Perception, 2002, volume 31, pages 421 – 434

DOI:10.1068/p3314

# Fast temporal dynamics of visual cue integration

Jochen Triesch
Department of Cognitive Science, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA; e-mail: triesch@ucsd.edu

Dana H Ballard, Robert A Jacobs¶
Department of Computer Science (¶ Center for Visual Science), University of Rochester, PO Box 270226, Rochester, NY 14627, USA
Received 25 April 2001, in revised form 19 September 2001

**Abstract.** The integration of information from different sensors, cues, or modalities lies at the very heart of perception. We are studying adaptive phenomena in visual cue integration. To this end, we have designed a visual tracking task, where subjects track a target object among distractors and try to identify the target after an occlusion. Objects are defined by three different attributes (color, shape, size) which change randomly within a single trial. When the attributes differ in their reliability (two change frequently, one is stable), our results show that subjects dynamically adapt their processing. The results are consistent with the hypothesis that subjects rapidly re-weight the information provided by the different cues by emphasizing the information from the stable cue. This effect seems to be automatic, ie not requiring subjects' awareness of the differential reliabilities of the cues. The hypothesized re-weighting seems to take place in about 1 s. Our results suggest that cue integration can exhibit adaptive phenomena on a very fast time scale. We propose a probabilistic model with temporal dynamics that accounts for the observed effect.

## 1 Introduction

A fundamental question in neuroscience is how the brain integrates information derived from different sensors, cues, and modalities into coherent percepts. Many researchers have looked at this question from behavioral, computational, and neurophysiological viewpoints. Unfortunately, different integration strategies have been observed in different experiments. Examples are weighted averaging (von Holst 1950; Bruno and Cutting 1988) or extensions thereof (Landy et al 1995), multiplicative interactions (Stein and Meredith 1993), Boolean logic (Newman and Hartline 1982), fuzzy logic (Massaro and Friedman 1990), and linear and nonlinear Bayesian inference (Yuille and Bülthoff 1996). It seems that the way in which different cues are integrated depends on the cues involved, the nature of the task, characteristics of the sensory environment, and prior experience and knowledge of the observers. We believe that an important reason for investigators being unable to identify "the cue-integration strategy" is that observers do not use a single, immutable strategy. Rather, they use a collection of context-sensitive strategies that are adaptable in an experience-dependent manner.

Evidence of the adaptability of cue-integration strategies has been mentioned quite early (von Holst 1950), but has recently begun to accumulate. Jacobs and Fine (1999) used a cue-conflict experimental paradigm to show that observers' cue-combination strategies for visual depth are adaptable as a function of training; subjects adjusted their cue-combination rules to use a visual cue more heavily after training in which that cue was informative than after training in which the cue was irrelevant. Moreover, these researchers showed that observers can learn multiple cue-combination rules, and can learn to apply each rule in its appropriate context. Ernst et al (2000) studied the adaptability of observers' cue-integration strategies using a virtual-reality environment that allowed subjects to interact with viewed objects by touching them. They showed that subjects' estimates of visual slant relied more heavily on a visual cue when the cue was consistent with haptic feedback than when it was inconsistent with this feedback. Also using a virtual-reality environment, Atkins et al (2001) showed that observers compare visual and haptic percepts in order to evaluate the relative reliabilities of visual cues, and use these reliabilities in order to determine how to combine cues during three-dimensional visual perception. In addition, observers are able to learn multiple, context-sensitive cue-integration strategies by comparing visual and haptic percepts.

Although it now seems clear that observers' visual cue-integration strategies are adaptable, little is known about this adaptation process. For example, the studies reviewed above demonstrate adaptation on a relatively long time scale (hours or days). They do not address the question whether or not a faster adaptation can influence on-line processing by modifying observers' cue-integration rules on shorter time scales. Triesch and von der Malsburg (2001) recently proposed that the driving force behind these adaptive phenomena in sensory integration could be a mechanism that makes different cues constantly try to agree on a coherent percept while at the same time adapting towards what is being agreed on. They suggest that this scheme might be very useful on shorter time scales too, predicting fast re-weighting of cues based on their mutual agreement.

We studied the fast temporal dynamics of visual cue integration and report here the results of an experiment in which we used a tracking task. Subjects tracked a target object among distractors and identified the target after an occlusion. Objects were defined by three visual attributes (color, shape, and size). In each trial, two of the attributes were unreliable, in the sense that their values changed frequently within a trial, whereas the remaining attribute was reliable, ie its value did not change. The results suggest that subjects rapidly re-weighted the different cues *during the course of each trial* by emphasizing the information provided by the reliable cue and by discounting the information provided by the unreliable cues. Most of this re-weighting took place in about 1 s and, thus, the results show that cue integration can exhibit adaptive phenomena on a very fast time scale. The experimental results are successfully accounted for by a probabilistic model with temporal dynamics for cue weights.

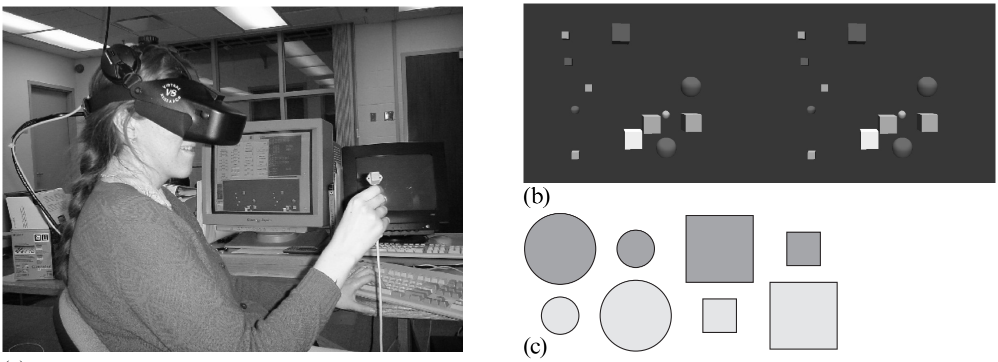
The paper is organized as follows. In section 2 we describe the experimental setup. Our results are presented in section 3. We propose a probabilistic model to account for the effect in section 4. Finally, our results are discussed in a broader context in section 5.

