, & Wolpert, 2002; van Beers, Sittig, & Denier van der Gon, 1999).

In this study, we use a motion extrapolation task to investigate uncertainty estimation directly. The initial visual information available to the observer comprises samples of location and velocity along the visible part of the trajectory and the rule of combination may be more complex than a simple weighted average. Regardless of the exact formulation of the rule of combination, we are interested in whether human observers can explicitly estimate the uncertainty corresponding to their visual estimates of the extrapolated motion.

2. General methods

2.1. Apparatus

Stimuli were presented on a flat-screen Apple Studio Display monitor (1280×1024 pixels; 33.8×27 cm). Observers sat approximately 60 cm from the monitor and viewed the display monocularly, wearing an eye patch over their left eye. The center of the screen was approximately at eye level.

2.2. Stimuli

In the first experiment, the visual stimulus consisted of a single dot (3 pixels × 3 pixels), moving from left to right across the screen toward a gray vertical strip (the occluder) that spanned the height of the screen (see Fig. 1). The dot originated at a fixed point halfway up the screen and slightly displaced from the left edge. The horizontal distance from the initial point to the occluder was 250 pixels (6.3 cm). The dot moved at a constant speed in a direction that changed at discrete temporal intervals (~5/s) during movement. To describe the motion, we denote the direction of movement on time step i by the angle θ_i, $i = 1, 2, 3, \ldots$ The first time step of the motion sequence was horizontal and to the right $\theta_0 = 0°$. On each subsequent time step, we added a random directional perturbation Δ_i: $\theta_{i+1} = \theta_i + \Delta_i$. The Δ_i, $i = 1, 2, 3, \ldots$ were independent, identically distributed random variables drawn from a von Mises distribution (described below). At discrete temporal interval i the dot was displaced in direction θ_i by an amount proportional to velocity v. The generation process is illustrated in Fig. 2. The random walk process is Markovian: at each time step, the

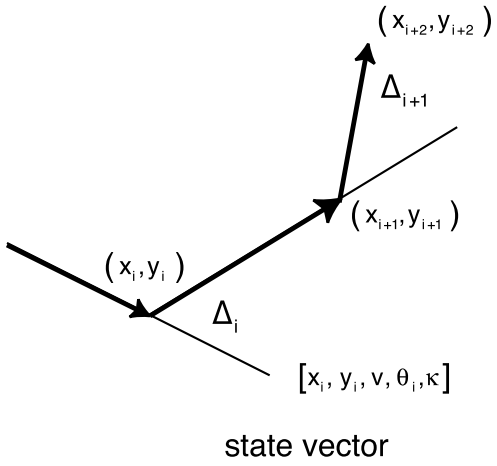

Fig. 2. Generation of the random walk. At each point in time, the dot has a state vector $[x_i, y_i, v, \theta_i, \kappa]$. The dot first moves forward a fixed distance determined by its speed v. The direction θ_i is then randomly perturbed by adding a von Mises random variable Δ_i with mean and directional reliability κ. The resulting random walk is isotropic in the sense that, if we replace the initial orientation $\theta_0 = 0°$ by a non-zero value θ then the distribution of possible future evolutions of the random walk is simply rotated by θ. In both experiments, the speed remains constant across the duration of the random walk: the size of each step is identical.

state vector $[x_i, y_i, \theta_i, v, \kappa]$ captures all of the information about the past movement of the point. The probability that the point will be at any location at any specified future time is determined by this state vector and the future trajectory of the point is otherwise independent of the past.

The von Mises distribution, frequently used as the circular analogue of the Gaussian distribution, is a symmetric unimodal distribution on the circle that has two parameters, $\mu \in (-\pi, \pi)$ and $\kappa \geqslant 0$. Its probability density function (for angular direction d) is given by

$$f_{\mathrm{vM}}(d) = \frac{1}{2\pi I_0(\kappa)} \mathrm{e}^{-\kappa \cos(d-\mu)}, \tag{4}$$

where $I_0(\kappa)$ is the 0th order modified Bessel function of the first kind (Evans, Hastings, & Peacock, 2000, Chapter 41). In sampling Δ_i, we set the location parameter $\mu = 0$ so the dot is as likely to turn clockwise as counterclockwise at each time step. The parameter κ is a shape parameter. When κ is large, the von Mises distribution resembles a Gaussian centered on μ. As κ approaches 0, the von Mises distribution approaches the uniform distribution on the circle. The variance of the von Mises distribution varies inversely with κ and, when κ is large, the variance of the von Mises distribution is approximately proportional to $1/\kappa$. Accordingly, we will refer to as the *directional reliability* of the motion path. It is the independent variable of main interest in Section 3. On trials where the directional reliability κ was small, the dot would typically take a more erratic path than on trials where it was larger.

