Stream specificity and asymmetries in feature binding and content-addressable access in visual encoding and memory

Duong L. Huynh
Department of Electrical and Computer Engineering,
University of Houston, Houston, TX, USA

Srimant P. Tripathy
School of Optometry and Vision Science,
University of Bradford, Bradford, UK

Harold E. Bedell
Center for Neuro-Engineering and Cognitive Science,
University of Houston, Houston, TX, USA
College of Optometry, University of Houston,
Houston, TX, USA

Haluk Öğmen
Department of Electrical and Computer Engineering,
University of Houston, Houston, TX, USA
Center for Neuro-Engineering and Cognitive Science,
University of Houston, Houston, TX, USA

Human memory is content addressable—i.e., contents of the memory can be accessed using partial information about the bound features of a stored item. In this study, we used a cross-feature cuing technique to examine how the human visual system encodes, binds, and retains information about multiple stimulus features within a set of moving objects. We sought to characterize the roles of three different features (position, color, and direction of motion, the latter two of which are processed preferentially within the ventral and dorsal visual streams, respectively) in the construction and maintenance of object representations. We investigated the extent to which these features are bound together across the following processing stages: during stimulus encoding, sensory (iconic) memory, and visual short-term memory. Whereas all features examined here can serve as cues for addressing content, their effectiveness shows asymmetries and varies according to cue–report pairings and the stage of information processing and storage. Position-based indexing theories predict that position should be more effective as a cue compared to other features. While we found a privileged role for position as a cue at the stimulus-encoding stage, position was not the privileged cue at the sensory and visual short-term memory stages. Instead, the pattern that emerged from our findings is one that mirrors the parallel processing streams in the visual system. This stream-specific binding and cuing effectiveness manifests itself in all three stages of information processing examined here. Finally, we find that the Leaky Flask model proposed in our previous study is applicable to all three features.

Introduction

Our visual world is complex and usually cluttered with a large number of objects, many of which can be simultaneously in motion. The ability to track and identify multiple moving objects in a visual scene is therefore essential for navigating and interacting with the surroundings. Significant research in vision science has been devoted to understanding the functional architecture of the visual system that enables the accomplishment of these tasks. Despite its relatively effortless execution, the underlying processes form a complicated computational problem for three reasons (Öğmen & Herzog, 2010). First, the tasks intrinsically demand that identities of objects be established and remain tolerant to continual changes in their defining features produced, for example, by different perspective views of an object when it moves. For example, the frontal and side views of an approaching animal can be drastically different. This problem was recognized by the gestalt psychologist Joseph Ternus, who called it the "problem of phenomenal identity" (Ternus, 1926,

Citation: Huynh, D., Tripathy, S. P., Bedell, H. E., & Öğmen, H. (2015). Stream specificity and asymmetries in feature binding and content-addressable access in visual encoding and memory. *Journal of Vision, 15*(13):14, 1–32, doi:10.1167/15.13.14.

doi: 10.1167/15.13.14 Received April 2, 2015; published September 18, 2015 ISSN 1534-7362 © 2015 ARVO

1938). This problem has also been known as the feature-invariance problem. Second, while we can establish and maintain the identities of objects in a feature-invariant manner, these features are not completely discounted, since our percept at any instant consists of objects with their *specific* features. Hence, along with feature-invariant representations, the brain should also have *feature-specific* representations. Finally, as objects move and occlude each other, features of different objects and features of the background can continuously blend with each other. Hence, the visual system should be able to correctly attribute features to objects (*feature-attribution problem*; Öğmen, Otto, & Herzog, 2006) and bind together different features of a given object (*feature-binding problem*; Revonsuo & Newman, 1999). We reason that addressing these problems requires a detailed understanding about how individual features contribute to the construction and maintenance of an object's identity and how they are temporally related.

Much of relevant research has focused on the topic of feature binding. Classically, it has been established that the processing of different visual features is mediated by distributed cortical areas. For example, evidence from previous electrophysiological recordings and functional-imaging studies suggests that surface features (e.g., color, form) of an object are processed in areas along the ventral pathway (Zeki, 1976; Desimone, Schein, Moran, & Ungerleider, 1985), while the object's spatiotemporal features (e.g., direction of motion) are processed in areas along the dorsal pathway (Maunsell & Van Essen, 1983; Born & Bradley, 2005). This anatomical and functional segregation leads to the questions of how and where in the brain visual features come to be integrated in the first place. Over the years, many solutions to the problem of feature binding have been proposed. According to Riesenhuber and Poggio (1999), perceptual binding is based on "grand-mother" neurons or neuron populations that are functionally selective for specific combinations of features. However, given that visual features such as form or color can all take on innumerable values, this way of binding processing seems impractical in many cases. Moreover, there has been evidence showing that the same neuron population can be activated by many different combinations of features (Basole, White, & Fitzpatrick, 2003). Alternatively, some researchers have suggested that bound features are represented via synchronized spiking activities of neurons from different brain areas (Von Der Malsburg, 1981; Gray, König, Engel, & Singer, 1989), but conflicting results have been reported as well (Fries, Neuenschwander, Engel, Goebel, & Singer, 2001; Thiele & Stoner, 2003; Dong, Mihalas, Qiu, von der Heydt, & Niebur, 2008).

In other studies, the pivotal role of visual attention in binding is emphasized not only in perception (Treisman & Gelade, 1980; Treisman & Schmidt, 1982; Braet & Humphreys, 2009; Hyun, Woodman, & Luck, 2009; Reeves, Santhi, & DeCaro, 2005) but also in visual short-term memory (Wolfe, 1999; Rensink, 2000; Wheeler & Treisman, 2002; Fougnie & Marois, 2009; but see Allen, Baddeley, & Hitch, 2006; VanRullen, 2009). For example, Hyun and colleagues (2009) used a visual-search paradigm and measured the lateralized N2pc event-related-potential component, a hypothesized indicator of an observer's allocation of attention to the contralateral hemifield. The amplitude of N2pc was found to be larger in a task that required observers to bind the color of the target to its location compared to that in a task that simply required observers to detect a target color. This indicates that the binding of an object's color to its location demands attentional resources. Saiki (2003a, 2003b) introduced the Multiple Object Permanence Tracking paradigm, in which the stimulus consisted of a rotating pattern (three dots of different colors forming a triangle) behind a fan-shaped occluder (which also rotated in some experiments). There were three conditions. In the replacement condition, one of the dots in the pattern was replaced by a dot of a different color; in the switch condition, two of the dots in the pattern briefly switched colors; and in the normal condition, there was neither replacement nor switch. Observers had to report if there was any irregularity (replacement or switch) in the stimulus. Performance was good for detecting replacement but was substantially poorer for detecting switches. These results suggest that binding of color and location is poor under dynamic conditions even when there are only three objects, in contrast to static conditions (Luck & Vogel, 1997) where feature-location binding of four objects was possible.

Combined, these findings do not provide a coherent picture concerning how object representations emerge and how bound features become associated with these representations.

Pylyshyn and Storm (1988) devised a Multiple-Object Tracking paradigm to investigate how the identities of objects are maintained in the absence of all features but position and motion. Observers were presented with a multiple-object motion stimulus and were instructed at the beginning of each trial to selectively track a specific subset of objects, the target set. After a period of linear motion, they were tested on their ability to distinguish which objects belonged to this target set. Observers could typically track four or five targets with 85% accuracy. In experiments of this type, because all objects appear identical during the motion period, information about object identities obtained initially must be maintained until responses are made based on other stimulus features, i.e., position and motion. The researchers proposed the Fingers of INSTantiation model, in which low-level internal pointers, referred to as visual indices, attach themselves to the tracked objects and move automatically with the objects. This position-based indexing mechanism occurs preattentively and forms the basis for subsequent cognitive operations on indexed objects (for a modified version of this model, see Alvarez & Franconeri, 2007). However, the contribution of other stimulus features, such as direction of motion, color, or shape, has also been examined in many studies employing variants of the Multiple-Object Tracking paradigm, and the results suggest that they all can play a role in indexing moving objects (Keane & Pylyshyn, 2006; Fencsik, Klieger, & Horowitz, 2007; Horowitz, Fine, Fencsik, Yurgenson, & Wolfe, 2007; Makovski & Jiang, 2007; Iordanescu, Grabowecky, & Suzuki, 2009).

Another position-based indexing model was proposed by Kahneman, Treisman, and Gibbs (1992). They used an "object-specific priming" paradigm and proposed that object representations are created by opening "object files" and are maintained by indexing these object files by their instantaneous position. According to this view, successive states of objects are updated by inserting features into their object files. One problem with this model is that in order to open an object file, one needs access to the features that define a distinct object; however, features cannot be accessed if an object file is not already opened. Another problem with this model is that in order to decide which feature to insert in which object file, one needs to know the spatial extent of each object. The spatial extent in turn requires the use of features other than position, as one needs to compute boundary and surface features of the object to determine where in space the object starts and where it ends. Hence, while spatial position can play an important role in indexing objects and maintaining their identities, other features must also play a role in these processes.

Despite the aforementioned evidence that features other than position can play a role in indexing objects, position remains as the predominant cue in experiments designed to study stimulus encoding and storage. Typically, in these experiments, an array of items is shown briefly, followed by a retention period. At the end of the retention period, a position cue is used to prompt the subject to report one or more features of the stimulus that was presented at the corresponding position. However, human memory is *content addressable* in a more general sense, in that we can access memory by using any partial information about its contents. Hence a general study of content-addressable memory should address how any given feature of the stored item can allow the retrieval of other features of that item. From this perspective, the efficiency of the memory is not characterized just by how well features are stored but also by how well they can be recalled by partial contents (cues). In the present study, we used a cross-feature cuing technique to investigate the efficiency of information encoding and retention in the form of content-addressable memory. Specifically, we sought to characterize the roles of three different features (*position*, *color*, and *direction of motion*) in the construction and maintenance of object representations across different processing stages, from perceptual encoding to sensory (iconic) memory and visual short-term memory (VSTM). We hypothesize that the effectiveness of cross-cuing—i.e., how well one feature can be retrieved, given another—reflects the ability of a feature to serve as an index for content-addressable memory as well as the strength of the binding between those two features. We expect that the effectiveness of cross-cuing will be affected by the number of objects in the display as well as the delay in cuing the feature that identifies the object to be reported.

In the cross-cuing approach, observers reported one feature of a single object, its final position, direction of motion, or color, while one of the other two features was cued. Using this approach, we investigated processing for each stimulus feature and the relationship between these features across different early stages of visual information processing. Experiment 1, in which the target was cued immediately after all objects stopped moving and disappeared, aimed to characterize the stimulus-encoding stage prior to memory stages. While the observers had to hold in memory information about this cued target during the adjustment phase, having a single target item and no delay after stimulus offset minimized the involvement of memorization in their performance. Experiment 2 with varying cue delays aimed to tap into iconic memory and VSTM. We varied the set size in Experiment 1, but not in Experiment 2—unlike in our previous study (Öğmen et al., 2013)—because our pilot data showed that for a set size of 1, the drop of performance over time was negligible (flat line), although the effect of cue delay was statistically significant. This is consistent with our previous finding (Öğmen et al., 2013). In Experiment 2, only a set size of 6 was chosen for all feature combinations; when compared with the findings for a set size of 1, performance at this set size as a function of cue delay was sufficient to assess the loss of information during each processing stage.

Methods

Participants

The first author and three observers who were unaware of the purpose of the study, had normal or corrected-to-normal vision, and had no color deficiency (according to self-reports and the online version of the

Ishihara test) participated in all experiments. Observers were not informed about the specific purposes of the experiments. Experiments were conducted according to a protocol adhering to the Declaration of Helsinki and approved by the University of Houston Committee for the Protection of Human Subjects.

Apparatus

A Visual Stimulus Generator system (Cambridge Research Systems, Rochester, UK) with a VSG2/3 video card driving a NANAO FlexScan color monitor (20 in., 100 Hz) was used to create and display stimuli; programming was implemented in C++. The screen resolution was 800 × 600 pixels, of which 656 × 492 pixels (18.5° × 14°, or 1.7 arcmin/pixel in terms of visual angle) were used for object display. The screen edges were visible during the experiments, but the border of the display area was not. Observers used a computer mouse to give their response, and their heads were kept still on a head/chin rest at a distance of 1.0 m in front of the monitor.

Experiment 1: Stimulus encoding

Stimuli

Objects were circular disks of different readily distinguishable colors that were randomly selected from the CIE L*a*b* color system (perceptually uniform). A constraint was used so that the color separation of any two objects was not less than 17° in L*a*b* space (with this separation, colors were salient and readily distinguishable; also see Selection operator $S^{T}_{i=1;i\neq t}[.]$, paragraph later, for further justification of this separation angle). The diameter of each object was chosen to subtend a visual angle of 1°. Objects were displayed on a gray (40 cd/m^2) background.

Procedure

Observers started each trial by clicking the mouse. Objects of a specific set size appeared on the screen at random, nonoverlapping locations. All objects remained stationary on the screen for 1 s, then started moving along linear trajectories at a constant speed of 5°/s, each in an independent, randomly selected direction. Like the color-separation constraint, the motion directions were constrained so that no two objects had motion directions closer than 17°. In our previous study using a similar task (Shooner, Tripathy, Bedell, & Öğmen, 2010), we used two durations of motion, 200 ms and 5 s. The results showed that at 200 ms, motion information was sufficiently suprathreshold that increasing the duration of motion did not improve encoding or storage of motion. Hence in this study motion duration was fixed at 200 ms for all trials. Objects did not interact with each other when their positions overlapped—i.e., they moved across each other without any change in their motion. However, if an object hit the invisible border of the display area, it would bounce back and change its direction of motion by reversing either the horizontal or vertical component of the velocity vector.

After the objects stopped moving, they were all removed from the display. One of them was randomly selected to be the target for report. In this experiment, the cue was given immediately after the disappearance of the objects, and observers provided their responses using a mouse.

In separate blocks, the procedure described was applied to six types of conditions corresponding to the following six combinations of cued and reported features (Figure 1a through c illustrates three of them):

(a) Position/motion direction
(b) Position/color
(c) Motion direction/position
(d) Motion direction/color
(e) Color/position
(f) Color/motion direction

At the beginning of each block, observers were informed about which feature would serve as the cue and which feature they would be reporting. This blocked design is consistent with a real-world scenario in which the observer has a predefined goal and analyzes the stimulus based on preselected feature dimensions. On the other hand, a randomized design would better capture the situation in arbitrary environments in which the observer does not have any a priori attentional focus on stimulus features. Our goal in using the blocked design was to remove the uncertainty about the cue and the item to be reported, so as to study cue–report combinations in their strongest form. Removing this uncertainty allowed observers to focus their attention on the relevant feature dimensions. On the one hand, this allowed stronger association/binding between the attended feature dimensions; on the other hand, it reduced the noise in the data that would arise from trial-to-trial variability. In this experiment, the cue was delivered immediately at stimulus offset so as to access the stimulus-encoding stage prior to memory. We acknowledge that, since it takes time for the observer to detect and decode the cue, an inevitable time delay occurs after the stimulus offset. However, given that the cue type was predictable and the stimuli were highly salient, and since observers were reporting a single item, we expect the effect of sensory memory to be minimal.

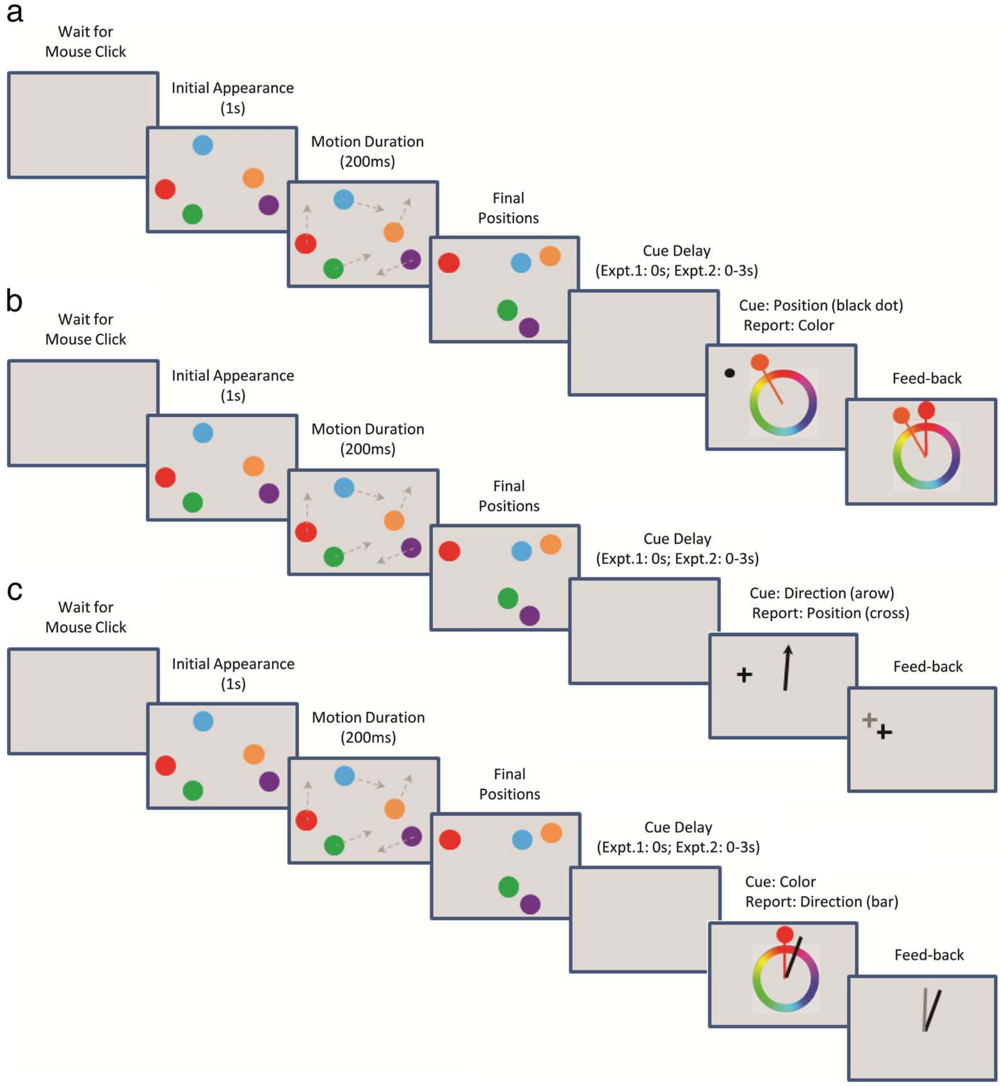


Figure 1. Time course of a trial in Experiments 1 (no cue delay) and 2 (varying cue delay) for the cases of (a) cuing the target's terminal position and reporting its color (position was cued by a small black dot, and observers reported the cued target's color on a color wheel); (b) cuing the target's final direction of motion and reporting its position (direction of motion was cued by a black arrow extending from the center of the screen, and observers adjusted the mouse cursor's location—represented by a cross—to report the cued target's position); and (c) cuing the target's color and reporting its direction of motion (color was cued on a color wheel, and observers adjusted the black bar—which extended from the center of the screen—to report the cued target's direction of motion).

In the cases where position was cued (conditions a and b), a small black dot was presented at the terminal position of the target object to prompt the observer's response (Figure 1a). For motion direction to be cued (conditions c and d), an arrow—a black line segment extending from the center of the screen—was shown, with its orientation indicating the target's motion direction (Figure 1b).[1] If the target had bounced, its final direction of motion was cued.