## 2 General methods

### 2.1 *Stimuli and apparatus*

We performed the experiment using a virtual-reality apparatus (Pelz et al 1999). A visual environment was rendered by a Silicon Graphics workstation on a pair of displays embedded in head-mounted goggles (see figure 1). The environment consisted of twelve virtual objects. Objects were defined by shape, color, and size attributes. Each attribute could take one of two possible values. The possible object shape values were sphere and cube, the possible color values were yellow and green, and the possible size values were small (2 cm) and big (3 cm). Subjects freely adjusted their viewing distance to the objects by altering their seating position during the experiment. Usual distance was of the order of 60 cm; thus the average visual angle of small and large objects was 2 and 3 deg, respectively.

Objects moved in a square region of the frontoparallel plane. Each side of the region was 30 cm in length. The initial velocity of an object was selected at random (mean = 24.5 cm $s^{-1}$; for the typical viewing distance this corresponds to an angular velocity of roughly 20 deg $s^{-1}$). Objects bounced when they collided with the sides of the square region or with each other. Because objects exchanged energy when they collided, their velocities changed during the course of a trial but the total energy of the system remained constant.



(a)

**Figure 1.** (a) Subjects view the scene in a head-mounted display. Responses are made by touching an object in virtual reality with a pointing device. (b) Example of the stimulus; the view for the left eye is shown. (c) Objects are defined by the three attributes: color, shape, and size, each of which can take two values; they can be thought of lying in a three-dimensional space of possible objects.

## 2.2 *Procedure*

A subject's task was to visually track one of the objects and to identify it after an artificial occlusion. At the start of each trial, two choices were made. First, one of the twelve objects was randomly selected to be the object to be tracked. This object is referred to as the target; the remaining objects are referred to as distractors. Second, one of the visual attributes was randomly selected to be reliable. For each object, the value of the reliable attribute did not change during the course of a trial. In contrast, the other attributes were unreliable in the sense that their values were modified during a trial. For example, suppose that color was selected as the reliable attribute. The shapes and sizes of the target and distractor objects changed during the trial, but the colors of these objects did not.

A trial consisted of four phases. During the *preparation phase*, the object to be tracked was uniquely defined by its white color. This phase ended after 2 s when the target turned either yellow or green. In the *tracking phase*, all objects randomly changed the values of their unreliable attributes according to a geometric distribution (ie at each time step there was a constant probability of an attribute changing its value). The parameter of the geometric distribution was chosen such that an attribute changed its value within 0.4 s with probability 0.5. The duration of the tracking phase was sampled from a uniform distribution ranging between 2.5 and 4.5 s. If a subject lost track of the target, he or she could abort the trial by pressing a key on the keyboard. After the tracking phase, all objects turned invisible for 0.5 s. This *occlusion phase* simulated a short occlusion of the visual scene during which the objects kept moving. Finally, in the *response phase*, the objects reappeared and moved for another 1 s before they stopped. Subjects then indicated which object they thought was the target by touching one of the twelve objects with a three-dimensional pointing device.

In order to estimate how much the objects' attributes (and their reliabilities) influenced subjects' decisions we used the following manipulation without informing subjects about this: Immediately before the objects reappeared at the end of the occlusion phase, the target and a second randomly chosen object were removed and replaced by two *candidate objects*, which were placed close to where the target had been with similar velocities to what the target velocity had been. The trajectories of the two candidates deviated from the trajectory of the target by the same amount. Note that, for the purpose of tracking the target through an occlusion, subjects could use the trajectory of the object as well as its attributes as cues. By introducing two competing candidates whose trajectories differed from the trajectory of the target by an equal amount we could study the effect of the different attributes of the candidates in isolation, because effects of the trajectory should have canceled out.

The two candidate objects had exactly two attribute values in common with the target at the end of the tracking phase, and one attribute value in common with each other. In other words, both candidates differed from the last appearance of the target in one attribute, but this changed attribute was a different one for the two candidates. For example, if the last attribute values of the target were (sphere, yellow, small), then the values of the candidates may have been (sphere, yellow, big) and (cube, yellow, small). In this example, the subject's response would reflect whether the subject's cue-integration strategy placed greater emphasis on the information provided by the shape cue or the size cue at the end of the trial. If subjects selected neither candidate but a different object, the trial was disregarded, because subjects could have picked this object for a variety of reasons other than the object attributes (subjects lost track of the target, trajectory of the chosen object matched that of the target better than the trajectories of the two candidates) and its choice does not allow clear inferences about the relative weighting of the cues. For a single subject this happened on average in $(16 \pm 5)\%$ of all trials.