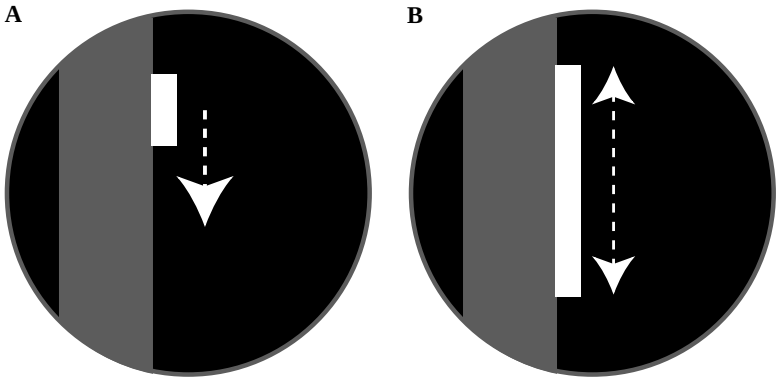

Fig. 3. The observer's tasks. (A) After the dot disappears behind the occluder, observers were asked to move a marker up and down along the edge of the occluder to indicate a best estimate of where the dot would emerge (indicated by dashed line and arrow). (B) They were then asked to select a 'capture region' around that estimate that would be sure to contain the dot.

2.3. Task

Observers watched the dot travel toward the occluder. When the dot reached the occluder, it disappeared. Observers were then asked to make two judgments: (1) a point estimate of where the dot would emerge on the other side of the occluder (Fig. 3A), and (2) the smallest 'capture region' around that estimate (interval estimate) that would capture the dot (Fig. 3B). The width of the capture region, W_O, is the dependent variable of greatest interest in both experiments.

A word of explanation is in order concerning the instructions to the observers. We initially considered asking them to set confidence intervals that would capture a fixed proportion of trajectories of the re-emergence of the moving point. However, there is convincing experimental evidence that observers cannot reliably set confidence intervals at specified levels α (Plous, 1993)[2] and, consequently, we were concerned that naïve observers may have difficulty understanding what is meant by a 50 or 95% confidence interval. By instructing observers to set the smallest 'capture region' that would 'reliably' contain the dot, we are anticipating what they are likely to do in any case. Given our instructions they were free to choose a criterion level. In our analyses, we assumed only that the criterion level remained fixed across the interleaved conditions of each experiment. We tested whether the width of their chosen capture regions scaled correctly with experimental condition. That is, while there is no 'right' choice of capture region in any single condition of the experiment, we test whether the observer's choices across conditions are *consistent*.

3. Experiment 1

3.1. Methods

In the first experiment, we varied three factors: (1) the directional reliability κ of the dot path, (2) the width of the occluder, and (3) the speed of the dot. We used five directional reliability levels κ (50, 125, 200, 400, and 650), three occluder widths w (75, 105, and 135 pixels) and two dot speeds s (75 and 115 pixels/s). Each of the resulting 30 ($5 \times 3 \times 2$) conditions was presented twice per session. Observers participated in 3 sessions, with each session lasting approximately 10 min. Five observers (two authors, three naïve) participated.

3.2. Results

Observers reported that they found the task to be easy and natural. The results of the first experiment are illustrated in Fig. 4. Results are shown for one naïve observer (RAC) in Fig. 4A, and for all observers in Fig. 4B. The observers' capture region can be seen to vary both with the reliability of the path and with the width of the occluder. Observers' capture regions were larger on trials where paths were more erratic and on trials where the occluder was wider. A three-factor ANOVA with occluder width, path reliability and speed as factors demonstrated a significant main effect of occluder width and path reliability, but not for speed ($p > 0.01$). Speed trials were combined for the present analysis. A separate analysis concluded that there was not a significant difference between the data of the authors and the naïve observers. There may be a concern that the data were the result of a higher-level cognitive assessment of the underlying direction

[2] Plous (1993) asked subjects to set 90% confidence intervals for the date of specific historical events and found that subjects tended to set intervals that always contained the specified date, i.e., 100% confidence intervals. It is unlikely that observers in our experiment would do the same as a 100% confidence interval would correspond to the entire right edge of the occluder in every condition. Observers, in fact, did not do so.