For the conditions related to color, a color wheel containing all possible colors of objects was used. A total of 180 colors on the color wheel, with each assigned an angular value (resolution = 2°/color), were defined along a circle in the CIE L*a*b* color space. All colors had equal luminance, which was set at $L = 15$ cd/m^2 and varied in hue. The color wheel was centered at the white point ($a = 0.2044$ and $b = 0.4808$), and its radius was chosen to maximize the discriminability of the colors (approximately 2.0°).

When color was cued (conditions e and f), the color wheel was presented at the center of the screen together with a small disk of the target's color (Figure 1c). The small disk that showed the target's color was located outside of the color wheel and connected to it by a bar with an arrow (not shown) pointing towards the wheel center at the part of the wheel that matched the color of the target.

To report, observers moved the mouse cursor on the screen, causing the reporting indicator of a noncued object feature to appear. The direction indicator, which was to report the target's motion direction, was a black bar of fixed length extending either from the target's terminal location (condition a, not shown) or from the screen center (condition f; Figure 1c). The position indicator, which was to report the target's position (conditions c and e), was a small black cross (Figure 1b) representing the location of the mouse cursor as it was moved. The color indicator (conditions b and d) was designed in a similar manner as the case for cuing color, except that the small disk could be moved along a circle around the color wheel and its color changed corresponding to its location (Figure 1a). The observers were required to indicate where on the color wheel the highlighted color best matched the target's color.

On each trial, observers responded by moving the reporting indicator of each type to best match the corresponding target feature and clicking the mouse. An additional indicator of the same design but in a different color (gray when reporting motion direction or position, the target's actual color when reporting color) then appeared to provide the observers with feedback as to the correct response, and the trial ended. Observers started a new trial whenever they clicked the mouse again.

Design

For each combination of cued and reported features, seven set sizes (one, three, four, six, eight, nine, twelve objects) were tested. Each combination was run in five separate sessions to obtain a total of 100 trials per set size. Therefore, there were 20 trials per set size in each session, yielding $7 \times 20 = 140$ trials per session, or $140 \times 5 = 700$ trials per combination, for each observer. Trials of all set sizes were randomly interleaved within every session. Each session took an observer approximately 15 min to finish.

Experiment 2: Iconic memory and VSTM

In separate blocks, the six types of experimental conditions from Experiment 1 were run again in Experiment 2, with for the following changes:

- The number of objects was fixed at six in all conditions.
- On each trial, the cue was provided not immediately after object disappearance but after a variable-duration delay. The delay on each trial was randomly selected from seven different values (0, 50, 100, 250, 500, 1000, 3000 ms).

Again, there were 20 trials per delay condition in each session (or $20 \times 7 = 140$ trials per session). A total of five sessions yielded $140 \times 5 = 700$ trials for each of the six conditions. Observers followed the same steps as in Experiment 1.

Data analysis

Similar to the methods we used in a previous study (Öğmen et al., 2013), we investigated the qualitative and quantitative aspects of the observers' performance by implementing statistical modeling of the error distribution for each object feature being probed. We analyzed the error distributions using a hierarchical family of models and compared the performance of these models. Interpretation of the data in each cue–report condition was then based on estimates of the parameters given by the best-performing model.

We will proceed shortly to consider the details of each model. It should be noted here that, in the present study, we slightly modified the procedure of fitting the models to empirical data. The procedure we used before (Öğmen et al., 2013) involved binning the data in each experimental condition to generate a frequency histogram, which was rescaled to unit area to obtain the corresponding probability density function, after which a nonlinear least-squares optimization routine was employed to find the values of the model parameters that provided the best fit to the empirical probability density function. However, binning the data might introduce artifacts, as fitting results typically exhibit sensitivity to the bin size. In the current study, we therefore used cumulative distribution functions (cdfs) to specify the distribution of response errors. Using CDFs in the form of a stepwise monotonic increasing function, in which accumulation occurs at every single value of the error variable, removes the need for binning the data. The selection of the best-performing model was based on the adjusted R^2 values obtained for each model. The difference between the adjusted R^2 values was rather small; hence we also carried out Bayesian analysis in which the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) was used for finding maximum-likelihood estimates of the parameters. This algorithm is described in detail in Supplementary information 2. For model selection, we used the Akaike information criterion and the Bayesian information criterion. These two model selection methods provided the same outcome in 11 out of 12 cases. However, there were some differences between the least-squares and Bayesian approaches. They provided the same model in seven out of 12 cases, but different models in the remaining five cases. Hence we report the results from both least-squares (in the main text) and Bayesian approaches (in

Supplementary information 2). Notwithstanding the differences in model selection, the general qualitative trends obtained with least-squares and Bayesian approaches were essentially similar.

As detailed in the section Design, our experiments investigated three object features: position, direction of motion, and color. Of these, position was defined in 2-D coordinates, while direction of motion and color were drawn from circular spaces. Therefore, statistical modeling was carried out separately for the horizontal and vertical components of position errors and needed to take into account the circular properties of direction of motion and color parameter spaces.

Statistical modeling

Model 1: Gaussian

The simplest model in the family is the cdf of a circular (wrapped) Gaussian:

$$\mathrm{CDF}(\varepsilon) = \mathrm{CDF}\{\mathrm{G}(\varepsilon; \mu, \sigma)\}, \qquad (1)$$

where the cumulative distribution function CDF(ε) of error variable ε (ε = reported feature value − actual feature value) is given by a Gaussian distribution $G(\varepsilon;\mu,\sigma)$ whose parameters represent the accuracy (mean: μ) and the precision ($1/\sigma$, where σ is the standard deviation) of processing. The *precision* parameter $1/\sigma$ captures the qualitative aspect of performance, with smaller values of σ corresponding to higher qualities of encoding for the processed items.

For a practical implementation, the effect of multiple wraps was tested in Shooner et al. (2010). Three Gaussians were initially included in the sum, and the outcome was compared with that produced by only one Gaussian. The difference was negligible due to the small variance of the distributions, which meant using a *single Gaussian* was sufficient to model the empirical data, and the circular nature of features could be ignored. However, we consistently applied three wraps in all conditions of reporting color and motion direction in the current study for the following reasons: (a) The variance of our data was large in some conditions; (b) the wrapping effect could not be ignored on the misbinding component (see Model 3; Model 3 was not used by Shooner et al., 2010); and (c) we observed no difference between three and five wraps.

Model 2: Gaussian + Uniform

In this model (Zhang & Luck, 2008), the distribution of errors is represented by

$$\mathrm{CDF}(\varepsilon) = \mathrm{CDF}\{w.\mathrm{G}(\varepsilon; \mu, \sigma) + (1 - w) \\ .\mathrm{U}(-180, 180)\}, \qquad (2)$$

where the cumulative distribution function CDF(ε) is obtained from the corresponding probability density function that consists of two components:

(a) A Gaussian distribution $\mathrm{G}(\varepsilon;\mu,\sigma)$ described in the Gaussian model
(b) A uniform distribution U over the interval (−180, 180), which represents guessing

The weight of the uniform distribution $(1 - w)$ represents the proportion of trials in which observers base their responses on guesses rather than on the target information available. The weight w of the Gaussian captures the quantitative aspect of performance by providing a relative measure for the *intake* of encoding, with a larger value corresponding to a greater possibility that a response is based on having some access to information from the cued target.

Model 3: Gaussian + Uniform + Gaussian

This model (Bays et al., 2009) includes an additional term to account for misbinding errors—i.e., errors resulting from incorrect associations of features with objects, when observers get confused and report the features of another object instead of the selected target:

$$\mathrm{CDF}(\varepsilon) = \mathrm{CDF}\Big\{w.\mathrm{G}(\varepsilon; \mu_t, \sigma_t) + (1 - w - w_m) \\ .\mathrm{U}(-180, 180) \\ + w_m.\mathrm{S}^T_{i=1;i\neq t}\big[\mathrm{G}(\varepsilon; \mu_t + \varepsilon_{i,t}, \sigma_t)\big]\Big\}, \qquad (3)$$

where the first two terms represent the same Gaussian and uniform distributions as in the Gaussian + Uniform model and the third term represents errors stemming from misbinding reports. The selection operator $\mathrm{S}^T_{i=1;i\neq t}[.]$ determines which item from the set of $(T - 1)$ noncued objects is the one that generates the subject's response due to a misbinding error. The misbinding term is expected also to have a Gaussian distribution, with the same standard deviation as the first Gaussian but with the mean shifted from the first Gaussian by the difference $\varepsilon_{i,t}$ in the reported feature space between the cued target and the misbinding object. This is because the empirical cdf is always computed with respect to the cued target item. For the Gaussian component that describes misbinding, the wrapping effect cannot be ignored, especially when the misbinding object is shifted far away from the center of the first Gaussian. The weight w_m is to represent the proportion of trials in which misbinding occurs.

Selection operator $\mathrm{S}^T_{i=1;i\neq t}[.]$: Technically, the third term on the right-hand side of Equation 3 needs to be computed based on all noncued objects, as misbinding may potentially occur on any of them (Bays et al., 2009). A synthetic simulation that we performed, however, showed that doing so can result in misleading estimates for this component. In order to estimate spurious contributions from different terms, we can simulate experiments where the contribution of a given term is nulled. Under this condition, the analysis of the other term provides an estimate of the spurious contributions that would result from this term. We simulated three scenarios: In the first simulation, by using the actual stimuli without any constraints on the directions of motion,[2] we had the computer respond always to the cued target with zero error. This effectively transforms the probability distribution resulting from the first term on the right-hand side of Equation 3 to a delta Dirac function. The computed error distribution from the misbinding term provides an estimate of its spurious contribution (since the computer is always responding to the target). This distribution was found to be significant because when the number of objects is large, there is a good probability that there will be another target moving in a similar direction as the cued one. We did the same analysis for the other two cases—i.e., the computer responds to the noncued object that is closest to the cued target either in position (*closest cued feature*) or in direction of motion (*closest reported feature*)—and also found significant spurious contributions from the first term on the right-hand side of Equation 3.

To minimize the potential interference that results from directions or colors that are too close to each other, we constrained our stimuli so that no two items in any stimulus had a direction or color difference of less than 17°. This value was chosen to obtain a rough symmetry of the three feature dimensions, in which the constraint for position was nonoverlapping (center-to-center distance between any two objects must be larger than 1°). Accordingly, the ratio of the minimum to the maximum possible separation was equivalent along each feature dimension. Given these constraints, we analyzed two versions of this model—i.e., misbinding with the object that is closest to the cued target in either the cued feature space (closest cued feature) or the reported feature space (closest reported feature).

Goodness-of-fit measure and model comparison

The models described are different from each other in terms of decomposition approaches and complexity to account for different aspects of the data. Increasing the degrees of freedom usually makes a model better at fitting the observed data, but the cost is that the model also may capture random patterns of noise (overfitting) and hence, compared to a simpler model, be less likely to translate well to other data sets from the population. We computed adjusted coefficients of determination (adjusted R^2) to evaluate and compare the performance of the models.

Just like the coefficient of determination (R^2), adjusted R^2 reflects the efficacy of a model in reproducing empirical data by measuring the fraction of variation in the dependent variable that is explained by the independent variable(s). However, using adjusted R^2 is more appropriate when comparing multiple models with different degrees of freedom, as it compensates for the addition of independent variables if doing so does not significantly improve the explanatory power of a model. The following equation is used to compute adjusted R^2:

$$\textit{Adjusted } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}, \tag{4}$$

where n is the sample size and p is the number of independent variables (parameters) in the model.

Results

Experiment 1: Stimulus encoding

Analysis of performance

We computed the magnitude of response error on each trial as the absolute difference between the actual and reported values of the probed feature, and rescaled it to the range [0, 1] by using a transformation metric defined by the following equation:

$$TP = 1 - |\varepsilon|/MAX, \tag{5}$$

where ε is response error and MAX represents the maximum possible value of $|\varepsilon|$. For direction of motion and color, this is a constant value (180°). For position, since response error typically depends on the size of the display screen and slightly varies across observers, MAX takes the maximum magnitude of position error produced by each observer in each cue–report combination. Finally, TP represents transformed performance, which takes the values of 1.0 and 0.5 for perfect and chance levels of performance, respectively.

Figure 2 plots TP (left y-axis) and $|\varepsilon|$ (right y-axis) averaged across observers as a function of set size, with each panel showing the data for one reported feature and different symbol colors representing different cue types. A two-way repeated-measures ANOVA shows that, in all cases, the main effects of set size—position reported: $F(6, 18) = 109.092$, $p < 0.0001$, $\eta_p^2 = 0.973$; direction of motion reported: $F(6, 18) = 82.350$, $p < 0.0001$, $\eta_p^2 = 0.965$; color reported: $F(6, 18) = 127.519$, $p < 0.0001$, $\eta_p^2 = 0.977$—and cue type—position reported: $F(1, 3) = 18.080$, $p = 0.024$, $\eta_p^2 = 0.858$; direction of motion reported: $F(1, 3) = 34.547$, $p = 0.01$, $\eta_p^2 = 0.920$; color reported: $F(1, 3) = 188.679$, $p = 0.001$, $\eta_p^2 = 0.984$—are significant. The interaction between set size and cue type is significant when direction of motion is reported,

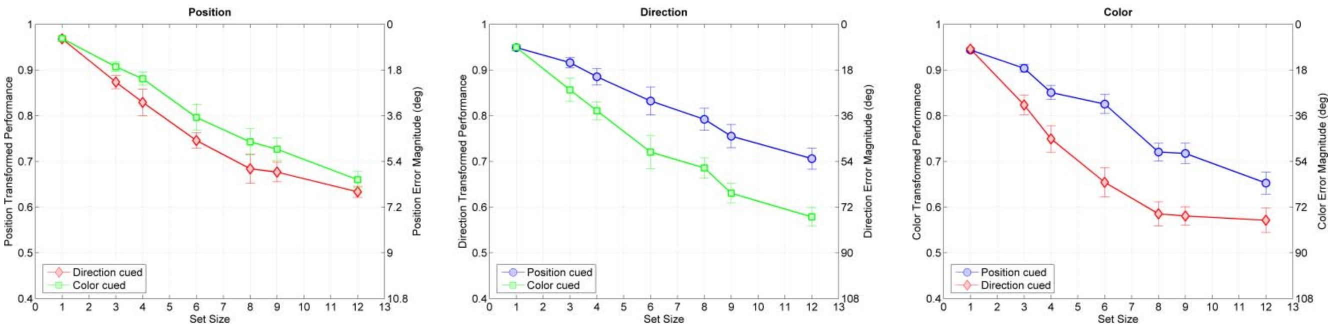


Figure 2. Transformed performance (left y-axes) and error magnitude (right y-axes) in Experiment 1, averaged across observers, as a function of set size for position (left panel), direction of motion (middle panel), and color (right panel). Each symbol color corresponds to a cue type (blue = cuing position, red = cuing direction of motion, green = cuing color). Error bars correspond to ±1 standard error of the mean.

$F(6, 18) = 6.341$, $p = 0.01$, $\eta_p^2 = 0.679$, and when color is reported, $F(6, 18) = 14.356$, $p < 0.0001$, $\eta_p^2 = 0.827$, but not when position is reported, $F(6, 18) = 1.244$, $p = 0.331$, $\eta_p^2 = 0.293$. For observer-specific effects, we used a two-way ANOVA separately for each observer to compare performance with respect to set size and cue type. The same pattern of results (Supplementary Figure S1.1) for each feature dimension is observed for all observers, except that the main effect of cue type is not significant when position is the reported attribute for one of the four observers (DHL).

Consider first the effect of cue type. We find that, for set sizes larger than 1, performance for direction of motion is better when cuing with position than when cuing with color. Similarly, the performance for reporting color is better when cuing with position than with direction of motion. On the other hand, the performance for position when cuing with direction of motion is not as good as when cuing with *color*. The similar performance obtained at a set size of 1 for the different cues is not surprising. On single-object trials, the observers can simply pay attention to the feature being probed instead of the cued feature, which is not useful for tracking. Furthermore, one can see that performance for each feature is very high and similar to each other (compare data points for set size 1 across the three panels). This indicates that, in isolation, features were matched to each other in salience. However, their pair-wise associative encoding as a function of set size was different, as shown by different drops in performance as a function of set size.

Figure 3 shows a rough sketch of our results for a more intuitive look into the role of each stimulus feature and the relationships between features. The size of the arrows linking two features represents the effectiveness of using the feature at the tail of the arrow to report the feature at the head of the arrow. For each case, the transformed performance averaged across set sizes is shown next to the corresponding arrow. The sketch highlights the following important points: First, position is an effective cue whether direction of motion or color is reported. In contrast, color and direction of motion are more effective as cues in reporting position than in reporting each other. Inspection of Figure 3 suggests that feature binding and content-addressable access in sensory encoding is reflective of the two parallel pathways: dorsal and ventral streams specialized in motion and color. The early parts of ventral and dorsal streams are retinotopically organized (e.g., Tootell, Silverman, Switkes, & De Valois, 1982; Sereno et al., 1995; Engel, Glover, & Wandell, 1997), and the retinotopic organization provides position information to both streams. However, when stimuli are in motion, retinotopic position information needs to be converted to nonretinotopic position information. Computation of visual attributes such as form (Öğmen et al., 2006), luminance (Shimozaki, Eckstein, & Thomas, 1999), color (Nishida, Watanabe, Kuriki, & Tokimoto, 2007; Cavanagh, Holcombe, & Chou, 2008), size (Kawabe,

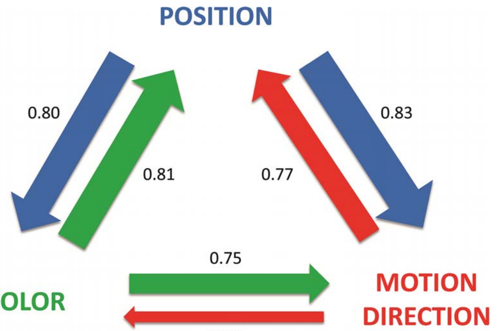


Figure 3. Diagram showing the effectiveness for each cue–report combination for the stimulus-encoding stage (Experiment 1). The size of the arrows represents the effectiveness of using one feature to recall another. The cue is at the tail of the arrow, while the reported feature is at the tip of the arrow. The number next to each arrow shows the average TP across set size for that particular cue–report combination. Different colors represent different cue types (blue = cuing position, red = cuing direction of motion, green = cuing color).