Altogether there were nine types of trials: any of the three attributes could be the reliable one, and any of the three attributes could be the one whose value both candidates shared with the target at the end of the tracking phase. The nine types of trials were presented in random order, forming a block. Subjects performed 30 blocks (270 trials). In two-thirds of the trials, both candidates had the same value of an unreliable attribute as each other and as the target, meaning that they differed in the reliable attribute and an unreliable attribute. We call this set of trials *reliable – unreliable trials*. In the remaining trials, both candidates shared the reliable cue with the target, so that they differed in the two unreliable attributes. We call this set of trials *unreliable – unreliable trials*. After the experiment, subjects filled in a questionnaire about the experiment.

### 2.3 *Subjects*

The ten subjects were students at the University of Rochester. They had normal or corrected-to-normal vision. They were naïve to the purposes of the experiment.

## 3 Results

The experiment was designed to evaluate whether or not subjects quickly adapt their cue-integration strategies on the basis of cue reliabilities. Our prediction was that subjects would quickly adapt their strategies so as to emphasize information provided by the reliable attribute and to discount information provided by unreliable attributes. We also predicted that subjects would tend to more equally weight the information provided by reliable and unreliable attributes when the value of an unreliable attribute had not changed for some significant duration of time.

### 3.1 *Reliable – unreliable trials*

In order to test these predictions, we first limit our analysis to the set of reliable – unreliable trials, which comprised two thirds of all trials. It is useful for us to consider this set because it allows for a direct comparison of how much subjects based their decisions on the reliable attribute versus an unreliable attribute. As a matter of terminology, we refer to the unreliable attribute in which the two candidates differed as the "relevant unreliable attribute". In addition, we refer to the candidate with the same value of the reliable attribute as the target as the "reliable-same candidate" and to the other candidate, the one with the same value of the relevant unreliable attribute as the target, as the "unreliable-same candidate". For instance, consider the case where the last attributes of the target were (sphere, red, small) and the reliable attribute was color, meaning that the unreliable attributes were shape and size. If the two candidates were (sphere, red, big) and (sphere, green, small), then size would be the relevant unreliable attribute, the first candidate would be the reliable-same candidate, since it matches the target in the reliable color attribute, and the second candidate would be the unreliable-same candidate, since it matches the target in the relevant unreliable size attribute.

The results of the analysis are shown on the left in figure 2. Shown are the number of trials (over all subjects) in which the subject chose the reliable-same candidate (dark bar) and the unreliable-same candidate (light bar). If subjects' cue-integration strategies were not sensitive to the reliability of cues, then both bars should be about equally high. However, subjects chose the reliable-same candidate much more often. We tested the statistical significance of this result as follows. Our null hypothesis is that cue reliability does not affect subjects' choices but that their decisions are based only on the attributes of the target immediately before it disappeared. Assuming independence of trials, the number of decisions for the reliable-same candidate should be binomially distributed according to $B(N, 0.5)$, where $N$ is the total number of trials considered. The probability of obtaining at least as many decisions for the reliable-same candidate as we observed under the null hypothesis, however, is $p \ll 0.01$. Thus we reject the null hypothesis. Thus, subjects' responses cannot depend only on the target's attributes immediately before the occlusion, but the recent history of attribute changes has to play a role. Our suggestion is that subjects dramatically re-weight different attributes on the basis of an estimate of how reliable they are. Because the tracking phase of a trial lasted from 2.5 to 4.5 s, we can conclude that the temporal dynamics of the adaptation process are fast enough to be important at this time scale.

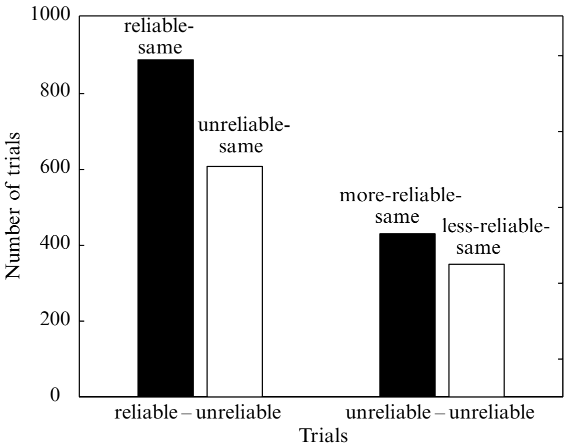



**Figure 2.** Left: Number of decisions for reliable-same candidate and unreliable-same candidate in the reliable – unreliable trials. Right: Number of decisions for more-reliable-same candidate and less-reliable-same candidate in the unreliable – unreliable trials.

In figure 3, the strength of the effect is shown on a subject-by-subject basis. When we perform the statistical analysis for subjects individually, the result is significant at the $p < 0.05$ level for eight of the ten subjects.

We also tested whether the observed effect could be due to a simple perceptual limitation: sometimes the last change of the unreliable attribute might have been so short before the occlusion that subjects did not notice the last change. We limited our analysis to the trials where the last change in the relevant unreliable attribute had been at least 100 ms before the occlusion, giving subjects 100 ms to notice the change. This time has been found to be sufficient for continuous object recognition (Potter 1976) and is well above time scales typically associated with serial visual search (Treisman and Gelade 1980). Again we found a significant trend for subjects choosing the reliable-same candidate ($p \ll 0.01$), suggesting that a perceptual limitation cannot account for the effect.