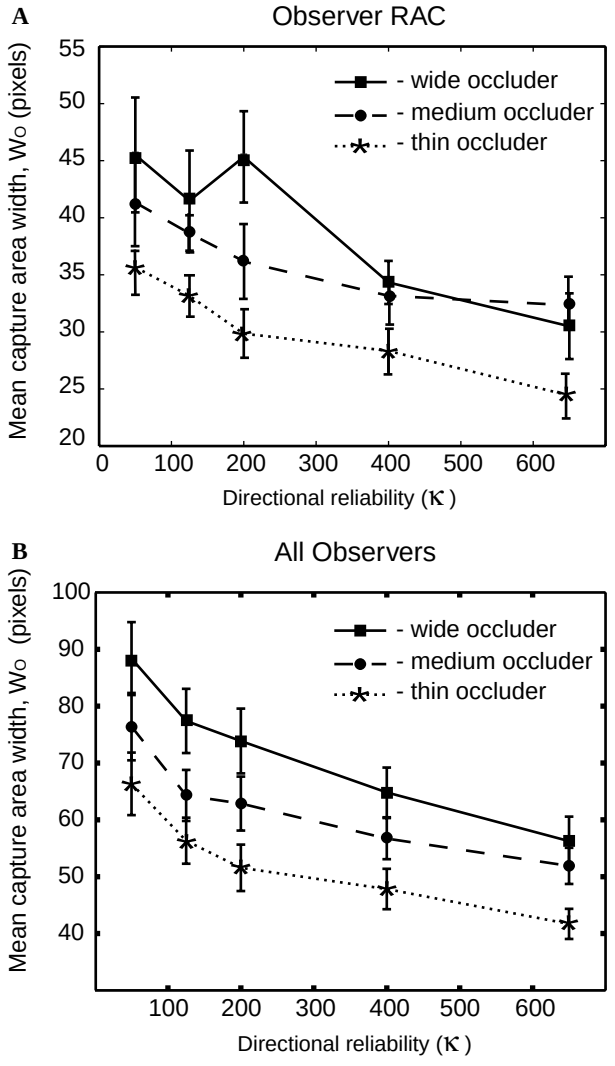

Fig. 4. Experiment 1: Capture region widths. (A) The mean capture region widths, W_O, in pixels, as a function of path reliability, κ are plotted for one naïve observer (RAC), and (B) for all observers. The different curves represent data collected at different occluder widths.

distribution as opposed to uncertainty in a perceptual estimate. Since none of our observers (authors included) knew how the variability of the motion path scaled with occluder width or the directional reliability parameter κ, implementation of such a cognitive strategy would seem unlikely.

Directional reliability κ was the variable of interest, and we found that human performance changed predictably as a function of κ. We next used Monte Carlo simulations to determine whether a representation of κ was sufficient to predict our observers' data. For each experimental condition, we simulated 100 dot trajectories (T_k, $k = 1,2,\ldots,100$) up to the point where the motion path disappeared under the occluder; at a particular y-coordinate y_o^k traveling in a specific direction θ_o^k. The future movement of the point under the occluder is determined by the state vector $[x_o, y_o^k, \theta_o^k, v, \kappa]$, where x_o is the x-coordinate of the occluder (fixed throughout the experiment) and v and κ are determined by the condition being simulated. For the kth trial, we simulated the continuation of the motion path from $[x_o, y_o^k, \theta_o^k, v, \kappa]$ under the occluder until the point re-emerged on the other side of the occluder and recorded the y-coordinate of the point of emergence y_e^k. We repeated the simulation of dot continuation under the occluder 1000 times for each of the 100 simulated dot trajectories. We evaluated the distribution of emergence points by computing the standard deviation W_S^k for the distribution. In the simulations, we discarded a small proportion of trials (less than 2%) in which the point had not emerged within 1400 time steps. We took this W_S^k (or any multiple of it) as an estimate of a veridical capture region width for the specific trial T_k.