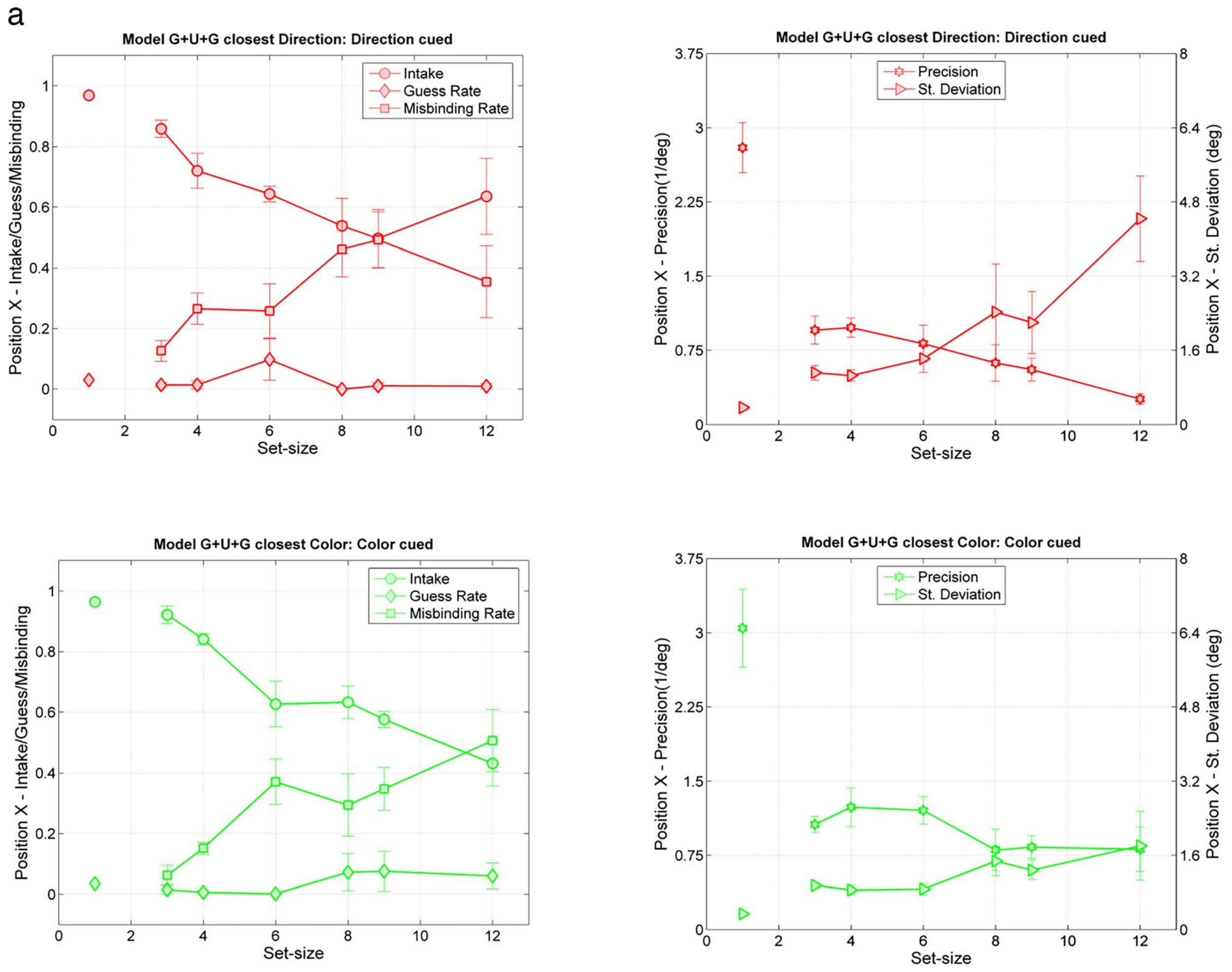


Figure 4. (a) Decomposition of performance for the horizontal component (X) of position in Experiment 1. Left panels show the results for intake (w) along with guessing ($1 - w - w_m$) and misbinding (w_m) rates; right panels show the results for precision ($1/\sigma$: left y-axes) and standard deviation (σ: right y-axes); upper and lower panels show the results for the cases of cuing direction of motion and color, respectively; and x-axes represent set size. Error bars correspond to ±1 standard error of the mean across observers. Parameters for each cue type are shown for the winning models (see top of each panel). In cases that either version of Model 3 won, data at a set size of 1 are taken from Model 2, because Model 3 is not applicable. (b) Same as (a), for the vertical component (Y) of position in Experiment 1.

2008), and motion (Cavanagh et al., 2008; Boi, Öğmen, Krummenacher, Otto, & Herzog, 2009) are shown to occur according to motion-based nonretinotopic reference frames, suggesting that nonretinotopic position information is available to both dorsal and ventral streams. While there is still a debate on more abstract representations of position in ventral and dorsal streams, such as a distinction between near versus far space (Lane, Ball, Smith, Schenk, & Ellison, 2013), it is reasonable to assume that, for the basic features examined in this study, position is a common attribute to both of these pathways. Hence binding and content-addressable access occur more effectively within each pathway than across pathways, as reflected by relatively weak connections between color and direction of motion.

As for the effect of set size, we observed a progressive degradation of performance with increasing set size for all cue–report combinations. The traditional view of information-processing bottlenecks is that stimulus encoding and sensory memory are high-capacity stages followed by VSTM, where the bottleneck occurs. In contrast to this view, Öğmen et al. (2013) showed that a substantial bottleneck occurs already at the stimulus-encoding stage for motion processing. The results here are consistent with this previous finding and generalize the observation of an early bottleneck for motion direction to all three of the feature dimensions investigated here.

Taken together, these results indicate (a) a more effective binding within dorsal and ventral pathways compared to across the pathways and (b) significant

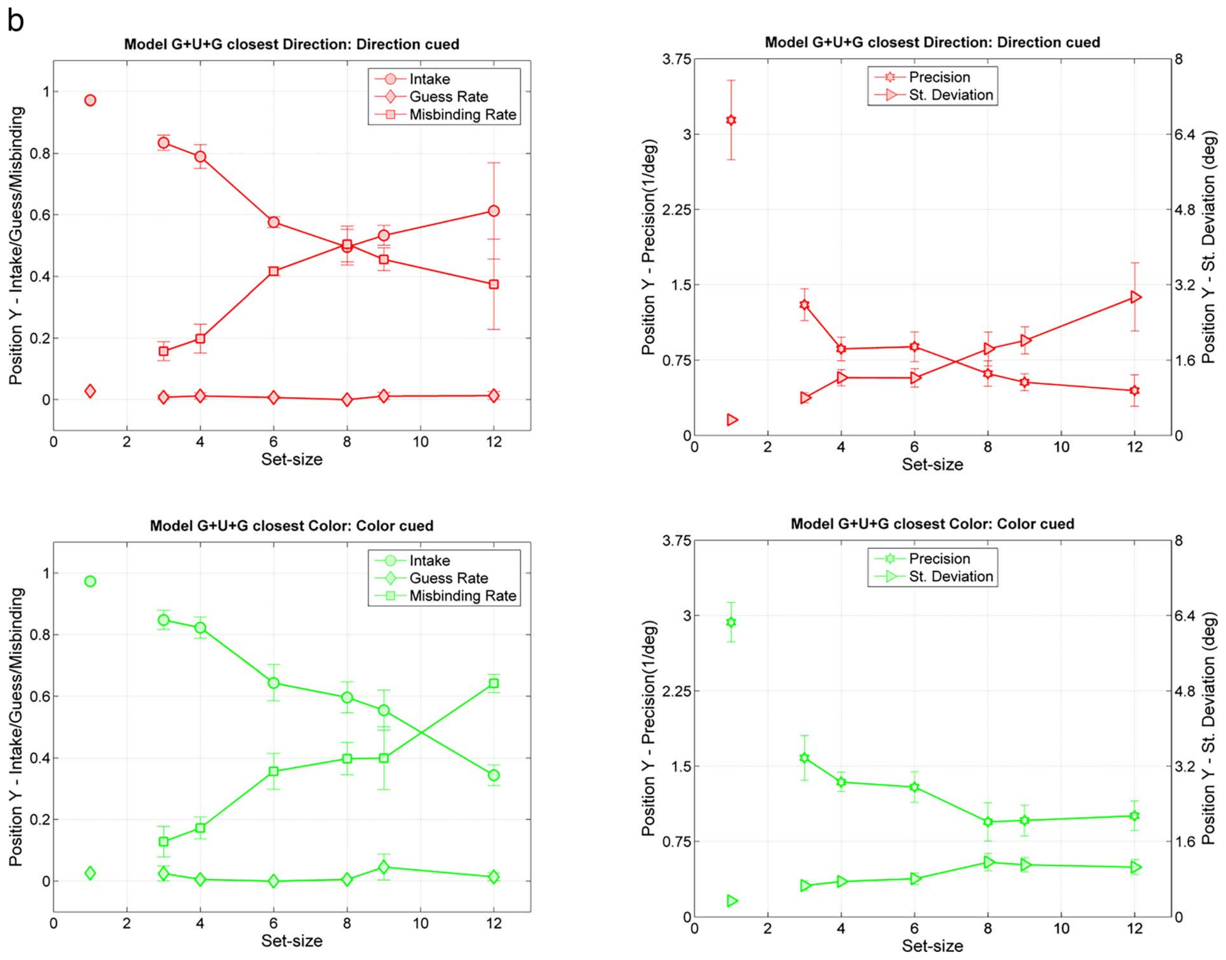


Figure 4. Continued.

capacity limits prior to sensory memory for all features and their associations.

Statistical modeling

As detailed already under Data analysis, we decomposed performance into *quantitative* and *qualitative* measures by fitting to our data Models 1 and 2 and two versions of Model 3. The averaged estimates for all relevant parameters in the models are plotted as a function of set size in Figures 4 through 6 for position (4a: horizontal component; 4b: vertical component), direction of motion, and color, respectively. In each figure, the left panels show the estimates for intake (w) along with guess ($1 - w - w_m$) and misbinding (w_m) rates; the right panels show the estimates for precision ($1/\sigma$: left y-axis) and standard deviation (σ : right y-axis); and the upper and lower panels show the results for the two different cue types. Note that, only the winning (highest adjusted R^2) model in each cue–report condition is shown here. Results of model-selection analyses for all conditions are shown in Table 4, and the adjusted R^2 values are given in Table 5.

Consistently across feature dimensions, we find that both intake and precision of encoding decrease with increasing set size. However, the relationships between standard deviation or intake and set size are more poorly approximated by straight lines than are the results for direction of motion found by Öğmen et al. (2013), and a lower asymptote seems to occur in some conditions. In addition, visual inspection of the effect of cue type on intake and precision of encoding for each stimulus feature suggests that the advantage of using one cue over another (e.g., using position vs. color as a cue for reporting the direction of motion) manifests itself through both the quantitative and qualitative aspects of performance. Although there were some differences between the least-squares and the Bayesian methods in the selected models, the same qualitative findings are also observed in Bayesian analysis (Supplementary information 2).

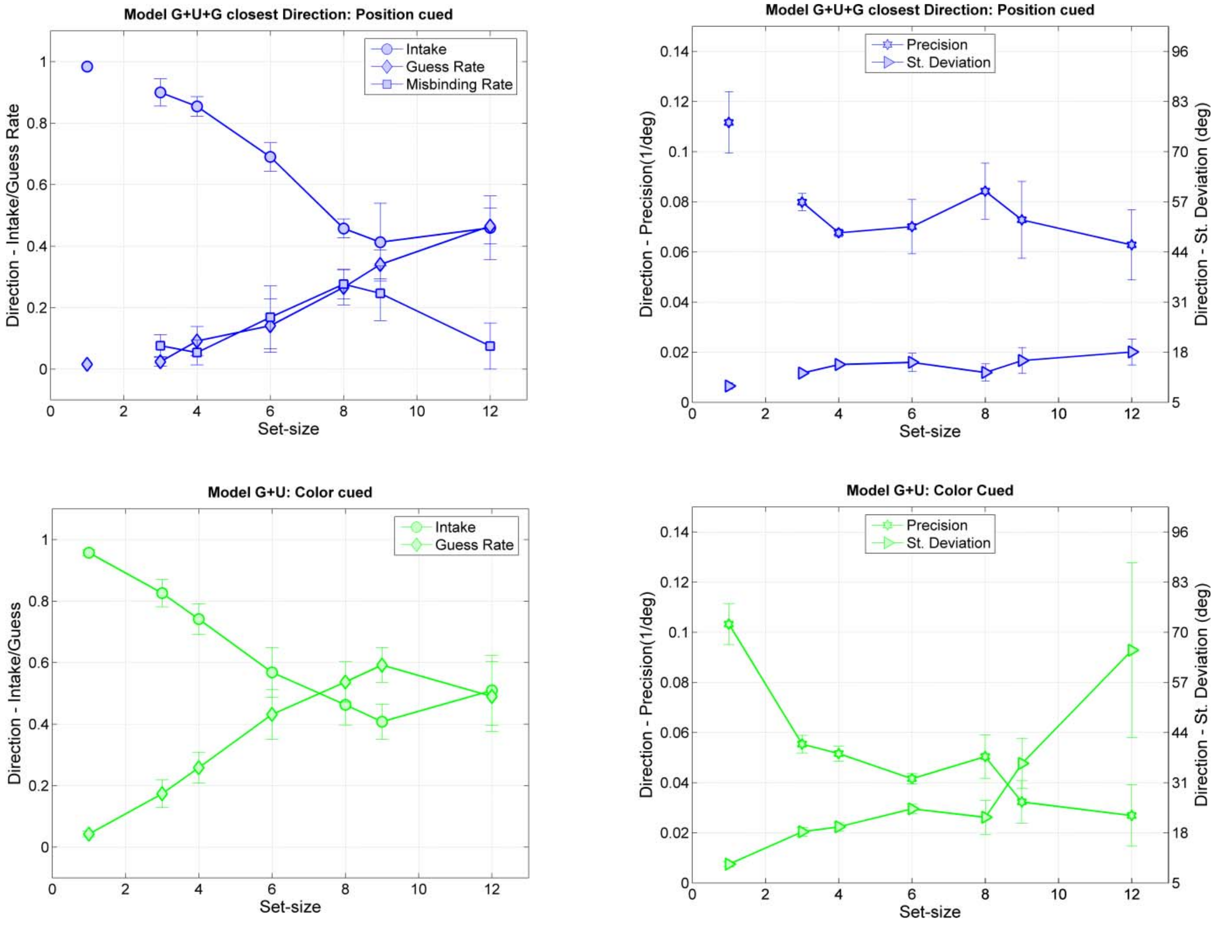


Figure 5. Decomposition of performance for direction of motion in Experiment 1. Left panels show the results for intake (w) along with guessing ($1 - w - w_m$) and misbinding (w_m) rates; right panels show the results for precision ($1/\sigma$: left y-axes) and standard deviation (σ: right y-axes); upper and lower panels show the results for the cases of cuing position and color, respectively; and x-axes represent set size. Error bars correspond to ±1 standard error of the mean. Data for each cue type are shown for the winning models (see top of each panel). In cases that either version of Model 3 won, data shown for a set size of 1 are taken from Model 2 (shown as an isolated data point at a set size of 1), because Model 3 is not applicable.

Experiment 2: Iconic memory and VSTM

Analysis of performance

Using the same procedure as in the previous experiment but with a varying-delay interval inserted between the termination of motion and the onset of the cue, this experiment allows us to investigate processing of information for different feature dimensions in subsequent memory stages. Figure 7 plots transformed performance TP (left y-axis) and magnitude of error $|\varepsilon|$ (right y-axis) averaged across observers as a function of cue delay and cue type for reporting position (left), direction of motion (middle), and color (right). For individual data, see Supplementary Figure S1.2. A two-way repeated-measures ANOVA shows that the main effect of cue delay is significant for each of the three features—position reported (with Huynh–Feldt correction for sphericity): $F(2.26, 6.78) = 9.967$, $p = 0.009$, $\eta_p^2 = 0.769$; direction of motion reported: $F(6, 18) = 15.271$, $p < 0.0001$, $\eta_p^2 = 0.836$; color reported: $F(6, 18) = 17.793$, $p < 0.0001$, $\eta_p^2 = 0.856$—whereas the main effect of cue type is marginally significant for position, $F(1, 3) = 9293$, $p = 0.055$, $\eta_p^2 = 0.756$; not significant for direction of motion, $F(1, 3) = 3.944$, $p = 0.141$, $\eta_p^2 = 0.568$; and significant for color, $F(1, 3) = 730.980$, $p < 0.0001$, $\eta_p^2 = 0.996$. The interaction between cue delay and cue type is not significant—position reported (with Huynh–Feldt correction for sphericity): $F(3.585, 10.756) = 0.805$, $p = 0.537$, $\eta_p^2 = 0.211$; direction of motion reported: $F(6, 18) = 2.385$, $p = 0.071$, $\eta_p^2 = 0.443$; color reported: $F(6, 18) = 0.300$, $p = 0.929$, $\eta_p^2 = 0.091$.

To further investigate how features are temporally related, we demarcated sensory memory from VSTM and analyzed the effect of cue type in each of these processing stages. One way to demarcate between the two memory systems is to fit to the empirical data an exponentially decaying model (Shooner et al., 2010; Öğmen et al., 2013). In this method, the duration of

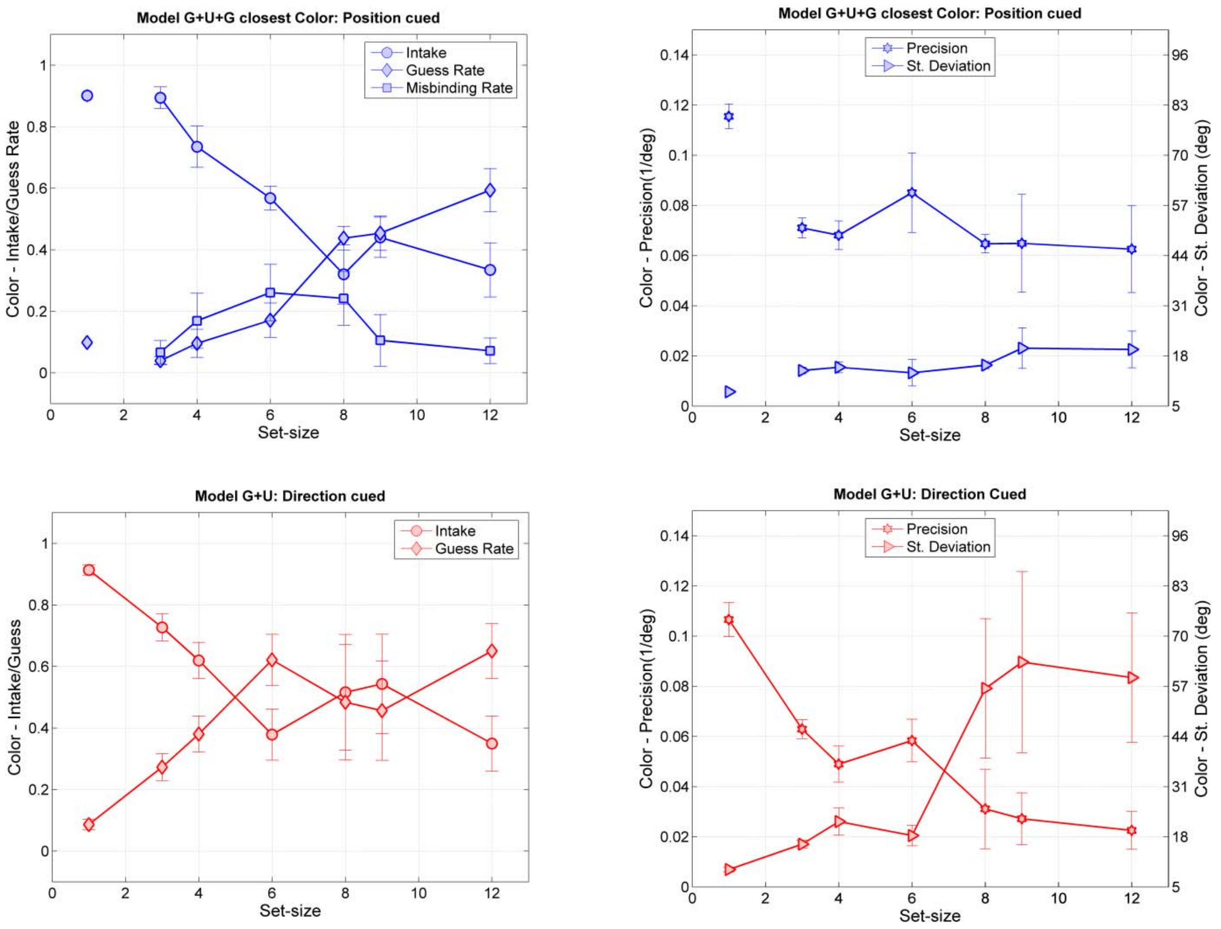


Figure 6. Decomposition of performance for color in Experiment 1. Left panels show the results for intake (w) along with guessing ($1 - w - w_m$) and misbinding (w_m) rates; right panels show the results for precision ($1/\sigma$: left y-axes) and standard deviation (σ: right y-axes); upper and lower panels show the results for the cases of cuing position and direction of motion, respectively; and x-axes represent set size. Error bars correspond to ±1 standard error of the mean. Data for each cue type are shown for the winning models (see top of each panel). In cases that either version of Model 3 won, data shown for a set size of 1 are taken from model 2 (shown as an isolated data point at a set size of 1), because Model 3 is not applicable.

the model's initial transient phase represents the lifetime of sensory memory, and VSTM takes place (or, perhaps, starts to dominate) when the model enters its steady phase. Our data, however, do not show a consistent exponential trend across subjects and conditions (see Supplementary Figure S1.2).[3] Therefore, we employed another demarcation approach in which exponential fits were not implemented. Transformed performance values (individual data) at different time samples were grouped into

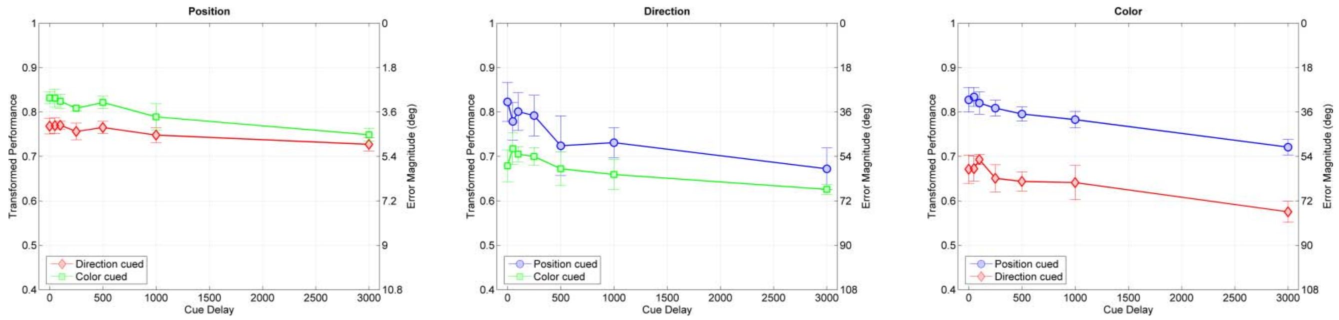


Figure 7. Transformed performance and error magnitude in Experiment 2, averaged across observers, as a function of cue delay for position (left panel), direction of motion (middle panel), and color (right panel). Each symbol color corresponds to a cue type (blue = cuing position, red = cuing direction of motion, magenta = cuing color). Error bars correspond to ±1 standard error of the mean.