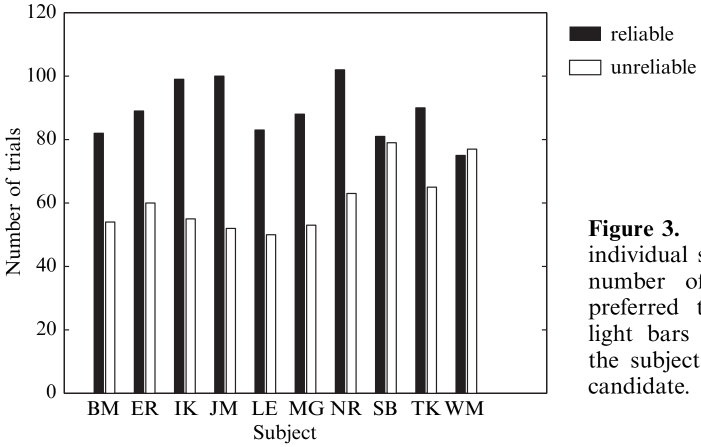



**Figure 3.** Strength of the effect for individual subjects. Dark bars show the number of trials where the subject preferred the reliable-same candidate, light bars the number of trials where the subject favored the unreliable-same candidate.

In order to learn more about the temporal dynamics of the adaptation process, we considered how subjects' responses depended on the amount of time that the relevant unreliable attribute had been unchanged on the target before the simulated occlusion. For example, suppose that on two trials color was the relevant unreliable attribute, and that on the first trial, the color of the target had not changed during the 100 ms prior to the occlusion, whereas on the second trial the color had not changed during the prior 1000 ms. We might reasonably expect that subjects will be more likely to choose the reliable-same candidate on the first trial because the relevant unreliable cue recently changed its value on this trial, whereas on the second trial this cue had been relatively stable. Figure 4 shows that this is indeed the case. The horizontal axis gives the time that the relevant unreliable cue was unchanged prior to the occlusion; the vertical axis gives the average ratio of trials in which a subject chose the reliable-same candidate (the error bars give the standard error of the mean). The results are that subjects tended to emphasize the information provided by the reliable cue, and to discount the information provided by the relevant unreliable cue when the relevant unreliable cue had changed soon before the occlusion. Importantly, when the relevant unreliable cue had not changed for more than approximately 0.3 s before the occlusion, subjects' performances were closer to chance. In this case, they did not seem to strongly distinguish between reliable and unreliable cues once the unreliable cue had been stable for this duration of time.

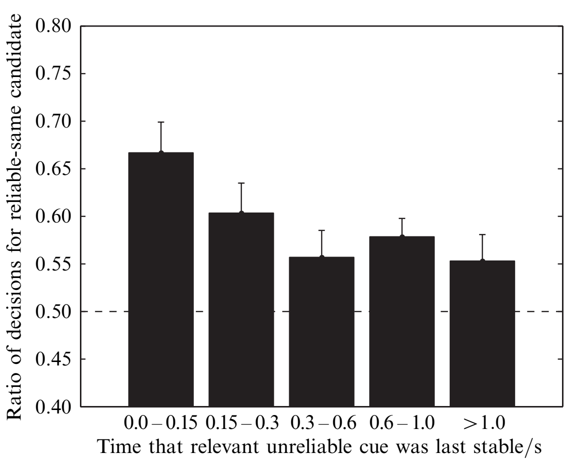



**Figure 4.** Ratio of decisions for the reliable-same candidate as a function of time that the relevant unreliable cue has last been stable.

### 3.2 *Unreliable – unreliable trials*

In the one-third of trials constituting the unreliable – unreliable set, subjects choose between two candidates which differ from the target in a different unreliable attribute. Generally, one of these two attributes has been stable for a longer amount of time. We call these attributes the *more reliable attribute* and *less reliable attribute*, respectively. We call the candidate that matches the target in the more reliable attribute the *more-reliable-same candidate*. The other candidate is the *less-reliable-same candidate*. We tested whether subjects' decisions were favoring either type of candidate. The result is shown in figure 2 right. Subjects prefer the more-reliable-same candidate over the less-reliable-same candidate. We performed the same statistical analysis to test the significance of this finding as described above and found that the probability of finding at least as many decisions for the more-reliable-same candidate under the null hypothesis that subjects do not take cue reliabilities into account is $p \ll 0.01$, leading to rejection of the null hypothesis. This result is consistent with the result from the reliable – unreliable trials and confirms our hypothesis of fast re-weighting of cues based on their reliability.

Again, we confirmed that the effect is not just due to a perceptual limitation by restricting our analysis to the trials where the last change in the less reliable attribute had been at least 100 ms before the occlusion, giving subjects 100 ms to notice the change. Again, we found a significant trend for subjects choosing the more-reliable-same candidate, $p \ll 0.01$.

### 3.3 *Biases for cues*

An interesting aspect of the data is that they reveal that subjects' cue-integration strategies were usually biased toward emphasizing the information provided by a particular cue regardless of the relative reliability of that cue. A subject, for instance, may have entered the experiment with a high relative sensitivity to information provided by the shape cue, and lower sensitivities to information provided by the color and size cues. For each attribute, we counted the number of trials among all the experimental trials in which a subject chose a candidate object that had the same attribute value as the target object at the end of the tracking period. For example, if at the end of the tracking period the target was (sphere, yellow, small) and the subject chose a candidate that was (sphere, yellow, big), then we would increment the count of the shape and color attributes, but not the count of the size attributes. According to this simple measure, seven out of ten subjects tended to emphasize the shape cue; two subjects emphasized the size cue; while one subject emphasized the color cue.

### 3.4 *Conscious strategies*

When each subject was asked at the end of the experiment whether he or she followed a particular strategy, only three subjects mentioned paying attention to the attributes of the target. However, when asked whether subjects noticed that there had always been one attribute that did not change its value during a trial, most subjects claimed to have noticed this in at least some of the trials. When asked whether they noticed anything else, two subjects suggested that the candidate was "split in two" during the occlusion.

