Psychological Review
2018, Vol. 125, No. 3, 391–408

0033-295X/18/$12.00 http://dx.doi.org/10.1037/rev0000099

The Capacity of Trans-Saccadic Memory in Visual Search

Nicholas J. Kleene and Melchi M. Michel
Rutgers University

Maintaining a continuous, stable perception of the visual world relies on the ability to integrate information from previous fixations with the current one. An essential component of this integration is trans-saccadic memory (TSM), memory for information across saccades. TSM capacity may play a limiting role in tasks requiring efficient trans-saccadic integration, such as multiple-fixation visual search tasks. We estimated TSM capacity and investigated its relationship to visual short-term memory (VSTM) using two visual search tasks, one in which participants maintained fixation while saccades were simulated and another where participants made a sequence of actual saccades. We derived a memory-limited ideal observer model to estimate lower-bounds on memory capacities from human search performance. Analysis of the single-fixation search task resulted in capacity estimates (4–8 bits) consistent with those reported for traditional VSTM tasks. However, analysis of the multiple-fixation search task resulted in capacity estimates (15–32 bits) significantly larger than those measured for VSTM. Our results suggest that TSM plays an important role in visual search tasks, that the effective capacity of TSM is greater than or equal to that of VSTM, and that the TSM capacity of human observers significantly limits performance in multiple-fixation visual search tasks.

Keywords: visual search, visual short-term memory, ideal observers, rate-distortion theory, trans-saccadic memory

Saccades, rapid eye-movements that occur multiple times per second, are our primary means of gathering visual information. Each fixation provides us with an isolated snapshot of the visual world. Generating a coherent percept of the world requires some sort of integration process to combine information across fixations. This integration process necessitates storing visual information from previous fixations in memory (termed trans-saccadic memory, TSM) so that it may be retrieved later. Any limitations on TSM capacity could hinder human performance in everyday tasks that require multiple fixations, such as driving, reading, or searching. The successful performance of such tasks in everyday life suggests that TSM must play a critical role in visual perception. The goal of the current study was to evaluate TSM capacity—specifically, its lower bound—to better understand how TSM limits performance in tasks that require multiple fixations.

Nicholas J. Kleene, Department of Psychology, Rutgers University; Melchi M. Michel, Department of Psychology & Center for Cognitive Science (RuCCS), Rutgers University.

Portions of the ideas and results reported in this article were presented at annual meetings of the Visual Sciences Society (in 2015, 2016, and 2017), and at the annual meeting of the Society for Mathematical Psychology (in 2016).

This research was supported by the National Science Foundation (NSF Award BCS 1456822 to Melchi M. Michel).

Correspondence concerning this article should be addressed to Melchi M. Michel, Department of Psychology & Center for Cognitive Science (RuCCS), Rutgers University, Piscataway, NJ 08854. E-mail: melchi.michel@rutgers.edu

Background

Early theories of TSM postulated that during integration sensory information would be superimposed across fixations in spatial coordinates (Irwin, Yantis, & Jonides, 1983), making it similar to iconic memory in that it had high resolution and a large capacity, at least relative to visual short-term memory (VSTM). However, when tasked with fusing two images separated by a saccade, participants proved unable to integrate across fixations in spatial coordinates (Irwin et al., 1983). TSM seemed more similar to conventional VSTM than to iconic memory. Later studies showed that TSM has the same time-course as VSTM and has comparable spatial resolution, with capacity estimates similar to the commonly held "3–4 objects" for VSTM (Irwin, 1991, 1992). More generally, current evidence tends to support the view that TSM and VSTM are largely mediated by a common memory system (e.g., Bays & Husain, 2008; Brady, Konkle, & Alvarez, 2011; Luck, 2008; Prime, Tsotsos, Keith, & Crawford, 2007). As such, a secondary objective of this study was to evaluate this hypothesis by comparing capacity estimates of TSM with those obtained for conventional VSTM.