We considered computing W_S^k using the actual value of the state vector $[x_o, y_o^k, \theta_o^k, v, \kappa]$ for each experimental trial that the observer saw and comparing it to the observer's capture width region for that trial. We did not do so for two reasons. First, the value W_S^k depends upon knowledge of the angle of arrival θ_o^k of the point at the occluder edge. We do not know how accurately an observer can estimate this angle given that it changes rapidly and the size of each movement step was small. Second, we wished to base our estimates of the veridical capture region width on a larger number of trials in each condition than we could ask observers to perform. Instead, we averaged the values $W_S^k, k = 1,2,3\ldots,100$ across the trials for a specific condition to obtain an aggregate value W_S that we compared to the corresponding mean of the observers' capture width settings for the same condition.

To compare these standard deviations with our human observer data, we regressed the mean width, W_O, of the observers' capture region settings for each condition against W_S and the occluder width w for that condition. If the observer were perfectly compensating for changes in the experimental conditions we would expect to find

$$W_O = aW_S + bw + c, \tag{5}$$

where w denotes the width of the occluder. In this equation, the coefficient a corresponds to the observer's choice of a criterion size (in SD units) of the capture area. The coefficients b and c correspond to the residual visual uncertainty that would remain even if the object traveled in a straight line to the edge of the occluder with no added directional uncertainty ($\kappa = \infty$). The inclusion of these terms is motivated by previous psychophysical results. In extrapolation, positional uncertainty (SD) is known to grow linearly with separation (Pavel, Cunningham, & Stone, 1992; Singh & Fulvio, 2005) and consequently we expect a linear growth in visual uncertainty with occluder width even when the directional uncertainty of the path and, consequently, W_S are 0.

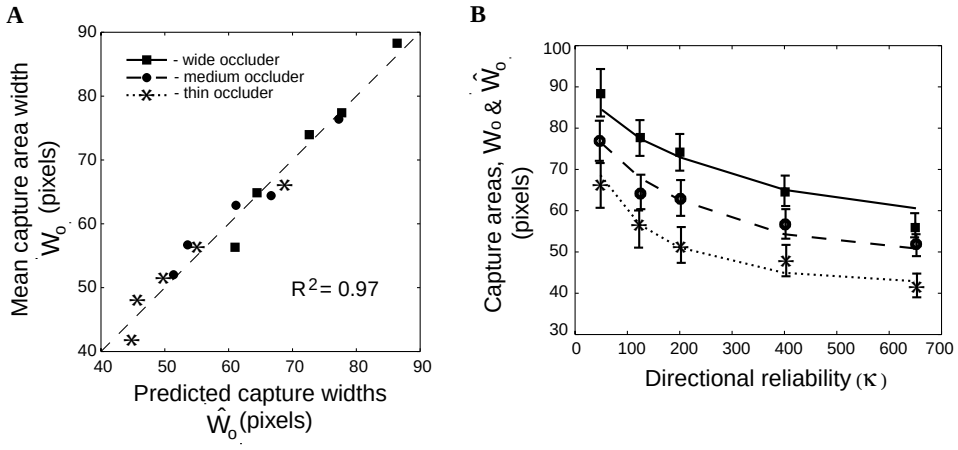

Fig. 5. Experiment 1: Consistency of settings. (A) The mean capture width regions are plotted against the capture width regions predicted by Eq. (5). Different symbols are used for different occluder widths (see legend). The data fall very close to the dashed identity line ($R^2 = 0.97$). (B) The mean capture width regions W_O are plotted versus κ. The same choices of symbols are used for different occluder widths as in Fig. 5A. The superimposed curves connect the predicted values of $\hat{W}_O$ plotted versus κ (see legend).

We fit the data to Eq. (5) by multiple regression with the result,

$$\hat{W}_O = 0.42W_S + 0.13w + 27.0. \quad (6)$$

Each of the three coefficients is significantly different from 0 at the 0.05 level. The data W_O are plotted versus $\hat{W}_O$ in Fig. 5A with different symbols for the three different occluder widths. The line fits the data very well ($R^2 = 0.97$). We used the regression equation to transform our human data (Fig. 5B, symbols) and simulated data (Fig. 5B, lines) into common units. Comparing the symbols and the lines, it can be seen that the simulation data closely follow the human data, indicating that the human observer data can be predicted simply by recovering an estimate of direction variance during the motion sequence. Also, W_O is approximately linear with respect to $k^{-1/2}$ which makes intuitive sense if k^{-1} is considered an approximation of the directional variance of the path. Overall, these data indicate that human observers correctly take path characteristics (notably directional reliability) and information about the scene (the occluder width) into account when extrapolating visual motion trajectories.