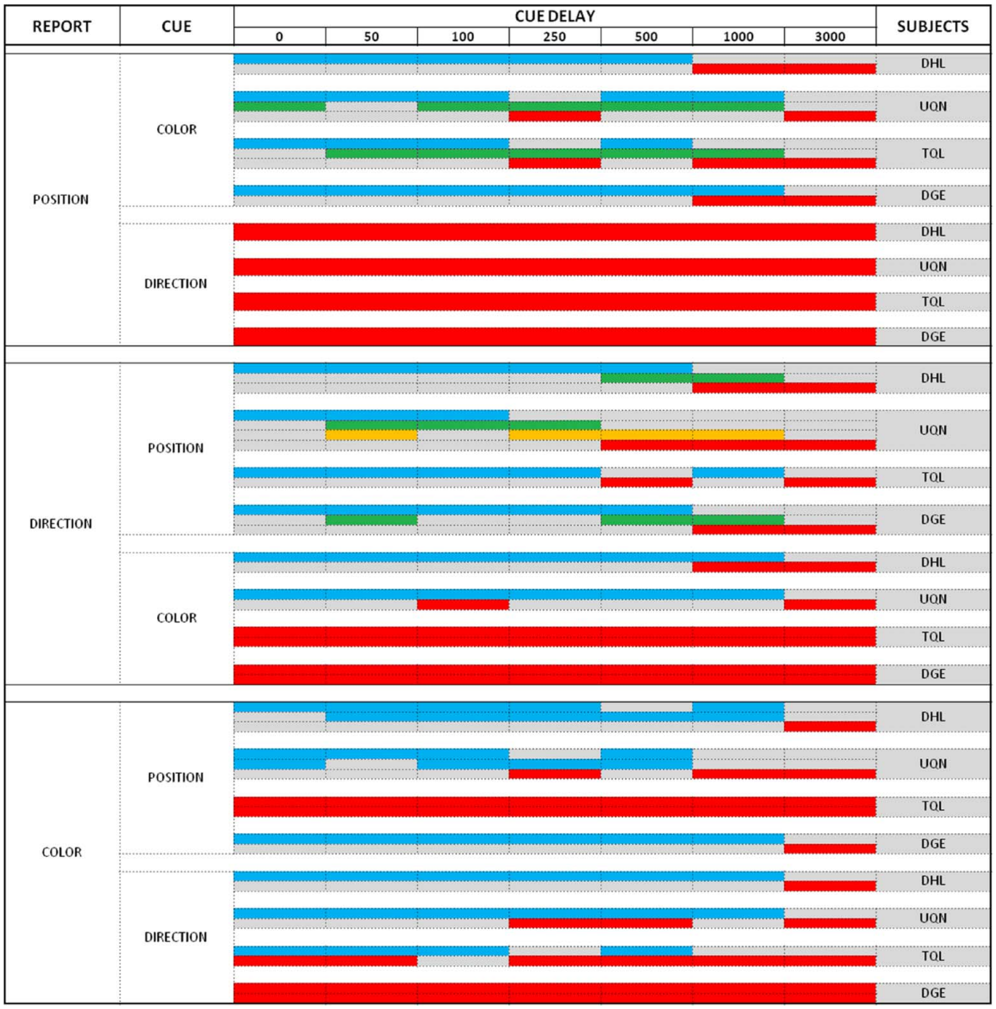
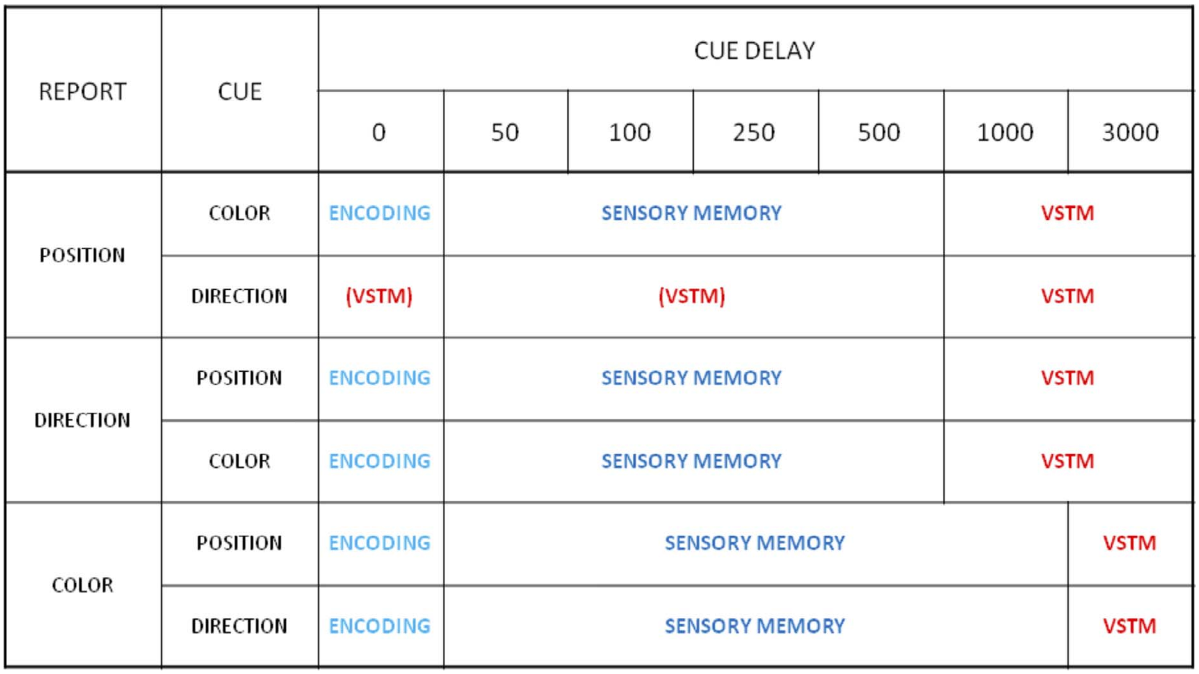


Table 1. Duncan's multiple-range test results in Experiment 2. For each observer, different colors (blue, green, yellow, and red) are used to represent different homogeneous subsets. Time samples belonging to the same subset as 3 s are always shown in red, while other colors are used for the remaining groups.

homogeneous subsets using Duncan's multiple-range test. Members of the same subsets have means that do not differ significantly from one another. In order to determine which subset corresponds to which memory stage, we used the following rationale. Our temporal samples for the cue delay consisted of the following values: 0, 50, 100, 250, 500, 1000, and 3000 ms. In general, the duration of sensory memory depends on stimulus parameters, and one cannot use an a priori fixed value to decide at which cue delay sensory memory has completely decayed (Coltheart, 1980). However, based on the large literature, one can assert that the duration of sensory memory is definitely shorter than 3 s (Sperling, 1960; Dick, 1974; Coltheart, 1980; Shooner et al., 2010). Hence, cue delays that belong to the same subset as 3 s are taken to represent VSTM. Those nonzero cue delays that fell into subsets that were significantly different from the one containing 3 s were taken to belong to sensory memory. In doing this, we have assumed that when performance does not reach an asymptote before or around 1 s but continues to drop, the asymptote is presumed to occur at some point between 1 and 3 s. When the effect of cue delay is not significant (flat data from 0 to 3 s), performance can be said to have dropped to the VSTM level at the very beginning. One explanation for this would be that at a set size of 6, a significant portion of information is already lost at the stimulus-encoding stage (see Figure 2) and the cued attribute being used does not provide a fast direct access to the remaining information. In other words, by the time the cue allowed access to the required information, nearly all of the additional information available during the initial encoding and sensory memory stages had faded away.



REPORT	CUE	CUE DELAY						
		0	50	100	250	500	1000	3000
POSITION	COLOR	ENCODING	SENSORY MEMORY				VSTM	
	DIRECTION	(VSTM)	(VSTM)				VSTM	
DIRECTION	POSITION	ENCODING	SENSORY MEMORY				VSTM	
	COLOR	ENCODING	SENSORY MEMORY				VSTM	
COLOR	POSITION	ENCODING	SENSORY MEMORY					VSTM
	DIRECTION	ENCODING	SENSORY MEMORY					VSTM

Table 2. Demarcation for different processing stages based on Duncan's test (individual data) and visual inspection of averaged data (Figure 9).

Although the multiple-range test was carried out on individual data to examine the idiosyncratic behavior of each subject (Table 1), our interpretation was based on both these test results and the general trend reflected by averaged data to minimize the noise in grouping. Consideration of the averaged data is especially necessary when the statistical-grouping results are not identical across subjects. For example, when position or direction of motion is cued and color is reported, the results in Figure 7 suggest that only performance at 3 s would correspond to VSTM. However, this is not the case for all subjects, as shown in Table 1. Table 2 shows our demarcation between sensory memory and VSTM for each cue–report combination. The effect of cue type for each memory system was then investigated accordingly using paired-samples *t*-tests with Bonferroni correction for multiple comparisons (two-tailed; $\alpha = 0.0167$; transformed performance at time samples attributed to the same memory systems were combined). We observed that the effect of cue type remains significant over time for color but vanishes for position and direction of motion. In the case of position, the vanishing occurs at VSTM, whereas for motion it already occurs in sensory memory (Table 3). Based on these changes in the effect of cue type, we elaborated the diagram in Figure 3 to represent the relationships between features in sensory memory and VSTM. The results for all three processing stages are reproduced in Figure 8, with the following implications: First, there is an asymmetry in the reciprocal relations between color and direction of motion, and this asymmetry persists in all three stages, with color being a more effective cue for direction of motion than vice versa.[4] Second, position systematically loses its advantage as the most efficient cue as one progresses from stimulus encoding to VSTM. Third, the reciprocal relation between color and direction of motion remains weak in all three stages. Hence, if the interpretation of this in terms of within-pathway bindings and associations is correct, then this pathway specificity is present not

	Memory system			
	Sensory memory		VSTM	
Report	$t(3)$	p	$t(3)$	p
Position	3.805	0.032	1.324	0.277
Direction	1.566	0.215	1.713	0.185
Color	20.156	$<$0.0001	6.771	0.007

Table 3. Paired-samples *t*-test results for the effect of cue type during the sensory-memory and VSTM stages.

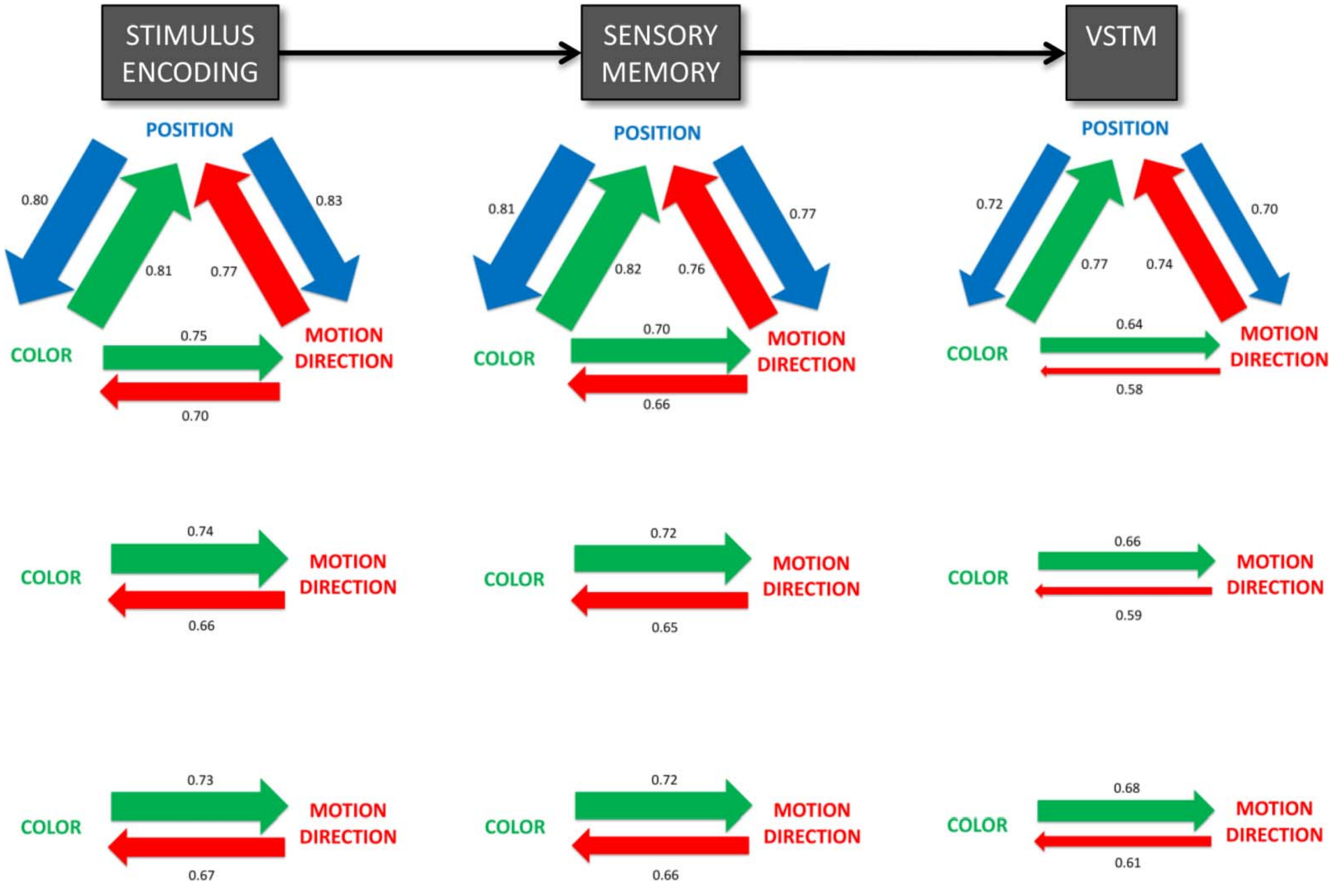


Figure 8. Top: Diagram showing the effectiveness for each cue–report combination for all three stages. Conventions are the same as those in Figure 3. The middle and bottom panels show cue effectiveness for the two replicated (middle) and control (bottom) conditions presented in Supplementary information 3.

just at the initial stimulus processing and encoding stage but also in the subsequent memory stages.

Statistical modeling

In Experiment 1, a bottleneck of processing was observed at the encoding stage for all features, as reflected by the degradation of performance with increasing set size. Analyzing the extent to which these performance degradations change over time provides information about the distribution of information loss across different processing stages. Figures 9 through 11 plot precision and intake as a function of cue delay and cue type for position (Figure 9a: horizontal component; Figure 9b: vertical component), direction of motion, and color, respectively. In each figure, the left panels show the results for intake w, the right panels show the results for precision $1/\sigma$, and the upper and lower panels with different marker colors show the results for two different cue types. Similar to Experiment 1, only the winning model in each cue–report condition was selected. Results of the model-selection analyses are shown in Tables 4 and 5. As mentioned earlier, the flat performance at a set size of 1 forms the baseline for our analysis and is represented by the horizontal lines extended from the single data points (at cue delay = 0 s) in Figures 9 through 11. These data points correspond to the intake or precision obtained in the case of a single object in Experiment 1. Consistently across features, we find that the major bottleneck for the *quality* of information (precision) resides at the stimulus-encoding stage rather than memory. At least 62% (78% with Bayesian analysis; see Supplementary information 2) of the total precision decay occurs during encoding. This replicates the finding for cuing position and reporting the direction of motion from Öğmen et al. (2013). The bottleneck for the *quantity* of information (intake) shows a wider range. The drop of intake at the encoding stage is also substantial in most cases ($\geq$66%), with its lowest value being (according to the least-squares analysis) 39% when position is cued and direction of motion is reported and (according to Bayesian analysis) 49% when position is cued and color is reported. Overall, these results indicate significant loss of information at the encoding stage both in terms of quality and quantity.