While few studies have explicitly investigated TSM, VSTM has been the focus of vast amounts of research, with particular emphasis given to measuring the capacity and the resolution of the encoded information (e.g., Luck & Vogel, 1997; Wilken & Ma, 2004; Zhang & Luck, 2008). Much of the work on estimating VSTM capacity has used change detection (e.g., Bays & Husain, 2008; Luck & Vogel, 1997; Wilken & Ma, 2004; Woodman, Vogel, & Luck, 2001) or continuous response paradigms (e.g., Fougnie & Alvarez, 2011; Wilken & Ma, 2004; Zhang & Luck, 2008). In change detection (or discrimination) tasks, participants

view a stimulus array, then, after an intervening blank or mask frame, view a new display and report whether (or how) one of the items changed (Phillips, 1974). Continuous response paradigms use the same stimuli as change detection tasks, but instead require reports of the value of a particular item, such as its color or orientation (Prinzmetal, Amiri, Allen, & Edwards, 1998; Prinzmetal, Nwachuku, Bodanski, Blumenfeld, & Shimizu, 1997). A major drawback of both approaches for assessing TSM is that they typically assess memory over temporal intervals far longer than the brief durations (~25–75 ms) characteristic of human saccades (Abrams, Meyer, & Kornblum, 1989; Bahill, Brockenbrough, & Troost, 1981). It is extremely rare for natural scenes to exhibit significant changes over the span of such short intervals, and when they do, there are often accompanying transients that act as cues to the change. This makes change detection tasks ill-suited for studying TSM.

An alternative approach is to use multiple-fixation visual search tasks that require trans-saccadic integration for efficient performance. In contrast to change detection tasks, visual search tasks usually consist of static displays in which participants are tasked with finding a target item. Critically, existing work on visual search has not been able to characterize how VSTM might mediate TSM (Eckstein, 2011). Instead, most models of visual search have focused on the influence of target-distractor similarity, display set size (e.g., Palmer, 1994; Verghese & Nakayama, 1994; Wolfe, 2007), or target position uncertainty (e.g., Eckstein, 1998; Michel & Geisler, 2011; Pelli, 1985).

Quantifying Memory Capacity

A major challenge in the study of TSM is determining its capacity. Studies of the capacity of VSTM have focused on two competing families of models (see Luck, 2008 for a review). The first of these are fixed-resolution, or *slots*, models in which VSTM is posited to hold a limited number of "items," typically 3–4 (Cowan, 2001; Luck & Vogel, 1997). Items that are held in memory are encoded with perfect resolution, but no information is stored for any additional items (Luck, 2008). As such, slots models are examples of high threshold models, where items are either represented perfectly, or not at all.

The prevailing alternative to slots models are variable resolution or *continuous resource* models. Within these models, VSTM capacity is viewed as an arbitrary resource that can be allocated flexibly among all the items in a visual display (Wilken & Ma, 2004). In contrast to the slots models, continuous resource models do not assume a limit on the number of items that can be encoded. Instead, they hypothesize a tradeoff between the number of items encoded and the fidelity with which they are encoded. As the set size (the number of items in the display) increases so does the memory encoding error associated with each item. This makes continuous resource models an example of low threshold models, in which all items may be represented, albeit imperfectly. A major difference between these two classes of models is how much noise is present in the encoded representations. Continuous resource models predict that encoding noise will increase as a function of set size, whereas slots models predict that each encoded item will have close to perfect resolution. In practice, however, the predictions of these two classes of models are often difficult to distinguish (e.g., see van den Berg, Awh, & Ma, 2014).

Slots models and continuous resources models also take very different approaches to comparing capacity estimates across different tasks and stimuli. In slots models, depending on the complexity of the stimuli (e.g., cubes vs. squares), capacity estimates have been shown to vary anywhere from about 1 to 5 items (Alvarez & Cavanagh, 2004). In continuous resource models, on the other hand, where capacity is characterized in terms of precision rather than the number of items, precision depends on the encoded feature (e.g., orientation, location, color), as well as on other details of the stimuli (Bays & Husain, 2008). A major difficulty in evaluating these results comes from how an item is defined; for example, should an item consist of a single feature dimension or a set of integrated features? Once again, the evidence is mixed, with support for considering items as integrated features (Xu, 2002), or for considering an item to consist of a single feature dimension (Alvarez & Cavanagh, 2004).