## 4 Computational model

In order to better understand the temporal dynamics of the subjects' cue-integration strategies, we developed a simple computational model that successfully accounts for subjects' responses. We assume that subjects are using four cues to identify the target after the occlusion: the trajectory of an object, and its shape, color, and size attributes. The trajectories of the two candidates differ from that of the target by exactly the same amount, so we can assume that information from the trajectory cue will not favor one of the two candidates over the other, but that it provides information distinguishing the candidate from the non-candidate objects. When subjects choose either candidate at the end of the trial, we can conclude that they must have favored this candidate over the other on the basis of the different features. When subjects choose neither candidate but one of the other objects, this may be due to the features and/or trajectory of the chosen object being more similar to those of the original object than those of both candidates. These trials are uninformative with respect to the relative contribution of the features, because the trajectory enters as a confounding factor. Since we do not have a model for how trajectory differences influence subjects' choices, we must restrict our analysis to the trials where subjects choose either candidate and focus on their features. This is not to say that we assume that subjects choose only among the two candidates and not among all twelve objects, but we assume that the probability of a subject favoring one of the two candidates over the other is independent of the other objects present.

Let $\nu$ be a set whose elements are the possible attribute values: sphere, cube, yellow, green, small, and big. In addition, let $\upsilon \in \nu$ denote an attribute value, and $f_\upsilon$ be a binary feature whose value is 1 if a candidate object has attribute value $\upsilon$; otherwise it is 0. For example, a candidate that is (sphere, yellow, small) would be indicated by the following feature values: $f_{\text{sphere}} = 1$, $f_{\text{cube}} = 0$, $f_{\text{yellow}} = 1$, $f_{\text{green}} = 0$, $f_{\text{small}} = 1$, and $f_{\text{big}} = 0$. We denote a particular combination of features by $f^i$, where $f^i$ represents the set of all features $f_\upsilon^i$, $\upsilon \in \nu$. Our model assumes that subjects estimate a probability $\hat{p}(f^i|T)$ that the target will have features $f^i$ after the occlusion. It further assumes that this probability estimate is proportional to a weighted sum of contributions from the different features:

$$\hat{p}(f^i|T) \propto \sum_{\upsilon \in \nu} w_\upsilon f_\upsilon^i \,, \tag{1}$$

where $w_\upsilon$ is the weight of feature $\upsilon$ at the time of the decision. The probability is computed by using the following normalization:

$$\hat{p}(f^i|T) = \sum_{\upsilon \in \nu} w_\upsilon f_\upsilon^i \Big/ \sum_j \sum_{\upsilon' \in \nu} w_{\upsilon'} f_{\upsilon'}^j \,, \tag{2}$$

where the sum over $j$ runs over all possible feature combinations that an object can have. We assume that the weights $w_\upsilon$ are a product of two quantities: (i) a reliability, $r_\upsilon$, that indicates how stable attribute value $\upsilon$ has been with respect to the target, and (ii) a bias, $b_\upsilon$, that indicates the bias in a subject's cue-integration strategy toward emphasizing information provided by attribute value $\upsilon$:

$$w_\upsilon = r_\upsilon b_\upsilon \,. \tag{3}$$

In regard to the temporal dynamics of the reliabilities $r_\upsilon$, we assume that these dynamics are given by a simple leaky integrator:

$$\tau \dot{r}_\upsilon(t) = f_\upsilon(t) - r_\upsilon(t) \,, \tag{4}$$

where $t$ denotes time. In discrete time, this equation may be rewritten as

$$r_\upsilon(t + \Delta t) = \gamma f_\upsilon(t) + (1 - \gamma) r_\upsilon(t) \,, \tag{5}$$

where $\gamma = \Delta t/\tau$. A reliability $r_\upsilon$ is increased at time $t$ if the candidate has attribute value $\upsilon$ at time $t$; otherwise it is decreased. In regard to the biases $b_\upsilon$, we assume that these are constant values for a particular subject. In addition, we assume that the biases for attribute values of the same attribute are equal, and that the biases associated with the different attributes are a set of non-negative numbers that sum to 1. For example, a possible set of biases is $b_{\text{shape}} \equiv b_{\text{cube}} = b_{\text{sphere}} = 0.5$; $b_{\text{color}} \equiv b_{\text{yellow}} = b_{\text{green}} = 0.2$; and $b_{\text{size}} \equiv b_{\text{small}} = b_{\text{big}} = 0.3$.

At the end of the trial the subject decides for one of the twelve objects. The feature combinations of the two candidates are denoted $f^1$ and $f^2$. We assume that the probability of the subject preferring candidate 1 over candidate 2, which we denote by $\hat{p}(1)$, is independent of the other (non-candidate) objects and is given by:

$$\hat{p}(1) = \frac{\hat{p}(f^1|T)}{\hat{p}(f^1|T) + \hat{p}(f^2|T)} \,. \tag{6}$$

To use the model to predict subjects' responses, we exposed the model to the same trials as the subjects. In each trial, the model used equation (5) during the tracking phase to update its estimates of the reliabilities. After the occlusion, the model was exposed to two candidate objects and it computed the probability that each candidate was the target. The model has three free parameters: the time constant $\tau$ and two independent attribute biases (recall that the biases sum to 1). Based on a subject's responses, we performed an exhaustive search of the parameter space to find the parameter values that allowed the model to best fit the subject's responses in a maximum-likelihood sense. Since we are only considering trials where the subject picked either of the two candidates, we can compute the Bernoulli likelihood, $L$, of a subject's decisions given a particular set of model parameters according to:

$$L = \prod_{n=1}^{N} p_n(1)^{R_n} p_n(2)^{1-R_n} \,, \tag{7}$$

where $n$ indexes the trials, $N$ is the number of trials where the subject selected either candidate, $p_n(i)$ is the probability of candidate $i$ being the target on trial $n$ as defined in equation (6), and $R_n$ describes the subject's response according to:

$$R_n = \begin{cases} 1 : \text{subject chose candidate 1} \\ 0 : \text{subject chose candidate 2} \end{cases} . \tag{8}$$