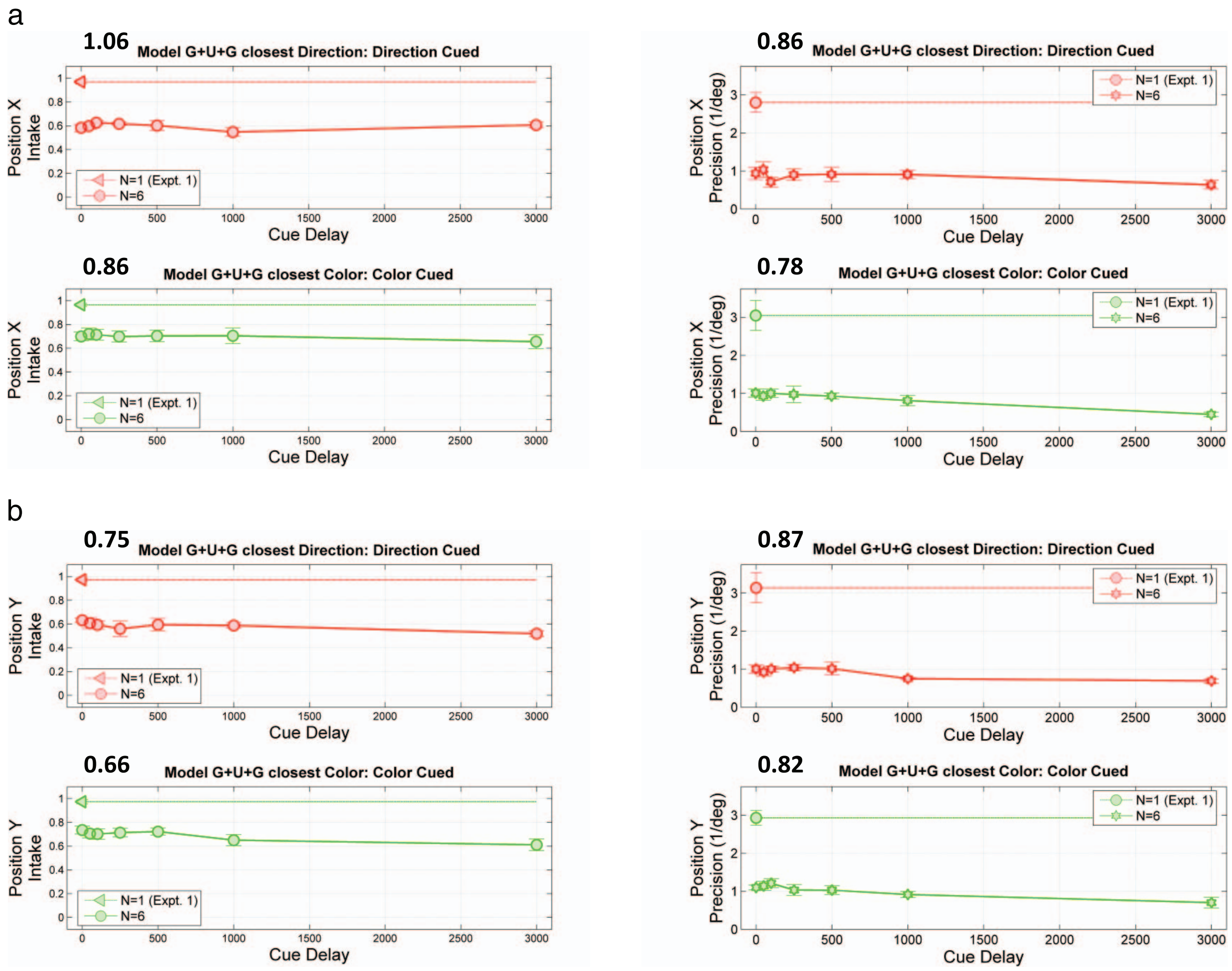


Figure 9. (a) Decomposition of performance for the horizontal component (X) of position in Experiment 2 (set size fixed at 6): Intake (w: left panels) and precision ($1/\sigma$: right panels) as a function of cue delay and cue type; upper and lower panels show the results for the cases of cuing direction of motion and color, respectively. Error bars correspond to ±1 standard error of the mean. Data for each cue type are shown for the winning models (see top of each panel). Data at a set size of 1, shown only at cue delay = 0 s, are taken from Experiment 1. Horizontal lines are to indicate that performance in this case is largely independent of cue delay. The numerical value above each panel indicates information loss for $N = 6$ targets during the encoding stage; this value is calculated as the ratio of the drop of intake or precision at cue delay = 0 s to that at cue delay = 3 s. Note that the value of 1.06 would imply a gain of information; however, this value is within the error range and hence should not be interpreted as information gain. (b) Decomposition of performance for the vertical component (Y) of position in Experiment 2 (set size fixed at 6). Same conventions as (a).

General discussion

Feature binding and content-addressable access

The visual system processes features of stimuli in specialized areas and pathways. However, these features need to be bound together in order to construct unified object representations. In addition, because human memory is content addressable, the binding of features also plays a critical role in memory access. By using a cross-cuing technique, we examined how features are bound together and how they can effectively allow access to each other during the initial stages of stimulus encoding as well as during subsequent stages of sensory memory and VSTM. Previous research has suggested that position is the index used to build and maintain unified object representations (Pylyshyn & Storm, 1988; Kahneman et al., 1992). While we found position to be a more effective cue, and hence index, for accessing other features in the initial stimulus-encoding stage, this privileged role is lost in sensory memory and in VSTM. In fact, in VSTM, color and direction of motion are more effective cues for

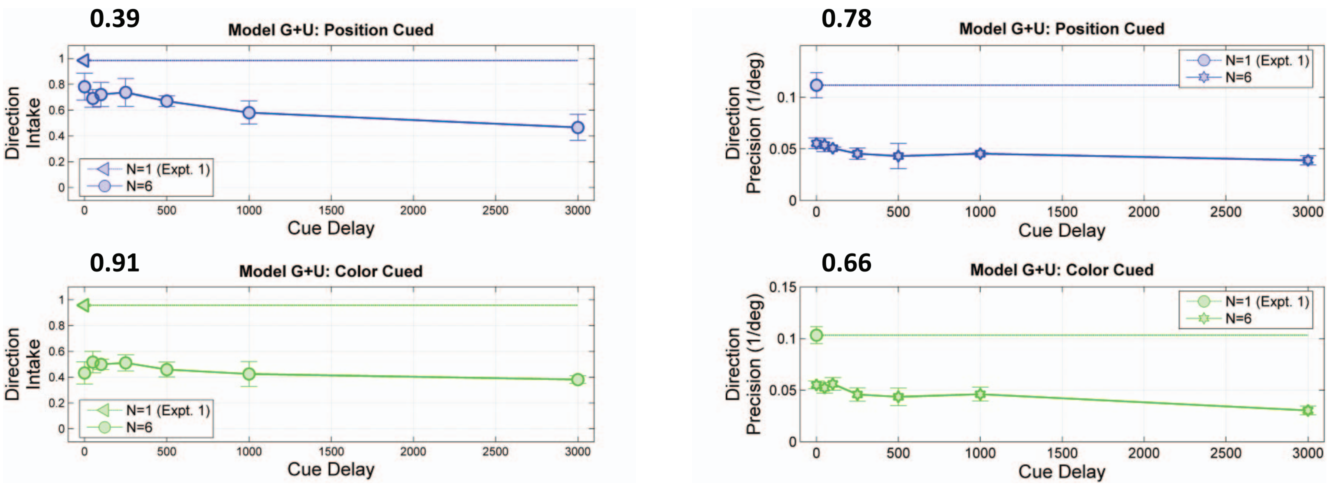


Figure 10. Decomposition of performance for direction of motion in Experiment 2 (set size fixed at 6): Intake (*w*: left panels) and precision ($1/\sigma$: right panels) as a function of cue delay and cue type; upper and lower panels show the results for the cases of cuing position and color, respectively. Error bars correspond to ±1 standard error of the mean. Parameters for each cue type are shown for the winning models (see top of each panel). Data at a set size of 1, shown only at cue delay = 0 s, are taken from Experiment 1. Horizontal lines are to indicate that performance in this case is largely independent of cue delay. The numerical value above each panel indicates information loss for $N = 6$ targets during the encoding stage; this value is calculated as the ratio of the drop of intake or precision at cue delay = 0 s to that at cue delay = 3 s.

position than vice versa. On the other hand, a characteristic that is present in all three stages of information processing is the effectiveness of binding and memory access that mirrors the parallel processing streams of the visual system: Within-stream mutual couplings (between position and color in the ventral stream, between position and direction of motion in the dorsal stream) are much stronger than the mutual coupling across streams (between color and direction of motion). Taken together, these results suggest that while the visual system effectively binds features of an object to construct unified object representations, the

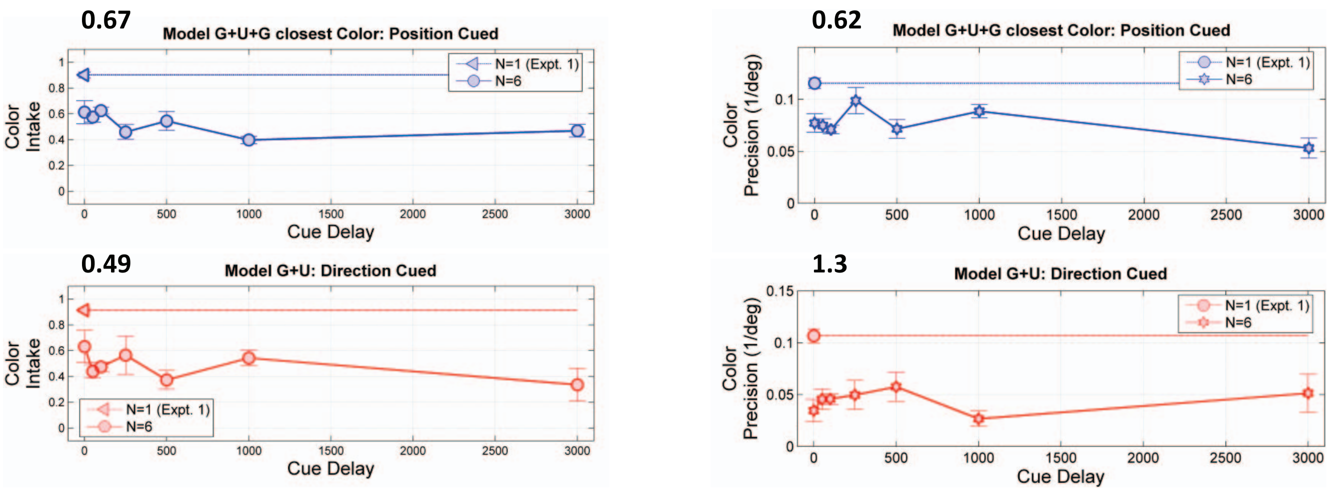


Figure 11. Decomposition of performance for color in Experiment 2 (set size fixed at 6): Intake (*w*: left panels) and precision ($1/\sigma$: right panels) as a function of cue delay and cue type; upper and lower panels show the results for the cases of cuing position and color, respectively. Error bars correspond to ±1 standard error of the mean. Data for each cue type are shown for the winning models (see top of each panel). Data at a set size of 1, shown only at cue delay = 0 s, are taken from Experiment 1. Horizontal lines are to indicate that performance in this case is largely independent of cue delay. The numerical value above each panel indicates information loss for $N = 6$ targets during the encoding stage; this value is calculated as the ratio of the drop of intake or precision at cue delay = 0 s to that at cue delay = 3 s. Note that the value 1.3 would imply a gain of information; however, this value is within the error range and hence should not be interpreted as information gain.

Experiment	Report	Cue	Mode with largest adjusted R^2
1	Position	Color	M3c
		Direction	M3c
	Direction	Position	M3r
		Color	M2
	Color	Position	M3r
		Direction	M2
2	Position	Color	M3c
		Direction	M3c
	Direction	Position	M2
		Color	M2
	Color	Position	M3r
		Direction	M2

Table 4. Model-selection results according to least-squares fitting method (adjusted R^2). A repeated-measures ANOVA shows that the main effect of model (M1, M2, M3c, M3r) is significant in all cases. The winning model was selected for each case based on follow-up planned comparisons between models. *Notes*: M1 = Gaussian. M2 = Gaussian + Uniform. M3c = Gaussian + Uniform + Gaussian *closest cued feature*. M3r = Gaussian + Uniform + Gaussian *closest reported feature*.

binding operations and the associated memory access are carried out primarily in specialized processing streams and hence reflect stream-specific quantitative differences.

Previous studies have also shown a relationship between the effectiveness of interactions between features and their putative neural correlates. For example, Kristjánsson (2006) studied priming along three feature dimensions—color, spatial frequency, and orientation—and showed that while spatial frequency and orientation did interact with each other, color did not interact with the other two features. One possible interpretation of these findings is that spatial frequency and orientation share neural correlates, whereas color is processed by largely independent mechanisms (Kristjánsson, 2006). Fougnie and Alvarez (2011) studied how color and orientation are stored in working memory and reported that recall errors for these two features are largely independent.[5] They suggested that the extent of overlap in neural coding for different features determines the degree of independence in their storage. The distinction between overlapping versus nonoverlapping neural correlates was made in these studies within the ventral stream (color vs. orientation/spatial frequency), whereas in our study we made the distinction by contrasting ventral and dorsal pathways. Notwithstanding this difference, a common theme that emerges from these previous investigations and our study is the influence of the neural architecture and correlates in determining the strength of associations or interactions between features.

Bottlenecks of information processing

Figure 12a depicts the traditional view of bottlenecks in visual processing. The stimulus-encoding stage is parallel and hence assumed to be of large capacity. Similarly, the traditional view of sensory memory is a large-capacity preattentive store, the contents of which decay rapidly. The hourglass analogy shows an initial large-capacity stimulus processing and encoding stage followed by a large-capacity sensory memory. The leak in the sensory-memory stage represents the rapid loss of information. Long-term memory is also thought to be of large capacity; hence the main bottleneck occurs in

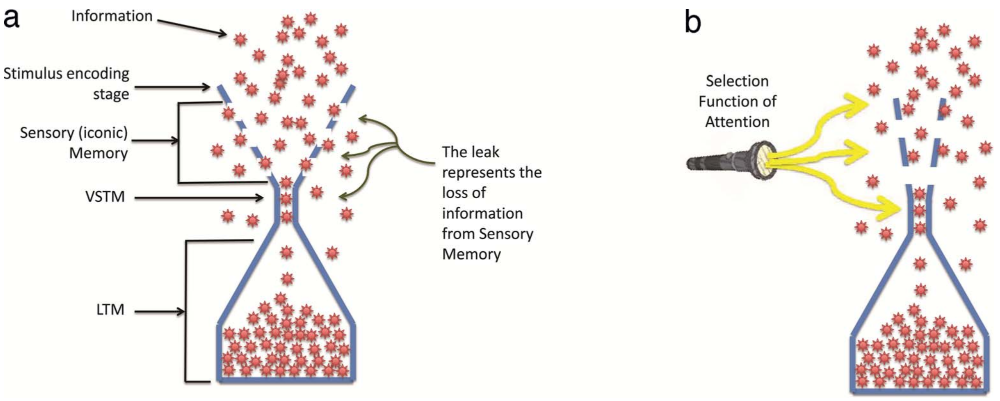

a
Information
Stimulus encoding stage
Sensory (iconic) Memory
VSTM
LTM
The leak represents the loss of information from Sensory Memory
b
Selection Function of Attention

Figure 12. (a) The Leaky Hourglass model of information-processing bottlenecks. According to this model the main bottleneck resides in VSTM. (b) The Leaky Flask model proposed by Öğmen et al. (2013). This model proposes significant bottlenecks already at the stimulus-encoding stage, and that attention interacts at all levels, including the sensory-memory stage. (Adapted from Öğmen H., Ekiz, O., Huynh, D., Tripathy, S. P., & Bedell, H. E. (2013). Bottlenecks of motion processing during a visual glance: The Leaky Flask model. *PLOS One*, *8*(12), e83671).

Report	Position *X*								Position *Y*							
Cue	Direction				Color				Direction				Color			
Model	M1	M2	M3c	M3r	M1	M2	M3c	M3r	M1	M2	M3c	M3r	M1	M2	M3c	M3r
Experiment 1 (Set size = *N*)																
$N = 3$	0.9946	0.9966	0.9971	0.9972	0.995	0.9959	0.996	0.996	0.9926	0.9954	0.9964	0.9965	0.9959	0.997	0.9959	0.997
$N = 4$	0.9882	0.9944	0.9964	0.996	0.9922	0.9954	0.9958	0.9962	0.9938	0.9963	0.9971	0.997	0.9935	0.9962	0.9971	0.997
$N = 6$	0.9784	0.9911	0.9928	0.9955	0.9804	0.9915	0.9971	0.9969	0.9846	0.9914	0.9957	0.9957	0.9815	0.9906	0.9964	0.9967
$N = 8$	0.9835	0.9893	0.9956	0.9903	0.9796	0.9901	0.9933	0.9919	0.9875	0.9901	0.9951	0.9911	0.982	0.9887	0.9933	0.9915
$N = 9$	0.9876	0.9911	0.9962	0.9924	0.9712	0.9877	0.9929	0.9889	0.9901	0.9931	0.9956	0.9948	0.9798	0.99	0.9955	0.9929
$N = 12$	0.9875	0.9883	0.989	0.9882	0.9809	0.9842	0.9901	0.9841	0.9951	0.9954	0.9961	0.9955	0.9833	0.9873	0.9955	0.9882
Experiment 2 (Cue delay = *D* ms)																
$D = 0$	0.9803	0.9896	0.9946	0.9941	0.9827	0.9916	0.995	0.9948	0.9845	0.9931	0.9966	0.997	0.9867	0.9921	0.9942	0.9936
$D = 50$	0.9783	0.9908	0.9951	0.996	0.9818	0.9879	0.9901	0.9896	0.9853	0.9923	0.9954	0.9954	0.9891	0.9946	0.9968	0.9961
$D = 100$	0.9849	0.9934	0.9965	0.9961	0.9846	0.9926	0.9957	0.995	0.9861	0.9924	0.9958	0.9942	0.9848	0.9914	0.9937	0.9941
$D = 250$	0.9826	0.9939	0.9973	0.996	0.9827	0.9928	0.9957	0.994	0.9805	0.9901	0.9959	0.9939	0.9888	0.9941	0.9965	0.9964
D = 500	0.9786	0.9898	0.995	0.9917	0.9834	0.9906	0.9955	0.9948	0.9824	0.9919	0.9969	0.9956	0.989	0.9946	0.9969	0.9969
$D = 1000$	0.9766	0.9905	0.997	0.9941	0.981	0.9922	0.994	0.9942	0.987	0.993	0.9973	0.9963	0.9871	0.9938	0.9972	0.9968
$D = 3000$	0.9835	0.9925	0.9962	0.9949	0.9853	0.9911	0.9922	0.9928	0.9845	0.9895	0.9935	0.9929	0.9885	0.9931	0.9954	0.9955

Table 5. Adjusted R^2 values averaged across observers for position (a), color (b), and direction of motion (c). The notation for the models is the same as in Table 4.

	Color								Direction							
	Position				Direction				Position				Color			
	M1	M2	M3c	M3r	M1	M2	M3c	M3r	M1	M2	M3c	M3r	M1	M2	M3c	M3r
$N = 3$	0.9959	0.9973	0.9973	0.9976	0.9864	0.9959	0.9964	0.9961	0.9945	0.9966	0.9966	0.9967	0.9927	0.9962	0.9963	0.9964
$N = 4$	0.9923	0.9969	0.9969	0.9973	0.9851	0.9961	0.9963	0.9968	0.9953	0.9974	0.9974	0.9974	0.9903	0.9969	0.997	0.9973
$N = 6$	0.99	0.9963	0.9963	0.9979	0.9855	0.9969	0.9969	0.9972	0.9894	0.994	0.9938	0.9947	0.9881	0.9967	0.9967	0.9968
$N = 8$	0.985	0.9956	0.996	0.9958	0.9929	0.9941	0.9955	0.9948	0.9874	0.9958	0.996	0.9968	0.9836	0.997	0.997	0.9971
$N = 9$	0.9846	0.9953	0.9952	0.9957	0.9917	0.9962	0.9962	0.9962	0.9882	0.9953	0.9954	0.9958	0.9871	0.9935	0.9945	0.994
$N = 12$	0.9868	0.9972	0.9975	0.9971	0.994	0.9955	0.9954	0.9955	0.9847	0.9973	0.9973	0.9973	0.9893	0.9943	0.9956	0.9952
$D = 0$	0.9896	0.9972	0.9974	0.9979	0.9848	0.9946	0.9956	0.995	0.9891	0.9962	0.9965	0.9962	0.9832	0.9976	0.998	0.9976
$D = 50$	0.9908	0.9964	0.9964	0.9975	0.9856	0.9964	0.9972	0.9969	0.9886	0.9955	0.9955	0.9955	0.985	0.9974	0.9975	0.998
$D = 100$	0.9868	0.9965	0.9968	0.9969	0.9832	0.9969	0.9972	0.9969	0.9895	0.9963	0.9967	0.9966	0.9792	0.9973	0.9974	0.9976
$D = 250$	0.988	0.995	0.9953	0.9966	0.9828	0.9961	0.9961	0.9962	0.9927	0.997	0.9973	0.997	0.9835	0.9981	0.9981	0.9983
$D = 500$	0.9904	0.9976	0.9976	0.9982	0.9849	0.9972	0.9974	0.9974	0.9901	0.995	0.995	0.9951	0.9836	0.9948	0.9966	0.9951
$D = 1000$	0.9882	0.9953	0.9955	0.9967	0.9904	0.9957	0.9958	0.9957	0.9873	0.9957	0.996	0.9958	0.985	0.9964	0.9964	0.9964
$D = 3000$	0.984	0.9955	0.9958	0.9956	0.9884	0.9935	0.9934	0.9934	0.9873	0.9963	0.9968	0.9962	0.9919	0.9963	0.9969	0.9966

Table 5. Extended.