Sims, Jacobs, and Knill (2012) recently introduced an alternative, normative, approach to overcome some of the limitations of the fixed- and variable-resolution models for quantifying memory capacity. Their proposed approach, which is based on rate distortion theory, conceives of memory as a capacity-limited channel. The notion of capacity in this approach is therefore quite general, and the approach has been shown to produce estimates of capacity that are comparable across visual features and task parameters (Sims et al., 2012).

Rate-distortion theory (Berger, 1971; Shannon, 1948, 1959) is a branch of information theory primarily concerned with the issue of lossy data compression, where inexact approximations are used to represent an encoded signal (often a video, image, or audio file) so that it can be reconstructed after transmission. Greater compression often results in greater distortion of the transmitted signal, degrading its quality. However, there is a tradeoff, as representing a more compressed signal requires fewer bits (binary units of information). This relationship between the information rate and distortion of a system (i.e., fewer bits $\rightarrow$ more distortion) is defined by the rate-distortion function (Shannon, 1959), which specifies the minimum bit-rate required to reconstruct a signal without exceeding a given distortion. Conversely, for a given distortion it is also possible to place a lower bound on the bit-rate of a system. The capacity of a communication channel is closely related to the information rate of a system. In particular, Shannon's channel coding theorem (Shannon, 1948) shows that a channel's capacity limits the maximum rate at which it can transmit information (i.e., a channel with a capacity of R bits per second can transmit information at a maximum bit rate of R). This means that any lower bound placed on the bit rate of a system identically bounds its transmission capacity.

The value of rate-distortion theory for characterizing memory becomes evident if we consider visual memory as a system for transmitting visual information across time. TSM, or VSTM, efficiently stores visual signals so they may be retrieved and reconstructed at a later time (Sims et al., 2012). In this conception, the bit-rate of a visual memory system represents its capacity (the amount of information that can be stored), while the distortion represents the noise that is introduced into memory representations because of the system's finite capacity. The rate-distortion theory framework allows us to quantify memory capacity in terms of bits, which are task-independent units of information, as opposed to "slots" or arbitrary resources.

Sims et al. (2012) used the rate-distortion approach to estimate VSTM capacity for change detection. Across several experiments, they systematically manipulated both the features to be memorized (i.e., orientation, line-length) and the degree of variability present in the distribution of feature values. They then used an ideal observer analysis to estimate the lower bound on VSTM capacity, with estimates ranging from 3 to 8 bits across participants and conditions. More important, because this information-theoretic approach automatically accounts for effects of the variability of the encoded feature values, the authors found no significant difference in the capacities estimated across tasks or variance conditions. Its success in producing consistent estimates across change-detection tasks suggests that this information theoretic approach holds promise as a task independent way of quantifying memory capacity.

In the current study, we combined ideas from two normative modeling approaches, the ideal observer framework for VSTM introduced by Sims et al. (2012), and the ideal searcher framework of Najemnik and Geisler (2005), to derive an ideal observer model for integration in TSM. We then used this model to estimate human TSM capacity, as well as to determine how TSM capacity limits performance in the visual search tasks. The ideal observer model used rate-distortion theory, in conjunction with human detection sensitivity and measured search performance, to place a lower bound on TSM capacity. Finally, we compared our TSM capacity estimates with those reported previously for VSTM to determine whether TSM and VSTM could be mediated by common or comparable mechanisms.

Overview of Tasks and Conditions

Two types of visual search tasks were performed in separate experiments. These will be referred to as the Fixed-Gaze experiment and the Gaze-Shift experiment. Each task used a different group of human participants. The critical difference between the experiments was that the Gaze-Shift experiment used natural saccades, while the Fixed-Gaze experiment simulated the retinal effects of natural saccades by means of manipulations to the display. Each experiment used the same type of search task, namely, finding a target signal in displays of visual noise. In each experiment the presentation on each trial consisted of a sequence of either natural (Gaze-Shift experiment) or simulated (Fixed-Gaze experiment) fixation intervals in which the display contained noise alone or target + noise.