The time constant $\tau$ was evaluated between 0 and 5 s in steps of 0.1 s, and the biases were evaluated between 0 and 1 in steps of 0.1 with the constraint that the three biases sum to 1. For each possible setting of the parameter values, we calculated the likelihood $L$ of a subject's responses. The parameter set which maximizes $L$ is the maximum-likelihood estimate for that subject.

The model estimates the probability that each candidate is the target. In the following, we will refer to the candidate that is assigned a higher probability by the model as the *favored candidate*. Subjects could use the probability estimates in at least two ways to make a decision. First, subjects could always pick the favored candidate, or, second, subjects could perform probability matching, ie pick a candidate probabilistically such that the probability of picking a candidate is proportional to its probability estimate of being the target.

Figure 5 compares model and subject responses. The horizontal axis gives the probability estimate for the favored candidate given by the model. The vertical axis gives the ratio of trials where the subject selected this candidate. For trials where one candidate (the favored one) is clearly preferred by the model, it is very likely to correctly predict the subject's response. The results are in good agreement with the idea that subjects estimate a probability for each candidate as suggested by the model and then perform probability matching to select one of them. The results could also be consistent with the subject always picking the favored candidate if the probability estimates are corrupted by noise. Our experiment does not allow us to distinguish between these hypotheses, although probability matching is a common observation (eg Kowler and Anton 1987).

The time constants $\tau$ that are estimated for different subjects have a median of 0.3 s but show considerable variability; they range from 0.0 – 3.1 s. To gain a better

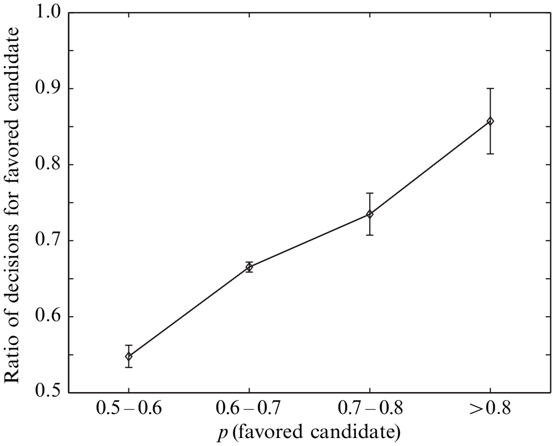



**Figure 5.** Consistency of model with subjects' responses. The ratio of trials where the subject picked the favored candidate is plotted as a function of the probability the model assigned to the favored candidate. For trials where the model strongly favors one candidate, it predicts subjects' responses very well.

understanding why the model estimated different time constants for different subjects, we performed several Monte Carlo simulations using the model. For this purpose, we created large sets of random stimuli and considered the decisions that the model would make for different time constants $\tau$ assuming probability matching. These results are limited to the set of reliable – unreliable trials. They are summarized in figure 6. The horizontal axis gives the time that the relevant unreliable cue was stable prior to the occlusion; the vertical axis gives the average ratio of trials in which the model chose the reliable-same candidate. The results show that, as the value of $\tau$ gets larger, the model is more likely to choose the reliable-same candidate even when the relevant unreliable cue has been unchanged for a significant duration of time. Consequently, the percentage of trials in which the reliable-same candidate is chosen is larger as $\tau$ gets larger. The Monte Carlo results are similar to the experimental results shown in figure 4.

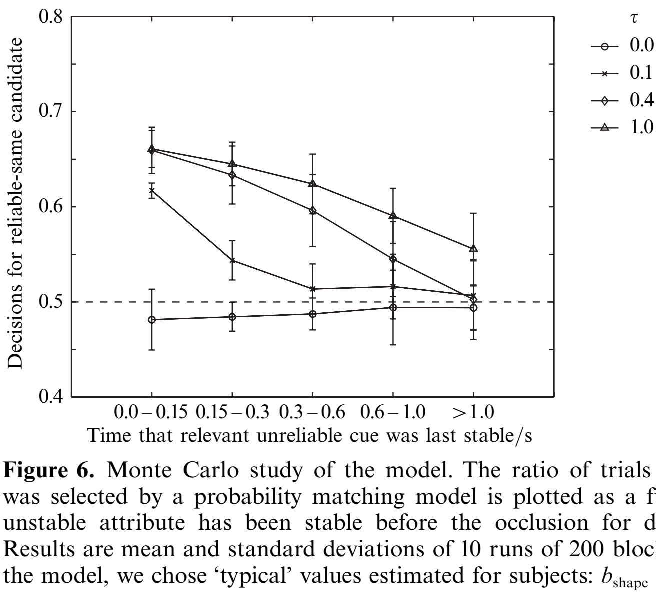



**Figure 6.** Monte Carlo study of the model. The ratio of trials where the stable-same candidate was selected by a probability matching model is plotted as a function of how long the relevant unstable attribute has been stable before the occlusion for different model time constants $\tau$. Results are mean and standard deviations of 10 runs of 200 blocks of trials each. For the biases of the model, we chose 'typical' values estimated for subjects: $b_{\text{shape}} = 0.5$, $b_{\text{color}} = 0.2$, and $b_{\text{size}} = 0.3$.