VSTM, as depicted in the hourglass analogy. In our recent work, in which we examined capacity limits for direction of motion, we found significant limits already at the stimulus-encoding stage (Öğmen et al., 2013). This prompted us to modify the Leaky Hourglass model and to propose instead the Leaky Flask model shown in Figure 12b, where the narrow rim of the flask represents the limited capacity of the encoding stage. In this model, attention was shown to play a role in all three stages of information processing. The involvement of attention in sensory memory is also supported by recent imaging (Ruff, Kristjánsson, & Driver, 2007) and psychophysical studies (Persuh, Genzer, & Melara, 2012). The results presented here generalize the Leaky Flask model to color and position, in addition to direction of motion.

Keywords: visual memory, feature binding, content-addressable memory, Leaky Flask

Acknowledgments

Commercial relationships: none.

Corresponding author: Haluk Öğmen.
Email: ogmen@uh.edu.
Address: Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA.

Footnotes

[1] The arrow provides a simple and direct indication of direction of motion. The reason we used an arrow rather than a moving stimulus is that the latter would require integration time for the determination of direction of motion. Since the delay after stimulus offset is a critical variable in assessing encoding and memory storage, we sought to minimize this additional delay.

[2] Our simulation was for the case of cuing position and reporting direction of motion. However, the outcome should be similar for the other feature combinations.

[3] The experimental design for one cue–report combination in Experiment 2 (i.e., when position is cued and direction of motion is reported) is similar to that in our previous study (Öğmen et al., 2013). The differences in the current study are that (a) no distractors were involved but all objects had the potential to be selected as the probed target and (b) objects were removed from the display before the cue came on. In fact, our average results in Experiment 2 (obtained for set size of 6) are about the same as in the closest condition in the previous study (see figure 6 of Öğmen et al., 2013; middle panel; target set size $T = 5$ and distractor set size $D = 0$). Based on our previous findings, we expected that the drop in performance would be more pronounced and that the steady phase would become more evident if set size were increased (e.g., to 9). However, only a set size of 6 was used throughout Experiment 2 of this study. This was to avoid floor effects, because for set sizes greater than 6, performance in some conditions is near chance level (see Figure 2).

[4] In our experimental design, subjects are presented with color information during the static preview period before motion begins. Hence, one may argue that subjects may be encoding color first, followed by motion, leading to the asymmetry found in the results. To test this hypothesis, we ran the control experiment presented in Supplementary Information 3, in which direction-of-motion information preceded color rather than vice versa. The results of the control experiment indicate that the asymmetry does not result from the temporal order or duration of features.

[5] In Fougnie and Alvarez's study (2011), a position cue was used and observers were asked to sequentially report color and orientation from VSTM. It was shown that in trials where the observer was at chance in reporting one feature, the report for the other feature was better than chance—a finding interpreted as independent storage of features. Independence of feature storage would argue against any form of binding specificity, whether across or within streams. First, let us point out that whereas data clearly refute the extreme form of binding, in which *all* features of an object are *always* bound together, there is ample evidence for feature binding in memory. In fact, our data show that all three features studied here are bound, at least in a pair-wise manner, since any feature can be used to recall any other feature. The apparent conflict between the findings of these studies can be resolved by noting that the transfer of information from iconic memory to VSTM is flexible and depends on task demands (e.g., Gegenfurtner & Sperling, 1993). As discussed under Methods, we used a blocked design to study pair-wise feature binding in its strongest form. Hence our task strongly promoted pair-wise binding of selected features. Fougnie and Alvarez (2011), on the other hand, used position as a cue, with color and orientation as reported features. Hence their task emphasized position–color and position–orientation bindings much more than color–orientation bindings. In fact, their results show that there was still color–orientation binding (features were largely but not completely independent), but this binding was weaker than those with position. With our blocked design, we examined each binding pair in its strongest form.

[6] The model of this form can be considered as a generalization of all models described earlier. Note that, for simplification, we did not include the summation operators to represent wrapped Gaussians in those models. However, as mentioned earlier, the wrapped form of the Gaussian must be used where applicable.

[7] Jensen's inequality: $\ln(\sum_{j=1}^{T} c_j) = \ln(\sum_{j=1}^{T} \frac{c_j}{p_j} p_j) \geq \sum_{j=1}^{T} p_j \ln(\frac{c_j}{p_j})$.

[8] Here, the second-order derivatives are found to be negative in all cases.

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135,* 298–313.

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13):14, 1–10, doi:10.1167/7.13.14. [PubMed] [Article]

Basole, A., White, L. E., & Fitzpatrick, D. (2003). Mapping of multiple features in the population response of visual cortex. *Nature, 423,* 986–990, doi:10.1038/nature01721.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10):7, 1–11, doi:10.1167/9.10.7. [PubMed] [Article]

Boi, M., Öğmen, H., Krummenacher, J., Otto, T. U., & Herzog, M. H. (2009). A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision, 9*(13):5, 1–11, doi:10.1167/9.13.5. [PubMed] [Article]

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28,* 157–189.

Braet, W., & Humphreys, G. W. (2009). The role of reentrant processes in feature binding: Evidence from neuropsychology and TMS on late onset illusory conjunctions. *Visual Cognition, 17,* 25–47.

Cavanagh, P., Holcombe, A. O., & Chou, W. (2008). Mobile computation: Spatiotemporal integration of the properties of objects in motion. *Journal of Vision, 8*(12):1, 1–23, doi:10.1167/8.12.1. [PubMed] [Article]

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*(1), 1–38.

Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. J. (1985). Contour, color, and shape analysis beyond the striate cortex. *Vision Research, 25,* 441–452.

Dick, A. O. (1974). Iconic memory and its relation to perceptual processing and other memory mechanisms. *Perception & Psychophysics, 16,* 575–596.

Dong, Y., Mihalas, S., Qiu, F., von der Heydt, R., & Niebur, E. (2008). Synchrony and the binding problem in macaque visual cortex. *Journal of Vision, 8*(7):30, 1–16, doi:10.1167/8.7.30. [PubMed] [Article]

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex, 7,* 181–192.

Fencsik, D. E., Klieger, S. B., & Horowitz, T. S. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Perception & Psychophysics, 69,* 567–577, doi:10.3758/BF03193914.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature store model. *Journal of Vision, 11*(12):3, 1–12, doi:10.1167/11.12.3. [PubMed] [Article]

Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition, 17*(1–2), 48–66.

Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R., & Singer, W. (2001). Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience, 4*(2), 194–200.

Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 845–866.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338,* 334–337.

Horowitz, T. S., Fine, E. M., Fencsik, D. E., Yurgenson, S., & Wolfe, J. M. (2007). Fixational eye movements are not an index of covert attention. *Psychological Science, 18,* 356–363.

Hyun, J. S., Woodman, G. F., & Luck, S. J. (2009). The role of attention in the binding of surface features to locations. *Visual Cognition, 17,* 10–24.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 9*(4):1, 1–12, doi:10.1167/9.4.1. [PubMed] [Article]

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175–219.

Kawabe, T. (2008). Spatiotemporal feature attribution for the perception of visual size. *Journal of Vision, 8*(8):7, 1–9, doi:10.1167/8.8.7. [PubMed] [Article]

Keane, B. P., & Pylyshyn, Z. W. (2006). Is trajectory extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology, 52*(4), 346–368.

Kristjánsson, Á. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision Research, 46,* 2554–2570.

Lane, A. R., Ball, K., Smith, D. T., Schenk, T., & Ellison, A. (2013). Near and far space: Understanding the neural mechanisms of spatial attention. *Human Brain Mapping, 34,* 356–366.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychonomic Bulletin and Review, 14*(6), 1072–1078.

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey: I. Selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology, 49,* 1127–1147.

Nishida, S., Watanabe, J., Kuriki, I., & Tokimoto, T. (2007). Human visual system integrates color signals along a motion trajectory. *Current Biology, 17*(4), 366–372.

Öğmen, H., Ekiz, O., Huynh, D., Tripathy, S. P., & Bedell, H. E. (2013). Bottlenecks of motion processing during a visual glance: The Leaky Flask model. *PLOS One, 8*(12), e83671.

Öğmen, H., & Herzog, M. H. (2010). The geometry of visual perception and the representation of information in the human visual system. *Proceedings of the IEEE, 98,* 479–492.

Öğmen, H., Otto, T. U., & Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research, 46,* 3234–3242.

Persuh, M., Genzer, B., & Melara, R. D. (2012). Iconic memory requires attention. *Frontiers in Human Neuroscience, 6,* 126, doi:10.3389/fnhum.2012.00126.

Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197.

Reeves, A., Santhi, N., & DeCaro, S. (2005). A random-ray model for speed and accuracy in perceptual experiments. *Spatial Vision, 18,* 73–83.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.

Revonsuo, A., & Newman, J. (1999). Binding and consciousness. *Consciousness and Cognition, 8,* 123–127.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Ruff, C. C., Kristjánsson, Á., & Driver, J. (2007). Readout from iconic memory and selective spatial attention involve similar neural processes. *Psychological Science, 18,* 901–909.

Saiki, J. (2003a). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3*(1):2 6–21, doi:10.1167/3.1.2.

Saiki, J. (2003b). Spatiotemporal characteristics of dynamic feature binding in visual working memory. *Vision Research, 43,* 2107–2123.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., & Tootell, R. B. (1995, May 12). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268,* 889–893.

Shimozaki, S. S., Eckstein, M., & Thomas, J. P. (1999). The maintenance of apparent luminance of an object. *Journal of Experimental Psychology: Human Perception and Performance, 25*(5), 1433–1453.

Shooner, C., Tripathy, S., Bedell, H., & Öğmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision, 10*(6):8, 1–20, doi: 10.1167/10.6.8. [PubMed] [Article]

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1–29.

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität [Experimental investigations of phenomenal identity]. *Psychologische Forschung, 7,* 81–136.

Ternus, J. (1938). The problem of phenomenal identity.

In W. D. Ellis (Ed.), *A source book of gestalt psychology* (pp. 149–160). London: Routledge and Kegan Paul.

Thiele, A., & Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature, 421*(6921), 366–370.

Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982, Nov 26). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science, 218,* 902–904.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cognition, 17*(1–2), 103–119, doi:10.1080/13506280802196451.

Von Der Malsburg, C. (1981). *The correlation theory of brain function.* Göttingen, Germany: Max-Planck Institute for Biophysical Chemistry.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.

Wolfe, J. M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235.

Zeki, S. (1976). The projections to the superior temporal sulcus from areas 17 and 18 in the rhesus monkey. *Proceedings of the Royal Society B, 193,* 199–207.

Supplementary information

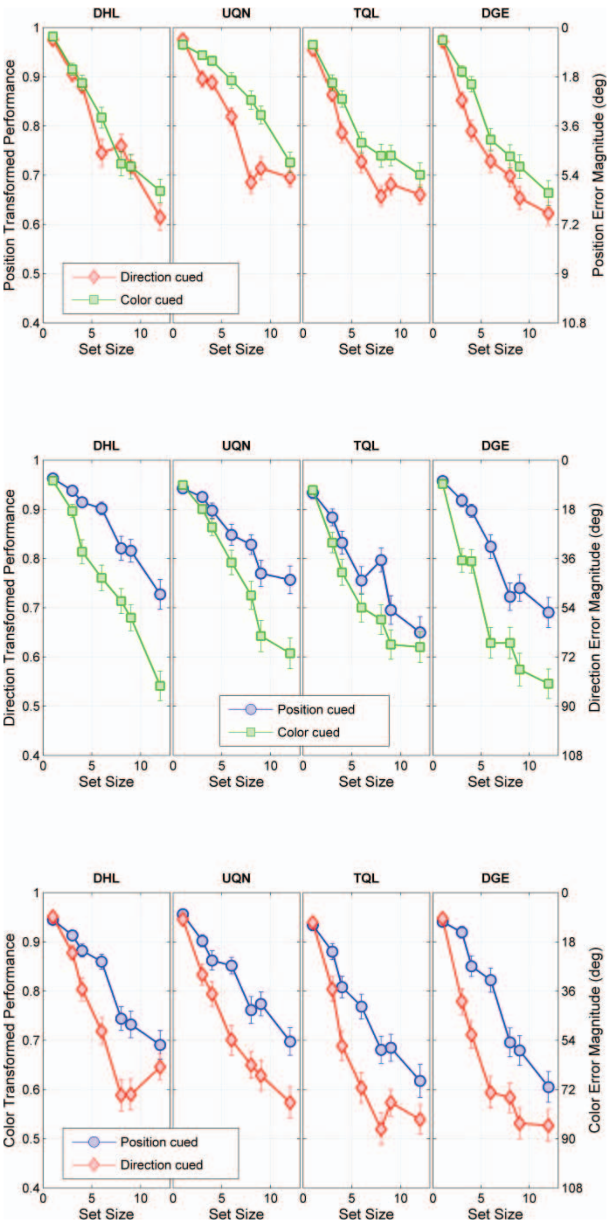
Supplementary information 1

Individual observer data for Experiments 1 and 2 are shown in Supplementary Figures S1.1 and S1.2, respectively.

Supplementary information 2

Bayesian method: Expectation-maximization (EM) algorithm

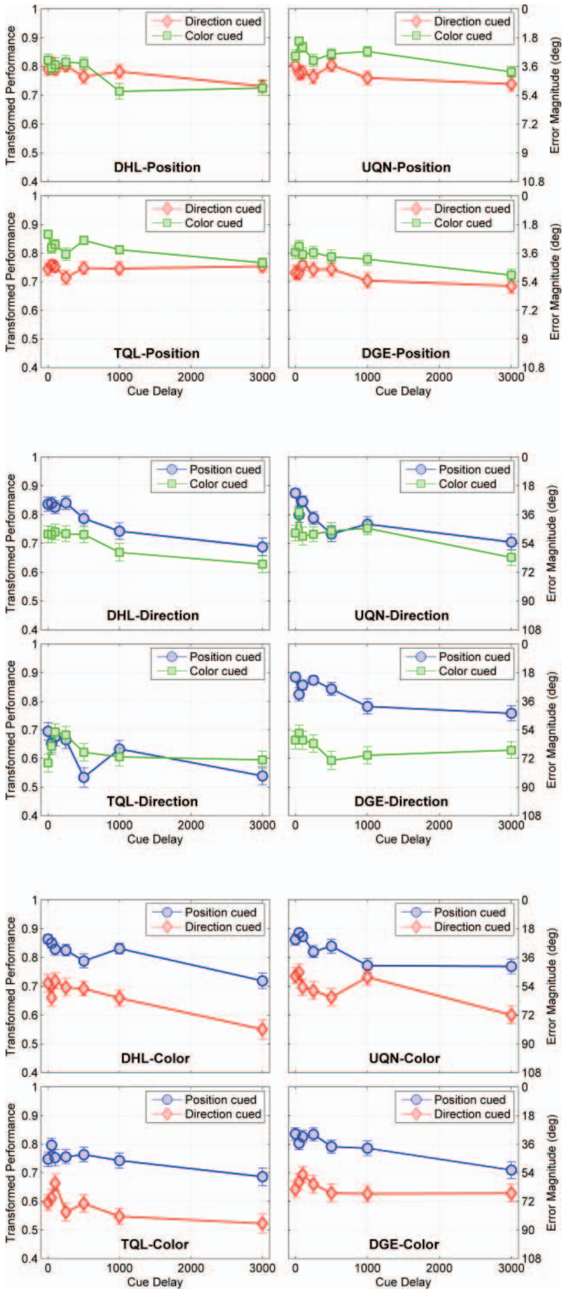
The EM algorithm starts with a certain initial estimate for the parameters whose values will be iteratively updated by means of two alternate steps until convergence is observed.



Supplementary Figure S1.1. Transformed performance (left y-axes) and error magnitude (right y-axes) for individual observers ($N = 4$) in Experiment 1, averaged across trials, as a function of set size for position (top), direction of motion (middle), and color (bottom). Each symbol color corresponds to a cue type (blue = cuing position, red = cuing direction of motion, green = cuing color). Error bars correspond to ±1 standard error of the mean for the specified observer.

1) The "E step" is to construct in the parameter space a likelihood (L) function that represents the probability that a given model has generated a set of data points. The expectation of L is then determined by evaluating its logarithm using the current estimate for the parameters.

Assume our model contains a mixture of two wrapped Gaussians and a uniform distribution as follows:[6]



Supplementary Figure S1.2. Transformed performance (left y-axes) and error magnitude (right y-axes) for individual observers ($N=4$) in Experiment 2, averaged across trials, as a function of cue delay for position (top), direction of motion (middle), and color (bottom). Each symbol color corresponds to a cue type (blue = cuing position, red = cuing direction of motion, green = cuing color). Error bars correspond to ±1 standard error of the mean.

$$p(\varepsilon) = w_1 . \sum_{m=-\infty}^{+\infty} G(\varepsilon; \mu_1 + m2\pi, \sigma_1) + w_2 . \sum_{n=-\infty}^{+\infty} G(\varepsilon; \mu_2 + n2\pi, \sigma_2) + w_3 . U(-180, 180), \tag{6}$$

where we have a set of seven parameters $\{w_1, w_2, w_3, \mu_1, \mu_2, \sigma_1, \sigma_2\}$, each of which has the same meaning as elaborated in the Statistical modeling section. The first three parameters are not independent of each other but sum to 1 ($w_1 + w_2 + w_3 = 1$). Similarly, μ_2 differs from μ_1 by the difference in the reported feature space between the cued target and the misbinding object (see Model 3 under Data analysis). We therefore substitute μ for μ_1 and let d_i be the difference between μ_1 and μ_2 on trial i. We also assume that $\sigma_1 = \sigma_2 = \sigma$ because subjects did not know whether they were reporting the target or a nontarget object on each trial.

Assume also that errors ε are produced independently across trials. From this, the likelihood function can be written as

$$L = \prod_{i=1}^{N} p(\varepsilon_i), \tag{7}$$

where N is the number of trials.

2) The "M step" is to find the optimal values for the parameters in the model, which are ones that maximize the L function.