4. Experiment 2

In a second experiment, we sought to determine which parts of the visible path the observer used in estimating capture regions by changing the reliability of the path within a trial. Observers might rely heavily on the contribution of the most recent part of the motion path when making their final judgments, or alternatively they could take into account information along the entire visual motion sequence. To investigate this, we conducted an experiment in which the directional reliability κ of the von Mises distribution was changed from low to high or vice versa at a specified point in the motion. The observer would see a motion path that varied erratically at first and then became more reliable or vice versa.

4.1. Methods

The change point could be at any of seven equispaced horizontal positions between the starting point and the occluder. Once the x-coordinate of the dot crossed the change point the value of changed either from low to high or high to low. The low κ value was set at 50 and the high κ value at 200. Speed was held constant at 115 pixels/s, and the occluder width was fixed at 115 pixels. Observers repeated each low/high and high/low condition 5 times, completing the data collection in one session lasting approximately 10 min. The same five observers (two authors, three naïve) participated.

4.2. Results

Fig. 6 shows the results of Experiment 2 for one naïve observer, RAC (Fig. 6A), and mean data for all observers (Fig. 6B). The points connected by solid lines form a plot of mean capture region width versus the point in the trial where κ went from high (reliable) to low (erratic). The earlier part of the trajectory is less erratic and the later the transition, the greater the proportion of the motion path with high directional reliability. In Fig. 6B, the mean capture region widths connected by solid lines decrease with increasing changeover point, as expected.

The points connected by dashed lines represent the trials in which κ went from low to high: the later part

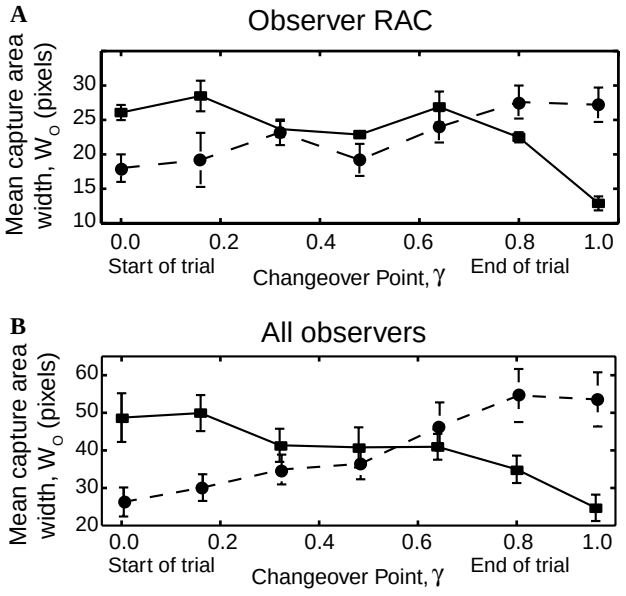

Fig. 6. Experiment 2: Results. Results are shown for one naïve observer (RAC), (A) and averaged across all observers, (B) plotted are subject capture region widths,, in pixels, as a function of the proportion of the path that had high (low) directional reliability during a given trial. The points connected by a solid curve correspond to trials where directional reliability κ went from high to low (the latter part of the motion trajectory is more erratic) and the points connected by a dashed curve correspond to trials where directional reliability κ went from low to high.

of the trajectory is now less erratic and the later the transition the greater the proportion of the motion path with low directional reliability. As expected, the data points connected by the dashed lines have an upward trend.

Suppose that observers used only the later part of the motion path to estimate directional reliability. Then, we would expect both of the curves in Fig. 6B to be horizontal over the early part of the motion path, up to the point where the observer began to accumulate information about directional reliability.

On the contrary, the smooth upward and downward trends in Fig. 6B indicate that observers are using information about the object path across the whole trial. Observers tended to set larger capture regions as the proportion of the whole motion path for which κ was high increased. This was true for the case when κ transitioned from low to high and vice versa. The apparent symmetry and rough linearity of the two curves suggests that observers are giving roughly equal weight to information about directional uncertainty at every point along the motion path.

The results of Experiment 2 confirm that human observers do have a representation of the uncertainty of their positional estimate and serve to demonstrate that they accumulate this information over long temporal intervals (our trials lasted up to 10 s). An estimate of visual motion uncertainty is a key component in the Monte Carlo simulations that predict our human observers' performance in a task involving occlusion. Can this estimate be obtained from what has been previously hypothesized about the organization of human motion processing?