### 4.1 *Model comparison*

To determine whether or not our model with dynamic re-weighting fits subjects' data better than a static model (a model without temporal dynamics; $\tau \rightarrow 0$) we performed a Bayesian model comparison (MacKay 1992). Model $M_1$ is the model introduced above and model $M_2$ is identical but with the time constant $\tau$ fixed at 0. Thus, model $M_1$ has three free parameters (two independent biases and the time constant $\tau$), while model $M_2$ has only two free parameters (the biases). In the following, we refer to the parameter sets of $M_1$ and $M_2$ as $U_1$ and $U_2$, respectively. With Bayesian statistics, the difference in complexity between the two models is properly accounted for during the comparison in the sense that the model with more parameters is penalized for its extra flexibility (MacKay 1992).

We are interested in the probability of $M_i$ ($i = 1, 2$) being the correct model given the data $D$:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{P(D)} \,. \tag{9}$$

On the assumption of uniform priors $p(M_1) = p(M_2) = 0.5$, $P(M_i|D)$ is proportional to $p(D|M_i)$ since $p(D)$ does not depend on the model. We compute $p(D|M_i)$ by considering $p(D|M_i, U_i)$ and marginalizing over the parameters $U_i$:

$$p(D|M_i) = \sum_{U_i} p(D|M_i, U_i)p(U_i) \,. \tag{10}$$

Here, $p(U_i)$ is the prior probability distribution over the model parameters and $P(D|M_i, U_i)$ is just the likelihood defined in equation (7). By using uniform priors for $p(U_i)$ we can calculate $p(D|M_i)$ and hence $p(M_i|D)$, correctly taking into account that the models have a differing number of free parameters.[(1)] The results are shown in figure 7. For eight out of ten subjects the dynamic model is clearly superior to the static model. Interestingly, subject SB, whose results seem to be most in favor of the static model, was one of two subjects (TK was the other) that verbally suggested at the end of the experiment that the target object was split in two. He reported to have chosen a candidate randomly whenever he noticed this, in which case we would not expect to see any effect, of course.

In summary, our analysis shows that the dynamic model is superior to the static model at describing subjects' cue-integration strategies and that temporal integration and rapid re-weighting of cues are necessary for accounting for subjects' responses.

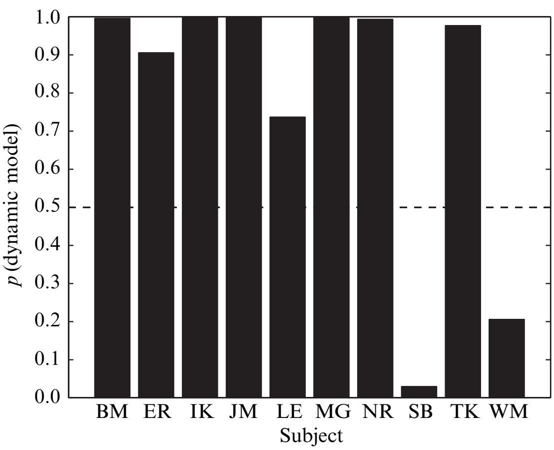



**Figure 7.** Comparison of the dynamic model $M_1$ to the static model $M_2$. The abscissa marks the subject; the ordinate gives the probability of $M_1$ being the correct model.

[(1)] The priors are also called Occam factors in the literature. The model with the smaller number of parameters has a bigger prior; in our case $p(U_2) > p(U_1)$.

## 5 Discussion

In summary, recent studies have demonstrated that visual cue-integration strategies are adaptable in an experience-dependent manner. These studies have shown adaptation on a relatively long time scale (hours or days), but have not considered the issue of whether or not adaptation can influence on-line processing by modifying observers' cue-integration strategies on a short time scale (seconds). Recently, Triesch and von der Malsburg (2001) demonstrated the benefits of very fast re-weighting of cues ($\tau = 0.5$ s) in an artificial face-tracking system, suggesting that a similar mechanism might be beneficial for human perception as well. Here we studied the fast temporal dynamics of visual cue integration by reporting the results of an experiment in which we used a tracking task. Subjects tracked a target object among distractors and identified the target after an occlusion. Objects were defined by three visual attributes (color, shape, and size). Two of the attributes were unreliable in the sense that their values changed frequently within a trial, whereas the remaining reliable attribute was stable. The results are compatible with the hypothesis that subjects rapidly re-weighted the different cues on each trial by emphasizing the information provided by the reliable cue and by discounting the information provided by the unreliable cues. A re-weighting of cues to reflect their current reliability and consistency had been proposed earlier (Landy et al 1995) but no attempts have been made to model the underlying mechanisms and no suggestions have been made regarding the time scale of such a re-weighting. In our experiment, most of the re-weighting took place in less than 1 s and, thus, the results show that cue integration can indeed exhibit adaptive phenomena on a very fast time scale.[(2)]

The issue whether adaptive changes in responses to multiple-cue stimuli are due to changes in observers' cue combination (as we have suggested) or due to changes within individual cues has been problematic for many studies on cue integration (Ernst et al 2000; Atkins et al 2001) and is problematic in this study, too. We have preferred the interpretation of a re-weighting of the information provided by the cues over changes within the cues themselves, but the experiment does not allow to distinguish between these alternatives. This distinction ultimately rests on where we define the processing of a cue to end and the cue-integration process to start.