To do that, we first take the logarithm of L:

$$\ln(L) = \ln\left(\prod_{i=1}^{N} p(\varepsilon_i)\right) = \sum_{i=1}^{N} \ln[p(\varepsilon_i)], \tag{8}$$

or

$$\ln(L) = \sum_{i=1}^{N} \ln\left[w_1 . \sum_{j=-1}^{1} G_1^j + w_2 . \sum_{j=-1}^{1} G_2^j + w_3 . U\right], \tag{9}$$

where

$$G_1^j = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\varepsilon_i - \mu - j2\pi)^2}{2\sigma^2}}$$

and

$$G_2^j = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\varepsilon_i - \mu - d_i - j2\pi)^2}{2\sigma^2}}$$

($j = -1, 0, 1$). For simplicity, we have dropped in Equation 8 the arguments of the Gaussian and uniform distributions, reduced the number of Gaussians of each component to three (for the reasons provided under Data analysis), and used the subscripts of the Gaussians to differentiate the two components of the

Experiment	Report	Cue	Model selection by AIC	Model selection by BIC	Final selection
1	Position	Color	M3c	M3c	M3c
		Direction	M3c	M3c	M3c
	Direction	Position	M2	M2	M2
		Color	M2	M2	M2
	Color	Position	M2	M2	M2
		Direction	M3c	M3c	M3c
2	Position	Color	M3c	M3c	M3c
		Direction	M3c	M3c	M3c
	Direction	Position	M2	M2	M2
		Color	M2	M2	M2
	Color	Position	M3c	M3c	M3c
		Direction	M3c	M2	M3c

Table 6. Model-selection results according to Bayesian method (AIC and BIC). A repeated-measures ANOVA shows that the main effect of model (M1, M2, M3c, M3r) is significant in all cases. The winning model was selected for each case based on follow-up planned comparisons between models. The notation for the models is the same as in Table 4.

model. The function $\ln(L)$ is then evaluated by using Jensen's inequality:[7]

$$\ln(L) \geq \sum_{i=1}^{N} \left\{ \sum_{j=-1}^{1} p^0(G_1^j|\varepsilon_i).ln(\frac{w_1.G_1^j}{p^0(G_1^j|\varepsilon_i)}) + \sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i).ln(\frac{w_2.G_2^j}{p^0(G_2^j|\varepsilon_i)}) + p^0(U|\varepsilon_i).ln(\frac{w_3.U}{p^0(U|\varepsilon_i)}) \right\}, \quad (10)$$

where $p^0(G_1^j|\varepsilon_i), p^0(G_2^j|\varepsilon_i)$, and $p^0(U|\varepsilon_i)$ represent the probabilities that a data point is most likely to be captured by the first Gaussian, the second Gaussian, and the uniform distributions in the model, respectively, given its value ε_i Note that the superscript 0 indicates the current status of the parameters that has sneaked into the inequality in the form of conditional probability. From Bayes's theorem we have

$$p^0\left(G_1^j|\varepsilon_i\right) = \frac{w_1^0.G(\varepsilon_i; \mu^0 + j2\pi, \sigma^0)}{p^0(\varepsilon_i)}, \quad (11)$$

$$p^0\left(G_2^j|\varepsilon_i\right) = \frac{w_2^0.G(\varepsilon_i; \mu^0 + d_i + j2\pi, \sigma^0)}{p^0(\varepsilon_i)}, \quad (12)$$

$$p^0(U|\varepsilon_i) = \frac{w_3^0.U}{p^0(\varepsilon_i)}. \quad (13)$$

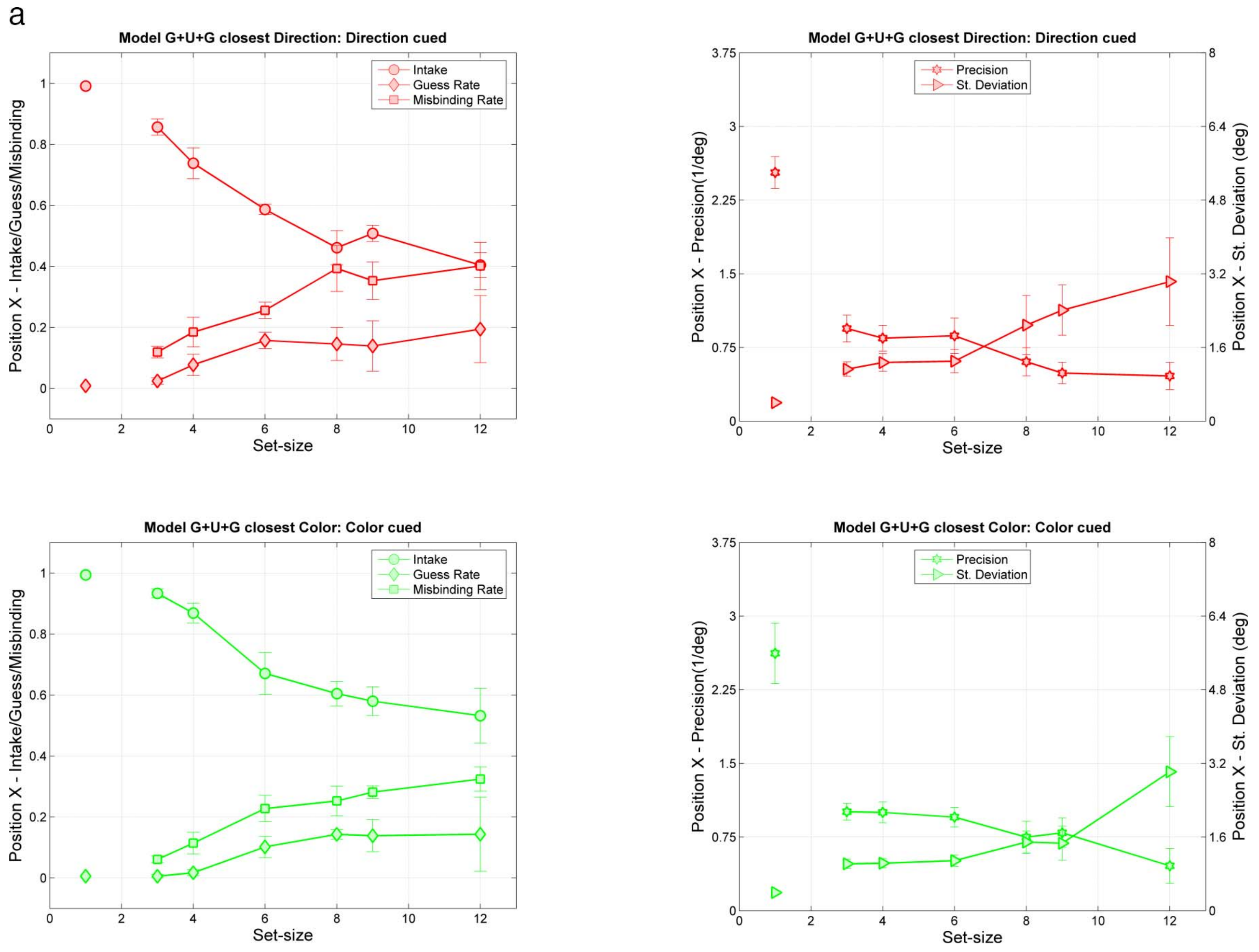
The right-hand side of Equation 10 is the lower bound of $\ln(L)$, so we want to maximize its value. The inequality can be rewritten as

$$\ln(L) \geq \sum_{i=1}^{N} \left\{ \sum_{j=-1}^{1} p^0(G_1^j|\varepsilon_i).\ln(w_1.G_1^j) + \sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i).\ln(w_2.G_2^j) + p^0(U|\varepsilon_i).\ln(w_3.U) \right\} - \sum_{i=1}^{N} \left\{ \sum_{j=-1}^{1} p^0(G_1^j|\varepsilon_i).\ln(p^0(G_1^j|\varepsilon_i)) + \sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i).\ln(p^0(G_2^j|\varepsilon_i)) + p^0(U|\varepsilon_i).\ln(p^0(U|\varepsilon_i)) \right\}. \quad (14)$$

Since the second summation is a constant, the problem boils down to finding the new values for the parameters that maximize the first summation S:

$$S = \sum_{i=1}^{N} \left\{ \sum_{j=-1}^{1} p^0(G_1^j|\varepsilon_i)\ln(w_1 G_1^j) + \sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i)\ln(w_2 G_2^j) + p^0(U|\varepsilon_i)\ln(w_3 U) \right\}. \quad (15)$$

We do so by taking partial derivatives of S with respect to each parameter, setting each derivative equal to 0, and solving the equations.[8] The results are (note that



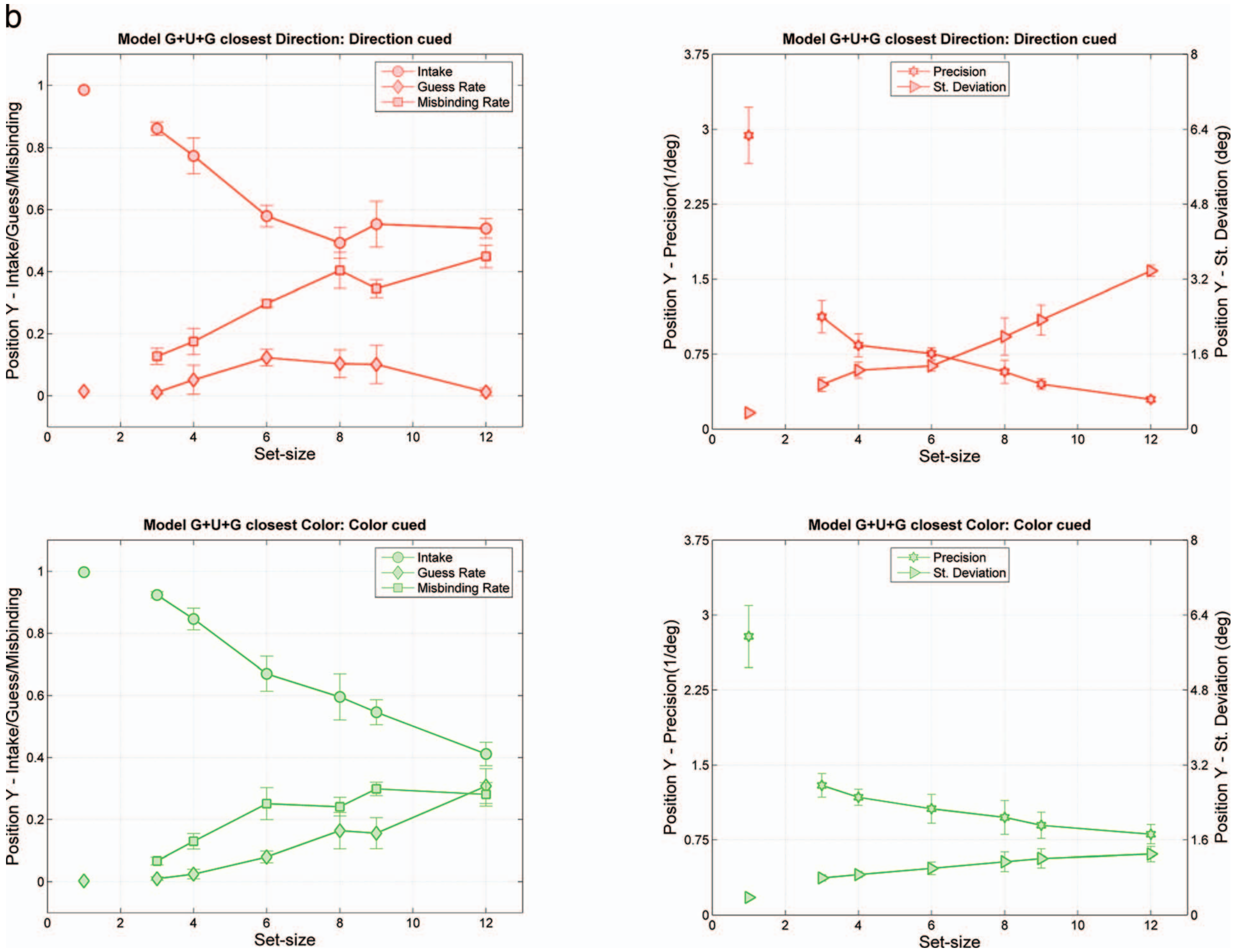
Supplementary Figure S2.1. (a) Same as Figure 4a but with data obtained from Bayesian analysis. (b) Same as Figure 4b but with data obtained from Bayesian analysis.

the superscript 1 indicates the updated values for the parameters):

$$\mu^1 = \Bigg[\sum_{i=1}^{N}\Bigg\{\Big(\varepsilon_i.\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i) + \sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i) + p^o(U|\varepsilon_i)\Big) + 2\pi.\Big(p^0(G_1^{-1}|\varepsilon_i) - p^0(G_1^1|\varepsilon_i) + p^0(G_2^{-1}|\varepsilon_i) - p^0(G_2^1|\varepsilon_i)\Big) - d_i.\sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i)\Bigg\}\Bigg] \div \Bigg[\sum_{i=1}^{N}\Bigg\{\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i) + \sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i)\Bigg\}\Bigg] \tag{16}$$

$$\sigma^1 = \Bigg\{\Bigg[\sum_{i=1}^{N}\Bigg\{\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i).(\varepsilon_i - \mu^1 - j2\pi)^2 + \sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i).(\varepsilon_i - \mu^1 - d_i - j2\pi)^2\Bigg\}\Bigg] \div \Bigg[\sum_{i=1}^{N}\Bigg\{\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i) + \sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i)\Bigg\}\Bigg]\Bigg\}^{1/2} \tag{17}$$

$$w_1^1 = \Bigg[\sum_{i=1}^{N}\Bigg\{\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i)\Bigg\}\Bigg] \div \Bigg[\sum_{i=1}^{N}\Bigg\{\sum_{j=-1}^{1}p^0(G_1^j|\varepsilon_i) + \sum_{j=-1}^{1}p^0(G_2^j|\varepsilon_i) + p^0(U|\varepsilon_i)\Bigg\}\Bigg], \tag{18}$$



Supplementary Figure S2.1. Continued.

$$w_2^1 = \left[\sum_{i=1}^{N}\left\{\sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i)\right\}\right] \div \left[\sum_{i=1}^{N}\left\{\sum_{j=-1}^{1} p^0(G_1^j|\varepsilon_i) + \sum_{j=-1}^{1} p^0(G_2^j|\varepsilon_i) + p^0(U|\varepsilon_i)\right\}\right], \quad (19)$$

$$w_3^1 = 1 - w_1^1 - w_2^1, \quad (20)$$

with $p^0(G_1^j|\varepsilon_i), p^0(G_2^j|\varepsilon_i)$, and $p^0(U|\varepsilon_i)$ given by Equations 11 through 13. These updated values become the current values in the next iteration, and the algorithm iterates these computations for the parameters until convergence to a certain local maximum of the likelihood function.

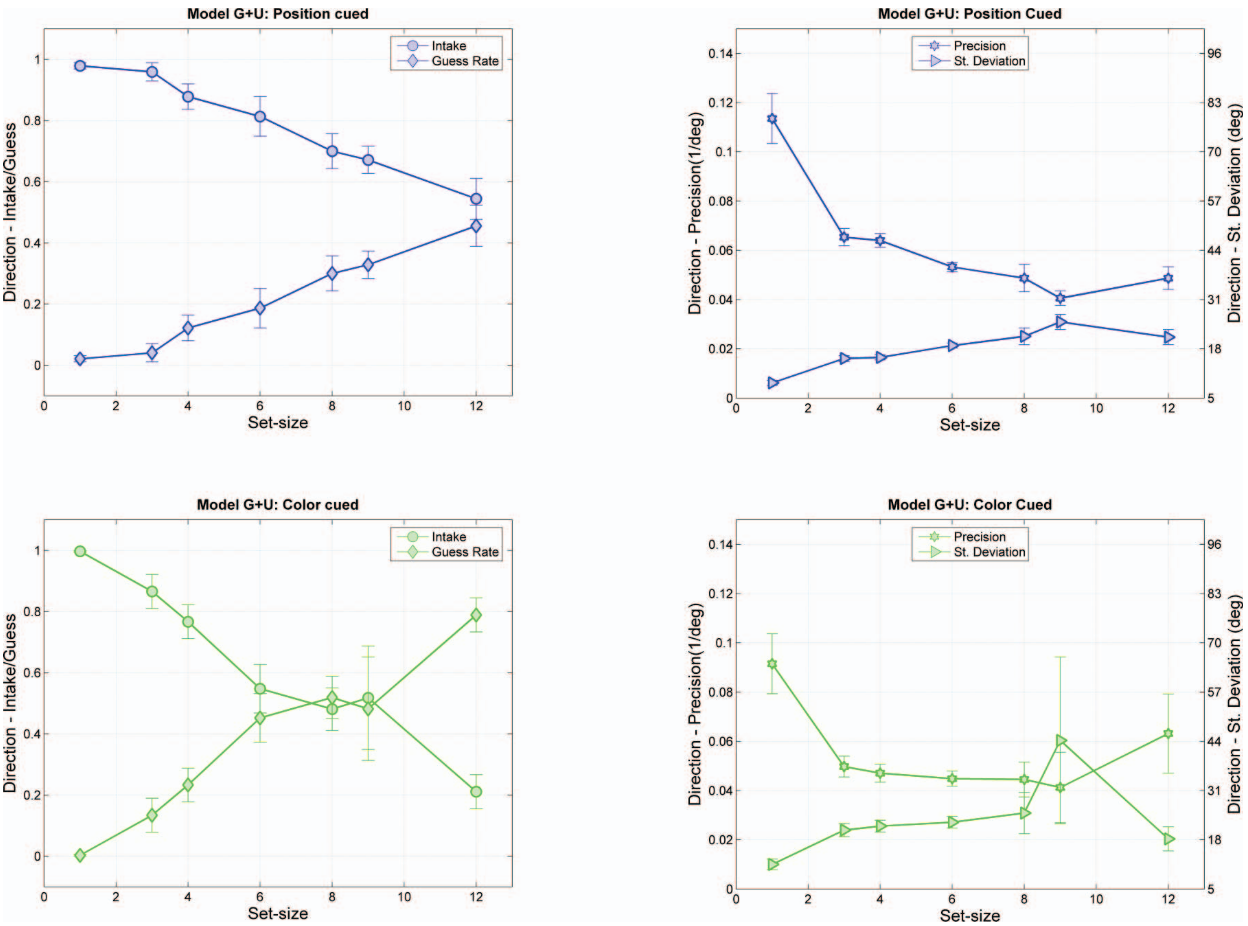
Bayesian-model comparison

We used penalized likelihood criteria of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for model selection. The AIC and BIC for a model are defined as

$$AIC = -2\ln(L) + 2p, \quad (21)$$

$$BIC = -2\ln(L) + p\ln(n), \quad (22)$$

where L represents the maximized value of the likelihood function of the model (obtained from the EM algorithm), p is the number of free parameters in the model, and n is sample size. These two criteria both try to balance a good fit with the parsimony of a model. Given a set of models, the selected model is the one with minimum AIC or BIC values. If two models yield AIC or BIC values that are insignificantly different from each other, the model with fewer parameters is preferred according to Occam's razor.



Supplementary Figure S2.2. Same as Figure 5 but with data obtained from Bayesian analysis.

In general, AIC and BIC values point to the same model. The only case in which they disagree is the condition of cuing direction of motion and reporting color in Experiment 2 (Table 6). In this condition, we obtained equivalent BICs for Gaussian + Uniform (Model 2) and Gaussian + Uniform + Gaussian *closest cued feature* (Model 3c) models ($p = 0.055$), so Model 2 should be selected. However, the smallest AIC was found for Model 3c. Therefore, the AIC and BIC taken together favor Model 3c.