Since their introduction by Adelson and Bergen (1985), motion-energy models have been used to account for human performance in a great number of motion perception tasks. This class of models uses physiologically plausible motion detectors as basis units for extracting spatiotemporal energy from a visual motion sequence. In relation to our experimental results, it is plausible that estimates of directional variance for increases in directional variability (k) could be derived from collections of motion-energy detectors. To account for our results using a motion-energy model, two important components are required. The first is that information can be accumulated over long temporal periods. The second is that the output of the detector can be used to determine variability during the accumulated motion sequence. In an Appendix A, we illustrate how an existing model of motion perception (Weiss, Simoncelli, & Adelson, 2002) may be extended and used to recover an estimate of the uncertainty of visual motion, a quantity which we have shown to be sufficient to predict observers' extrapolation data.

5. Discussion

Human observers do have access to a representation of uncertainty in a visual stimulus, and they accumulate this estimate of uncertainty over a remarkably long temporal interval. The implications of these results are discussed in the following paragraphs.

Recent studies in physiology have explored the representation of decision information in the brain (Barborica & Ferrera, 2003; Eskandar & Assad, 1999; Kim & Shadlen, 1999). Motion extrapolation tasks similar to the one used in the present study are used to disassociate sensory areas and motor planning areas, making use of certain fundamental aspects of motion processing. Visual areas MT and MST contain direction-selective neurons that represent motion information (Britten, Shadlen, Newsome, & Movshon, 1992; Newsome & Pare, 1988). When a visual motion stimulus is extinguished, neuron firing in MT and MST ceases (Seidemann, Zohary, & Newsome, 1998). However, in an extrapolation task, neurons associated with path prediction and subsequent decision-making must maintain their activity beyond this point. Neurons have been found in posterior parietal area LIP that seem to encode such information (Eskandar & Assad, 1999), and frontal eye field (FEF) neurons have been shown to emphasize the *storage* of speed information (Barborica & Ferrera, 2003). Thus, neurophysiological evidence points towards the conclusion that information about motion is stored for use in subsequent tasks and judgments.

Physiology studies to date have been interested mainly in positional information (i.e., what point will the monkey move its eyes to in order to intercept the target). The present work is consistent with and complementary to these studies. Here, we have shown that observers accumulate visual information about movement *uncertainty* over time and can use that information in a subsequent task. We emphasize that this is not solely information about mean speed or mean direction, but distinct, explicit information about uncertainty, a topic not previously addressed. Using a physiologically plausible model we have shown that human performance can use low-level motion-energy detectors to inform a decision task regarding the uncertainty of a visual estimate. The model is built on fundamental principles of motion integration; only requiring that the visual system make some measurement of dispersion of the accumulated velocity representation.

The results presented here, together with other work, support an emerging model of human visual processing as a statistical engine (Kersten, Mamassian, & Yuille, 2004; Kersten & Schrater, 2002; Knill & Richards, 1996; Landy et al., 1995; Mamassian, Landy, & Maloney, 2002; Maloney, 2002; Yuille & Bülthoff, 1996). Previous work in psychophysics has concentrated on the central tendency component of object property estimation, but for many tasks it is not enough to know the best point estimate and an estimate of variability is required. The results here suggest that human observers have the same access to this information as to standard point estimates that have thus far been the main focus of psychophysics. The visual system allows the observer not only to estimate where the hummingbird or shuttlecock will emerge from occlusion, but also to know whether there is any point in trying to catch it.

Acknowledgments

E.W.G. was supported by the National Science Foundation (USA), Fellowship 0107383. P.A.W. was supported by EPSRC Grant GR/R57157/01 and LTM was supported by NIH Grant EY08266 and HFSP Grant RG0109/1999-B.

Appendix A. Variability estimation and motion models

To illustrate how motion-energy models might acquire variability information, we extend an existing Bayesian model of motion perception (Weiss et al., 2002). Weiss et al. developed a Bayesian method to estimate a motion velocity v_p at each point p in a time-varying image $I(x,y,t)$ and used it to explain a variety of motion illusions. They first chose a likelihood function of the form

$$P\left[I(x_p,y_p,t) \mid v_p\right] \propto \exp\left(-\frac{1}{2\sigma^2}\int w_p(x,y)\left(I_x(x,y,t)v_x + I_y(x,y,t)v_y + I_t(x,y,t)\right)^2 \mathrm{d}x\,\mathrm{d}y\right),$$