There were two main conditions in each experiment, termed the Redundancy condition and the Uncertainty condition. In the Redundancy condition the target appeared in the same location during each natural (Gaze-Shift experiment) or simulated (Fixed-Gaze experiment) fixation interval. Optimal performance thus, requires integrating across all fixation intervals. Memory load (VSTM or TSM) increases as the number of possible locations increases and as the number of fixation intervals increases. In the Uncertainty condition the target appeared in only one of the fixation intervals. Optimal performance in this condition does not require integrating across fixation intervals (optimal performance can be obtained by extreme detector or "max-rule" models; Nolte & Jaarsma, 1967; Pelli, 1985; and see The Trans-Saccadic Integration Component).

Performance in both the Redundancy and Uncertainty conditions was compared with that of a constrained Ideal Observer model (see An Ideal Observer Model of Memory-Limited Search). Differences in the extent to which participants underperform the Ideal Observer in the two conditions are important in inferring the capacity of TSM.

A separate detection task was performed both before and after data were collected on the search tasks. The results of the prior detection task were used to select the contrast of the search target. The results of the post detection task were used in the development of the Ideal Observer model.

Method

Participants

Data were collected from a total of nine participants (5 in the Fixed-Gaze experiment and 4 in the Gaze-Shift experiment, all of whom were naïve to the aims of the study). All participants had normal or corrected-to-normal vision and were paid $10 per hour for their participation.

Apparatus

Stimuli were presented on a 22 in Philips 202P4 CRT monitor at a resolution of 1280 × 1024 pixels and a frame rate of 100 Hz. Participants were seated 70 cm from the display such that the display subtended 15.8° × 21.1° of visual angle. Stimuli were generated and presented using MATLAB software (Mathworks) and the Psychophysics Toolbox extensions (Brainard, 1997). Head position was fixed using a forehead and chin rest. Participants' right-eye was tracked using an EyeLink 1000 infrared eye tracker (SR Research, Kanata, Ontario, Canada). Gaze location was sampled from the eye tracker at 500 Hz in the detection task and at 1000 Hz in both search tasks.

Stimuli

Target and background. The target signal was a two cycle-per-degree Gabor, oriented 45° counterclockwise from vertical (see Figure 1). This relatively low target spatial frequency was chosen to minimize the effect of retinal eccentricity on visual

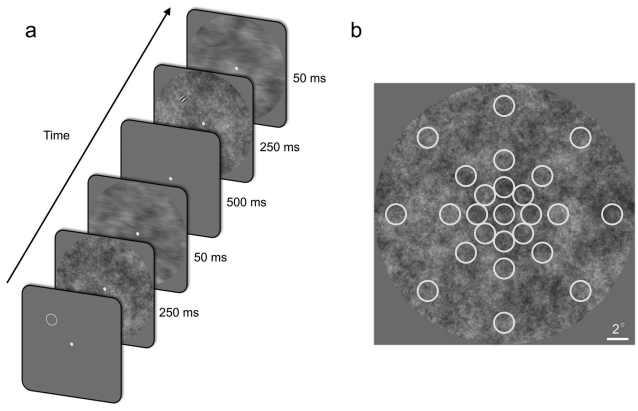

Figure 1. The 2IFC detection task. (a) Stimulus sequence for a trial of the detection task. In this example, the target is present in the second interval. (b) Visual field locations that were tested to construct visibility maps in the Gaze-Shift experiment.

sensitivity. The target was embedded in a circular noise mask (the background), 20° in diameter, which consisted of 1/*f* filtered noise with 10% RMS contrast. Each noise mask presented on a given trial in the Detection task or search tasks was generated independently. The area surrounding the background was set to the mean luminance of the display (40 cd/m^2). The same target and background were used in the Gaze-Shift experiment, except that the noise mask was 24° in diameter.