Bayesian results

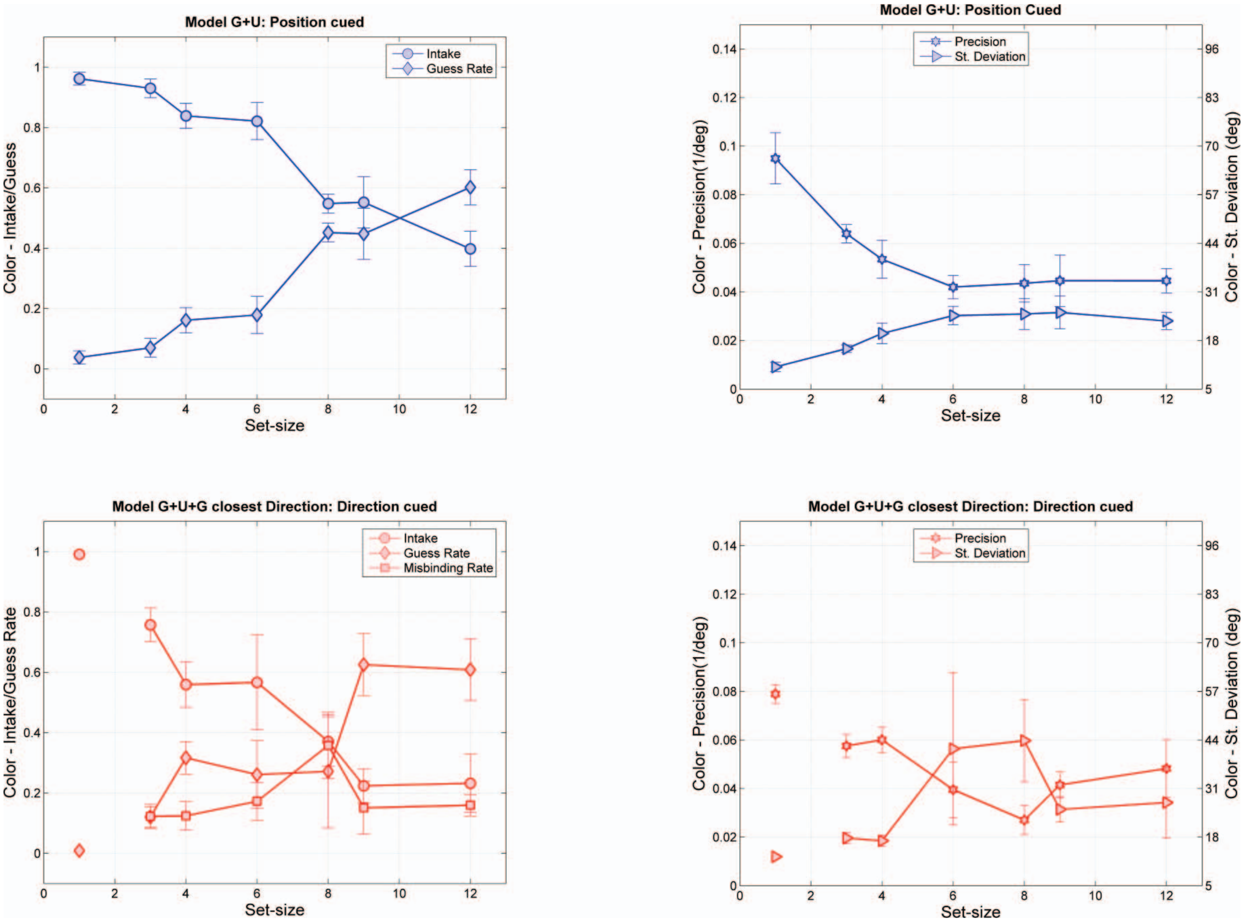
The results obtained from the models selected by Bayesian analysis are shown in Supplementary Figures S2.1 through S2.6.

Supplementary information 3

Control experiment

We conducted a control experiment to consider the possibility that the asymmetry we observed in cue effectiveness between color and direction of motion is due to the way color and motion information are presented. That is, consistently across our experiments, color onset always preceded motion onset in the presentation sequence. This temporal difference might render it more advantageous for the subjects to first encode color information, then add motion information as it becomes available, leading to color being the more effective cue.

In the control experiment, the temporal order of the two stimulus features was reversed—i.e., motion was presented first, followed by color. We repeated Experiment 2 with some slight modifications for the two conditions: (a) cuing color and reporting direction of motion and (b) cuing direction of motion and reporting color. All objects were presented in gray, rather than in unique colors, during the 1-s static previewing and the first 100 ms of motion. We increased the motion duration from 200 to 300 ms, of which the last 200 ms was left for color presentation. Increasing motion duration should make no difference to motion performance according to our previous findings (Shooner et al., 2010), but it allows enough time for color encoding and binding. The gray objects were equiluminant with object colors. All other conditions were kept the same as in Experiment 2 (see Methods). Data were collected on three observers,



Supplementary Figure S2.3. Same as Figure 6 but with data obtained from Bayesian analysis.

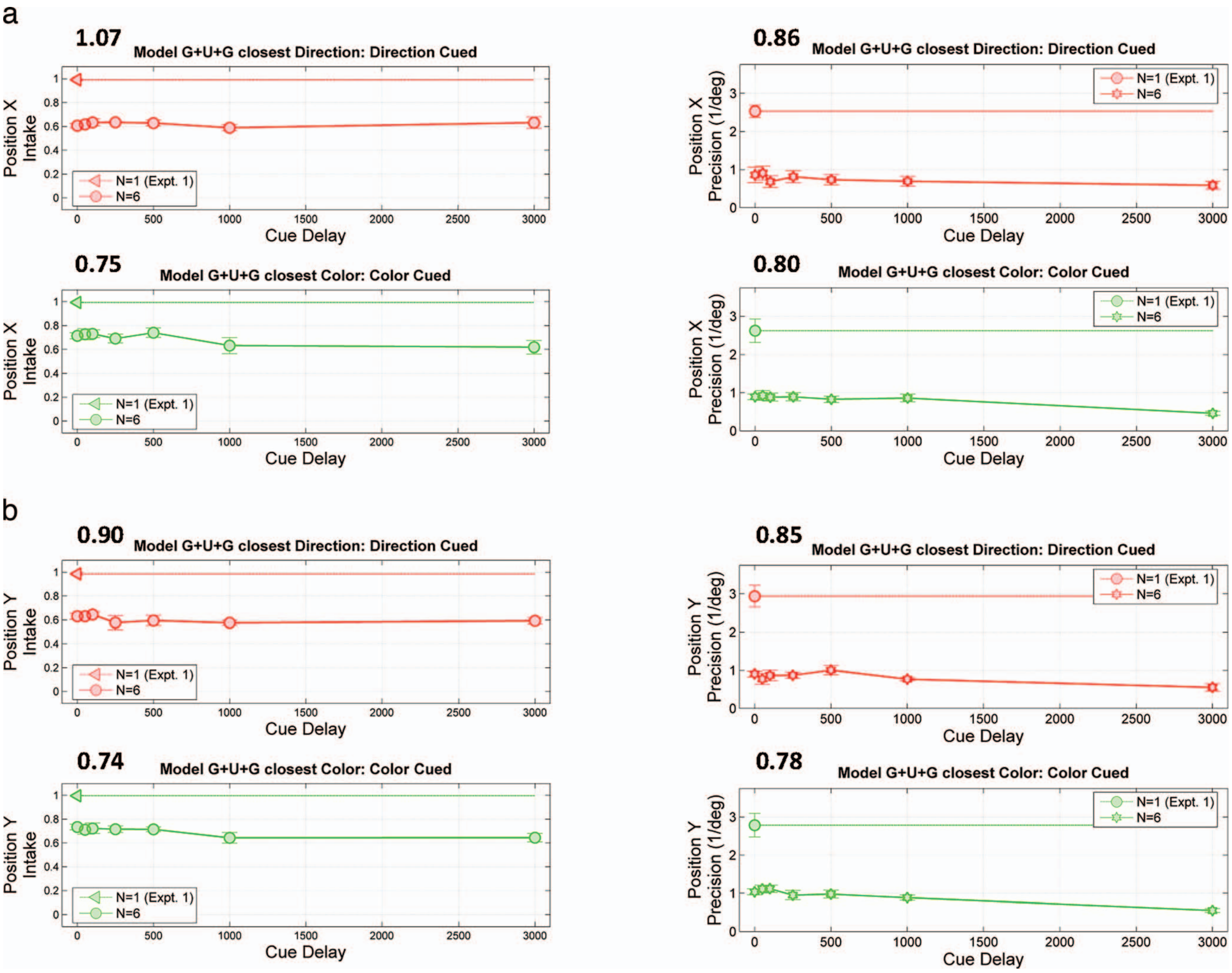
including DHL and two new observers who had normal vision. For comparison, the unmodified versions of the two cue–report conditions (same as in Experiment 2) were also run for each new observer.

We hypothesize that if the asymmetry found between color and direction of motion simply reflects the encoding order induced by the temporal difference between these features, the asymmetry is likely to be reversed in this control experiment (direction-of-motion cue should become more effective because it was presented first). Otherwise, if the same pattern of results emerges, we are confident that it reflects the natural properties of the two features.

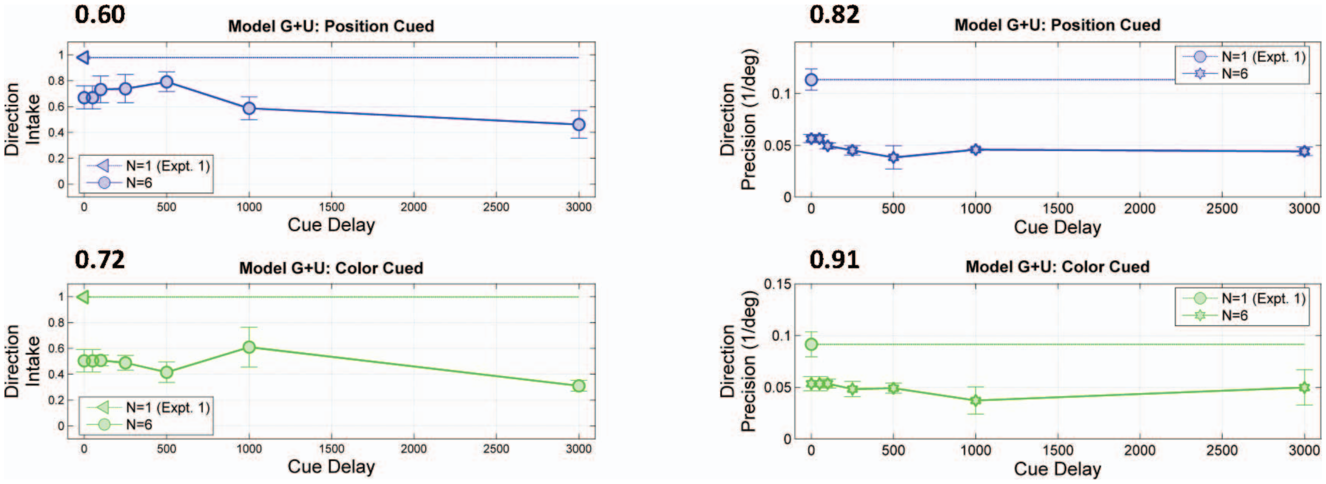
Results

Data for the replicated Experiment 2 and the control experiment are plotted in Supplementary Figure S3.1, in which similar patterns of results as in Experiment 2 are observed for both conditions. This is confirmed by a mixed-model ANOVA, with experiment and cue delay as between- and within-subjects factors, respectively, to compare main Experiment 2 with replicated Experiment 2—cuing color and reporting direction of motion: $F(1, 5) = 0.517$, $p = 0.504$, $\eta_p^2 = 0.094$; cuing direction of motion and reporting color: $F(1, 5) = 0.058$, $p = 0.820$, $\eta_p^2 = 0.011$—and a two-way repeated-measures ANOVA, with experiment and cue delay as two factors, to compare replicated Experiment 2 with the control Experiment—cuing color and reporting direction of motion: $F(1, 2) = 0.091$, $p = 0.792$, $\eta_p^2 = 0.043$; cuing direction of motion and reporting color: $F(1, 2) = 0.363$, $p = 0.608$, $\eta_p^2 = 0.154$. In all cases, no interactions between the two factors are significant. The results indicate that reversing the order of presentation onset of the two features did not change performance.

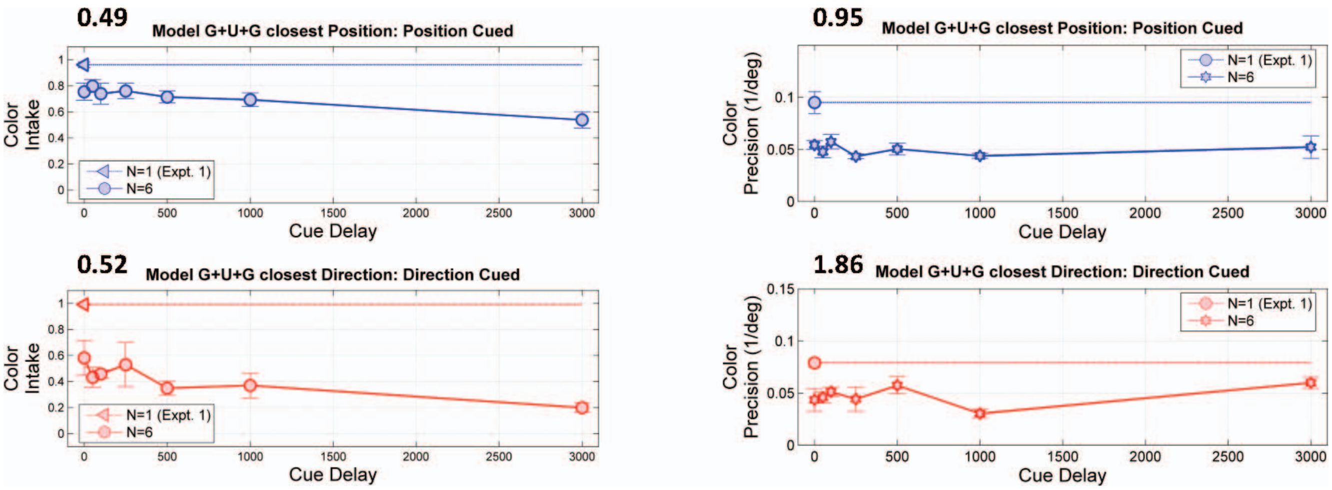
To look at the effects of reversing the presentation order on the asymmetry of the two features at each processing stage, we first applied the same procedure as in Experiment 2 (Duncan's multiple-range test on individual data combined with visual inspection of the mean data) on the control experiment's data to demarcate sensory memory from VSTM for each cue–report condition. We obtained the same demarcation results as in Table 2 (see rows 4 and 6 of Table 2). We



Supplementary Figure S2.4. (a) Same as Figure 9a but with data obtained from Bayesian analysis. (b) Same as Figure 9b but with data obtained from Bayesian analysis.

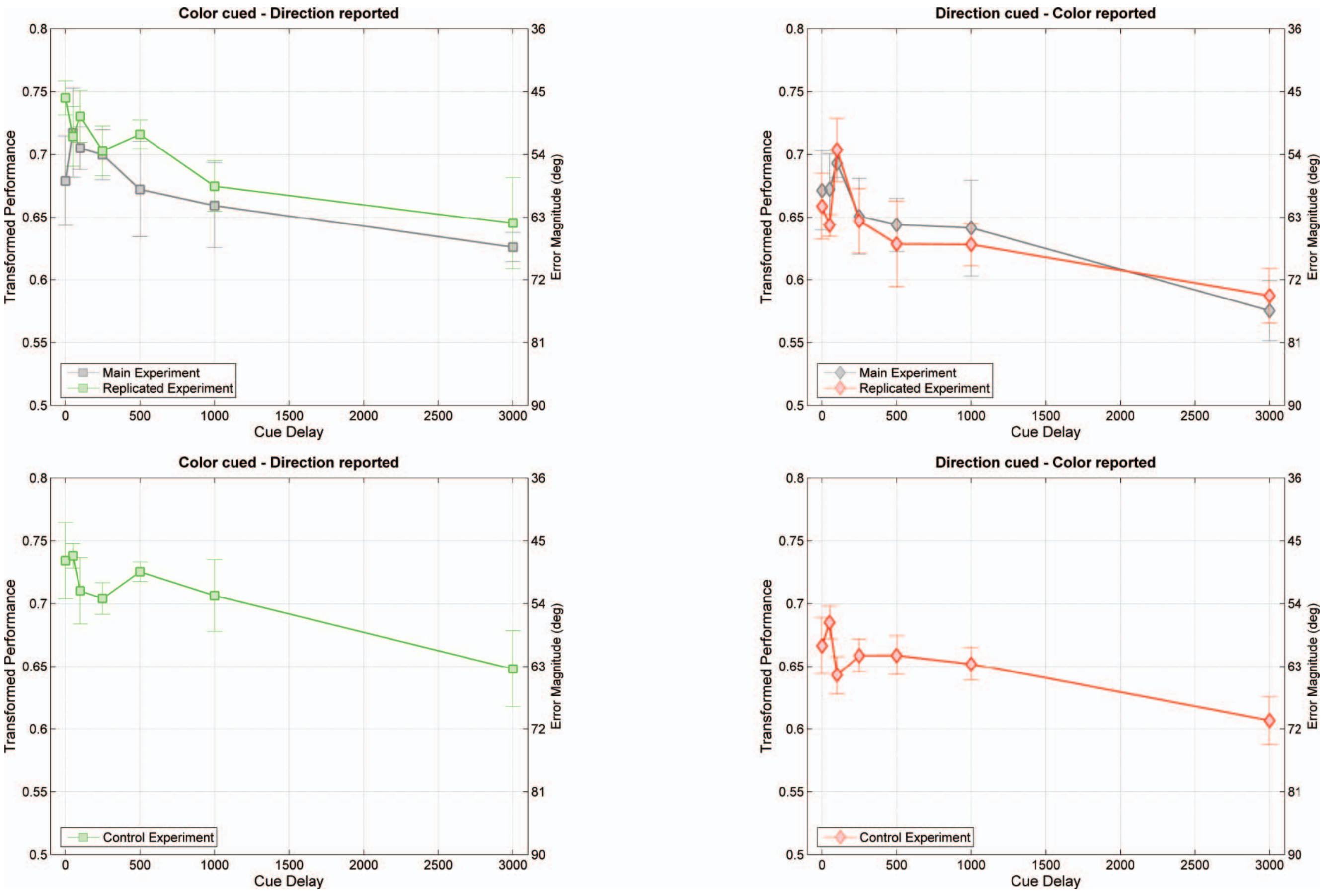


Supplementary Figure S2.5. Same as Figure 10 but with data obtained from Bayesian analysis.



Supplementary Figure S2.6. Same as Figure 11 but with data obtained from Bayesian analysis.

also followed the same logic as in Experiment 2 to reproduce Figure 8 for the two replicated and control conditions. These conditions are plotted together with the results of the main experiment in Figure 8. Across all experiments, we found that the asymmetry of color and direction of motion remains stable regardless of their temporal order. Hence, color is intrinsically the more effective cue compared to direction of motion.



Supplementary Figure S3.1. Data obtained from the replicated Experiment 2 (top) and the control (bottom) experiment for the two conditions: cueing color and reporting direction of motion (left panels; green) and cueing direction of motion and reporting color (right panels; red). Also in top panels (gray) are data from main Experiment 2: Transformed performance (left y-axes) and error magnitude (right y-axes) averaged across observers are shown as a function of cue delay. Error bars correspond to ±1 standard error of the mean.