where v_x and v_y are the x and y components of velocity, I_x, I_y, and I_t are the partial derivatives of the image $I(x,y,t)$ with respect to x, y, and t, respectively, and $w_p(x,y)$ is a motion detector centered on p. Assuming a prior on velocity v that favors slower speeds

$$P(v) \propto \exp(-\|v\|^2/2\sigma_p^2)$$

the resulting posterior distribution is the product of likelihood and prior,

$$P[v_p|I(x,y,t)] \propto P[I(x,y,t)|v_p]P[v_p].$$

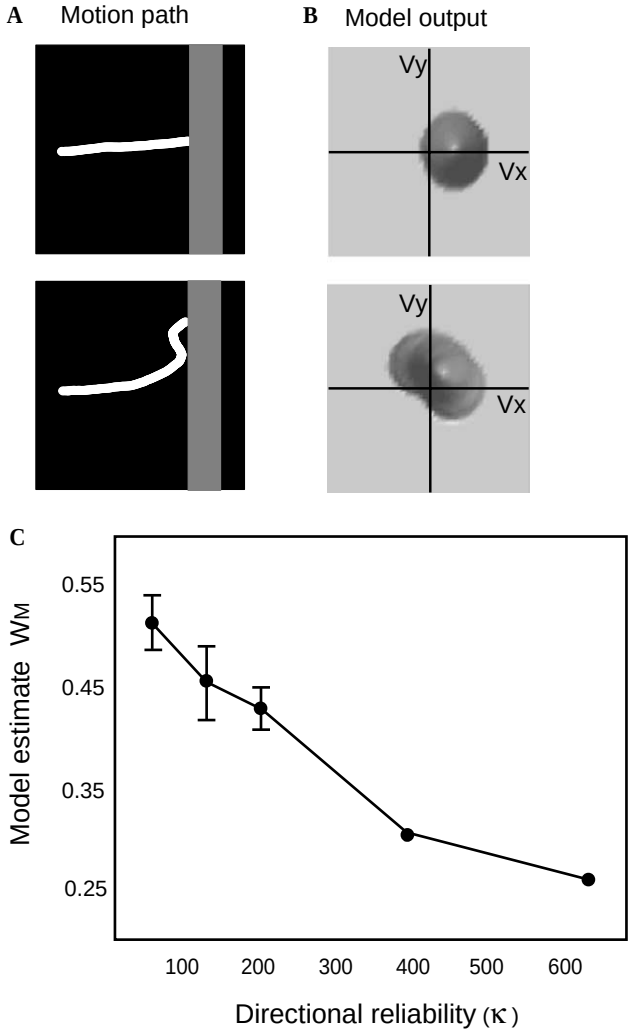

Fig. 7. Bayesian model. Two examples of trajectories (A) and the corresponding Bayesian model output (B). The model output is represented in velocity space with on the horizontal axis and on the vertical. The more erratic paths result in output where the summed posterior distribution has a larger angular spread. (C) Summary of model predictions W_M versus directional reliability κ.

The velocity at which the posterior distribution achieves its maximum is the MAP (maximum a posteriori) point estimate of velocity.

We used the Weiss et al. model to compute a MAP estimate of the velocity vector at each point along the trajectory of the dot before it reached the occluder. We placed motion-energy detectors at every point in the image and computed the posterior probability distribution for each temporal interval $P_\tau[v|I]$, $\tau = 1, 2, \ldots, T$ across the duration of the trial (until the dot reached the occluder). Following Weiss et al., we computed the MAP estimate of velocity $\hat{v}^\tau = (\hat{v}_x^\tau, \hat{v}_y^\tau)$ at each time step along the trajectory and then computed the angle $\hat{\phi}_\tau$ subtended by the horizontal axis and the MAP velocity vector. The distributions of the estimated velocity vectors are shown in Fig. 7 for two paths, one with lower variability, one with higher. We computed the circular standard deviation W_M of the angles $\hat{\phi}_\tau$, $\tau = 1, \ldots, T$ (Mardia & Jupp, 1999) and used it as our estimate of directional uncertainty derived from the model of Weiss et al. The circular standard deviation is a number between 0 and 1 that measures the spread of a collection of unit vectors or angles.