Procedure. Procedures are described below for the Detection task, the Fixed-Gaze localization search task, and the Gaze-Shift localization search task. During each laboratory visit participants ran 10 blocks of 50 trials each for a total of 500 trials. At the start of each block, participants completed a 5-point (detection and Fixed-Gaze search tasks) or a 9-point gaze calibration routine (Gaze-Shift search task). The calibration was repeated until the average test–retest measurement error across gaze points fell below 0.25°.

Detection task. Detection performance was measured both before and after the Fixed-Gaze experiment and the Gaze-Shift experiment. Participants fixated a central marker before each trial. A stroked circle cued the target location (see Figure 1). Participants started each trial when ready with a button press. After an interval of 100–400 ms (chosen randomly) two 250 ms intervals of 1/*f* filtered noise were presented, one (selected randomly) containing the target. Each interval was followed by a 50 ms simulated saccadic transient (Fixed-Gaze Search Task) and the two intervals were separated by a further delay of 500 ms. After viewing both intervals, participants indicated which contained the target. Auditory feedback was then presented. If participants blinked or moved their eyes at any time during the trial they were notified and the trial was discarded.

The target could appear in any one of eight possible directions relative to fixation. For the Fixed-Gaze experiment the eccentricity was 7 degrees. For the Gaze-Shift experiment the eccentricity was either 0, 2.5, 5, or 10 degrees, with trials blocked by eccentricity. Otherwise, the Fixed-Gaze and Gaze-Shift detection tasks were identical.

At the start of each block of trials, participants completed five practice trials in which the target was presented at 50% contrast. Data from these practice trials was not recorded. For the remaining trials in the block, target contrast was selected using an interleaved, adaptive procedure (Kontsevich & Tyler, 1999). Participants in the Fixed-Gaze experiment completed the detection task in approximately 5 sessions (4 pretest, 1 posttest), while those in the Gaze-Shift experiment completed it in approximately 10 sessions (6 pretest, 4 posttest). The first two sessions of the detection tasks served as practice. Data from these two sessions were excluded from the analysis.

Fixed-Gaze search task. The Fixed-Gaze search task involved integrating visual information across a sequence of simulated fixation intervals to localize the position of a Gabor target among eight possible locations. The fixation intervals were separated by *simulated saccadic transients* created by translating the display 2° horizontally on each frame (10 ms) for a total of five frames (50 ms). This 2° displacement was equal to the average translation of the retinal stimulus that would have occurred across 10 ms had the participant made an actual 10° saccade. The direction of translation for each simulated transient (leftward or rightward) was selected randomly. Each frame of the simulated transient was constructed by taking a sample of the noise background and convolving it with a 2° horizontal boxcar filter. The simulated transient was used in both the Detection task and the Fixed-Gaze search task. Figure 2 shows the target and background, as well as a frame of the simulated saccadic transient from the Fixed-Gaze task.

Each trial of the Fixed-Gaze task began with the participant fixating a marker at the center of the display (see Figure 3). Eight evenly spaced circular cues located 7 degrees from fixation indicated the possible locations of the target. Participants initiated each trial with a button press, and following an SOA of 100–400 ms, one, two, or four intervals of the target and background (to be referred to as fixation intervals) were presented. The duration of each fixation interval was 250 ms and each was followed by 50 ms of the simulated saccadic transient. In the Redundancy condition, the target appeared in the same location in each fixation interval. In the Uncertainty condition, the target appeared in only one randomly selected fixation interval.

A different 1/*f* background noise mask was independently generated for each interval. This was done to make the display noise temporally independent (i.e., independent across fixation intervals). Making the display noise temporally independent had two favorable consequences: (a) it maximized the amount of information that could be gained by integrating across fixation intervals, increasing the expected impact of limited memory capacity; and (b) it allowed us to treat the sensory + display noise as temporally independent in the ideal observer model (see The Sensory Representation Component).

At the end of each trial a response display appeared, with circular cues marking the eight potential target locations. Participants indicated the perceived target location by directing their gaze to the corresponding cue and pressing a key. Auditory feedback indicated whether the response was correct. If participants blinked or moved their eyes at any time before the appearance of the response display they were notified and the trial was discarded.