In experiments on depth cue integration, an often-made assumption is that individual cues compute a best depth estimate given the information provided to the cue and that these best depth estimates of individual cues are subsequently integrated. In our experiment, the situation is somewhat different because we interpret an individual cue not as computing a 'best estimate' for the quantity in question (which is the target?), but as providing more or less evidence for each of the objects depending on the features of the objects, and the subject chooses the object that accumulated most evidence. The classical view from the depth-cue-integration literature is not applicable here because we cannot 'average' entire objects or even only their shapes.

In our model, we try to capture the confluence of evidence from different cues by an additive combination. Such additive combination of evidence from very different cues like color, size, and shape has also recently been reported by Oyama and Simizu (1999) in an experiment on the effects of object similarity on apparent motion and perceptual grouping. In their experiment, differences in color, luminance, size, and shape of dots in grid-like display were found to contribute additively to subjects' perception in an apparent-motion and a perceptual-grouping task.

Although our task has a memory component due to the occlusion during which information about the target trajectory and attributes has to be retained, the memory requirements are well below the capacity of visual working memory which seems to

(2) Whether this re-weighting is called a form of adaptation of the system or part of the flexible on-line processing of a fixed non-adaptive system is a matter of personal taste.

store integrated objects in contrast to individual attributes (Vogel et al 2001). Hence, working-memory limitations should not interfere with the observed effect.

The median time constant for the adaptation estimated for the subjects was 0.3 s. Interestingly, this time is similar to the time scale that Motter (1994) has observed for the switching of responses of V4 neurons when stimuli in the receptive field of a neuron switch from being task-relevant to task-irrelevant or vice versa without a change in the stimulus in the classical receptive field of the neuron. This switching of a feature from being task-relevant to being task-irrelevant can be regarded as a special case of changes in the reliability of a cue for a given task. Hence, Motter's result may be understood as an enhancement/suppression of a cue that suddenly becomes reliable/unreliable. This view suggests that a similar mechanism might accompany the re-weighting of cues in the present experiment.

We expect the observed dynamics of cue integration to be a ubiquitous phenomenon, which affects sensory processing whenever cue changes in their reliabilities for the given task or their inherent noise properties (Triesch 2000), as is often the case in cue-conflict experimental paradigms, or task-switching paradigms. Indeed, it might be the case that researchers have witnessed these effects earlier, but simply misinterpreted them as noise in the cue-integration process.

**Acknowledgements.** This work was supported by NIH/PHS research grant P41 RR09283 and by NIH research grants R29-MH54770 and RO1-EY13149. We thank two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

**References**

Atkins J E, Fiser J, Jacobs R A, 2001 "Experience-dependent visual cue integration based on consistencies between visual and haptic percepts" *Vision Research* **41** 449 – 461

Bruno N, Cutting J E, 1988 "Minimodularity and the perception of layout" *Journal of Experimental Psychology* **117** 161 – 170

Ernst M O, Banks M S, Bülthoff H H, 2000 "Touch can change visual slant perception" *Nature Neuroscience* **3** 69 – 73

Holst E von, 1950 "Die Arbeitsweise des Statolithenapparates bei Fischen" *Zeitschrift für vergleichende Physiologie* **32** 60 – 120

Jacobs R A, Fine I, 1999 "Experience-dependent integration of texture and motion cues to depth" *Vision Research* **39** 4062 – 4075

Kowler E, Anton S, 1987 "Reading twisted text—implications for the role of saccades" *Vision Research* **27** 45 – 60

Landy M S, Maloney L T, Johnston E B, Young M, 1995 "Measurement and modeling of depth cue combination: In defense of weak fusion" *Vision Research* **35** 389 – 412

MacKay D J C, 1992 "Bayesian interpolation" *Neural Computation* **4** 415 – 447

Massaro D W, Friedman D, 1990 "Models of integration given multiple sources of information" *Psychological Review* **97** 225 – 252

Motter B C, 1994 "Neural correlates of feature selective memory and pop-out in extrastriate area V4" *Journal of Neuroscience* **14** 2190 – 2199

Newman E A, Hartline P H, 1982 "The infrared 'vision' of snakes" *Scientific American* **246**(3) 116 – 127

Oyama T, Simizu M, 1999 "Effects of similarity on apparent motion and perceptual grouping" *Perception* **28** 739 – 748

Pelz J B, Hayhoe M M, Ballard D H, Shrivastava A, Bayliss J D, Heyde M von der, 1999 "Development of a virtual laboratory for the study of complex human behavior", in *Proceedings of the SPIE—The International Society for Optical Engineering, 3639B, The Engineering Reality of Virtual Reality, San Jose, CA, USA* pp 416 – 426

Potter M C, 1976 "Short-term conceptual memory for pictures" *Journal of Experimental Psychology: Human Learning and Memory* **2** 509 – 522

Stein B, Meredith M A, 1993 *The Merging of the Senses* (Cambridge, MA: MIT Press)

Treisman A M, Gelade G, 1980 "A feature-integration theory of attention" *Cognitive Psychology* **12** 97 – 136

Triesch J, 2000 "Democratic integration: A theory of adaptive sensory integration", Technical Report NRL TR 00.1, National Research Laboratory for the Study of Brain and Behavior, University of Rochester, Rochester, NY, USA

Triesch J, Malsburg C von der, 2001 "Democratic integration: Self-organized integration of adaptive cues" *Neural Computation* **13** 2049 – 2074

Vogel E K, Woodman G F, Luck S J, 2001 "Storage of features, conjunctions, and objects in visual working memory" *Journal of Experimental Psychology: Human Perception and Performance* **27** 92 – 114

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