Fig. 7C is a summary plot of the model estimate W_M versus the directional reliability κ. The latter is evidently a 1-1 function of the former and an estimate of W_M is equivalent to an estimate of κ that is easily computed from low-level energy detectors. A regression of W_M on κ in log–log coordinates showed close correspondence between the output of our motion model and the uncertainty parameter of interest ($R^2 = 0.95$).

References

Adelson, E., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, 2*(2), 284–299.

Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception, 28*, 217–242.

Barborica, A., & Ferrera, V. P. (2003). Estimating invisible target speed from neuronal activity in monkey frontal eye field. *Nature Neuroscience, 6*(1), 66–74.

Becker, W., & Fuchs, A. F. (1985). Prediction in the oculomotor system: Smooth pursuit during transient disappearance of a visual target. *Experimental Brain Research, 57*, 562–575.

Bennett, S. J., & Barnes, G. R. (2003). Human ocular pursuit during the transient disappearance of a visual target. *Journal of Neurophysiology, 90*, 2504–2520.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*, 4745–4765.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, 4*(Suppl.), 102–118.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162–169.

Eskandar, E. N., & Assad, J. A. (1999). Dissociation of visual, motor and predictive signals in parietal cortex during visual guidance. *Nature Neuroscience, 2*(1), 88–93.

Evans, M., Hastings, N., & Peacock, B. (2000). von Mises distribution. In *Statistical distributions* (3rd ed., pp. 189–191). New York: Wiley.

Gharamani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. G. Morasso & V. Sanguineti (Eds.), *Self-organization, computational maps, and motor control*. Amsterdam: Elsevier Press.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39*, 3621–3629.

Jacobs, R. A. (2002). What determines visual cue reliability. *Trends in Cognitive Sciences, 6*(8), 343–350.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Chesire, A. (2003). Infants' perception of object trajectories. *Child Devlopment, 74*(1), 94–108.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271–304.

Kersten, D., & Schrater, P. (2002). Pattern inference theory: A probabilistic approach to vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world*. New York: Wiley.

Kim, J., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neuroscience, 2*(2), 176–185.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge, England: Cambridge University Press.

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

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*, 244–247.

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

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Integration of stereopsis and motion shape cues. *Vision Research, 35*, 389–412.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of perception and brain function* (pp. 13–36). Cambridge, MA: MIT Press.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world* (pp. 145–189). New York: Wiley.

Mardia, K., & Jupp, P. (1999). *Directional statistics*. New York: Wiley.

Mood, A., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Newsome, W., & Pare, E. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience, 8*, 2201–2211.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451–2468.

Pavel, M., Cunningham, H., & Stone, V. (1992). Extrapolation of linear motion. *Vision Research, 32*, 2177–2186.

Pola, J., & Wyatt, H. (1997). Offset dynamics of human smooth pursuit eye movements: Effects of target presence and subject attention. *Vision Research, 37*, 2579–2595.

Plous, S. (1993). *The psychology of judgment and decision making*. Philadelphia: Temple University Press.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*, 259–290.

Seidemann, E., Zohary, E., & Newsome, W. T. (1998). Temporal gating of neural signals during performance of a visual discrimination task. *Nature, 394*, 72–75.

Singh, M., & Fulvio, J. M. (2005). Visual extrapolation of contour geometry. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 939–944.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience, 5*(11), 1226–1235.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003a). Statistical decision theory and tradeoffs in motor response. *Spatial Vision, 16*, 255–275.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003b). Statistical decision theory and rapid, goal-directed movements. *Journal of Optical Society A, 20*, 1419–1433.

van Beers, R. J., Baraduc, P., & Wolpert, D. M. (2002). Role of uncertainty in sensorimotor control. *Philosophical Transactions of the Royal Society of London. Series B, 357*, 1137–1145.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position information: An experimentally supported model. *Journal of Neurophysiology, 81*, 1355–1364.

Verghese, P., & McKee, S. P. (2002). Predicting future motion. *Journal of Vision, 2*, 413–423.

Warren, P. A., Maloney, L. T., & Landy, M. S. (2002). Interpolating sampled contours in 3-D: Analyses of variability and bias. *Vision Research, 42*, 2431–2446.

Warren, P. A., Maloney, L. T., & Landy, M. S. (2004). Influence functions for 3D visual interpolation: A perturbation analysis. *Vision Research, 44*, 815–832.

Watamaniuk, S. N. J., & McKee, S. P. (1995). Seeing motion behind occluders. *Nature, 377*, 729–730.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 123–161). Cambridge, England: Cambridge University Press.