Trials were blocked according to: (a) target permanence condition (Redundancy or Uncertainty), and (b) number of fixation intervals (one, two, or four). Participants were informed of the condition before each block. As in the detection task, participants started each block by completing five practice trials with target contrast set to 50%. Data from these practice trials were not recorded. For the remaining trials in each block, target contrasts at each of the eight possible target locations were selected to achieve a sensitivity (d') of approximately 1.5 based on the measurements

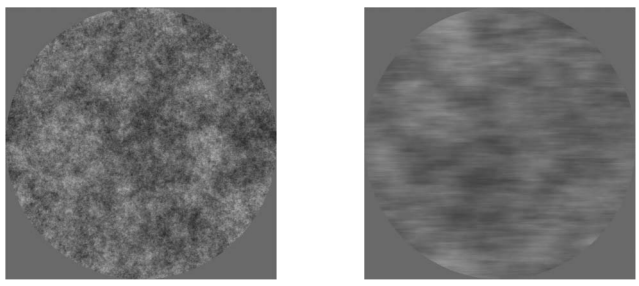

Figure 2. Displays used in the Fixed-Gaze experiment (left). A stimulus frame from a fixation interval (with target absent). (right) A frame from the simulated saccadic transient.

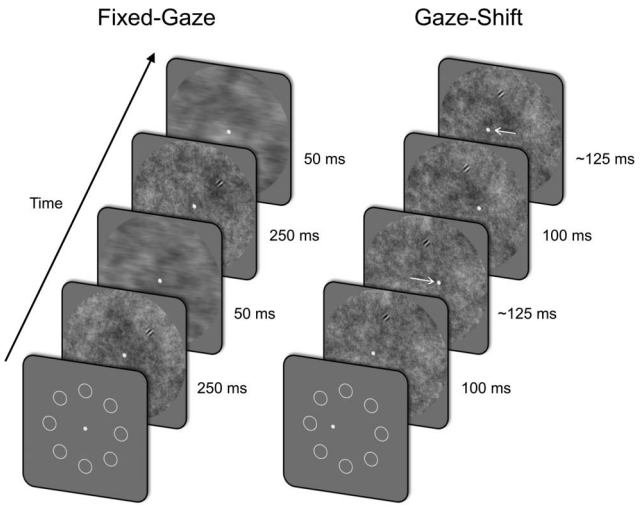

Figure 3. Stimulus sequence for a two interval Redundancy trial in the Fixed-Gaze (left) and Gaze-Shift (right) search tasks. In Uncertainty trials, the target appeared in only one of the two intervals. The white arrows in the right panel are schematic representations of gaze shifts and did not appear in the experimental display.

made in the pretest detection task (see Detection Task). Participants in the Fixed-Gaze search task completed approximately five sessions. The first session served as a practice day and was not analyzed. The analysis for each participant (see Fixed-Gaze Experiment) included only data from the last four sessions (40 blocks, consisting of 400 trials per condition).

Gaze-Shift search task. Like the Fixed-Gaze search task, the Gaze-Shift search task asked observers to localize the position of a Gabor target among eight possible target locations. However, it involved integrating visual information across a sequence of real saccades rather than across simulated saccadic transients. Participants began each trial by fixating a small marker 2.5° to the left or right of the display center (see Figure 3). The position of the initial fixation marker (left or right) was determined randomly on each trial. Eight circular cues evenly spaced 7.5° from the display center indicated the potential target locations. When ready, participants initiated the trial with a button press. One, two, or four fixation intervals were then presented, with the number of intervals remaining constant across a block of trials. Each fixation interval was presented for 100 ms, at which point the fixation marker was extinguished, a new horizontally displaced fixation marker appeared, and a tone was played. These events cued participants to saccade 5° to the left or right, depending on the initial displacement of the fixation marker (i.e., if the initial fixation marker was on the left, the first saccade would be to the right, the second would be to the left, etc.). Participants were asked to avoid making saccades before the onset of the cue and to initiate the saccade no later than 250 ms after the cue onset. Thus, saccades should be occurring 150–400 ms after the presentation of the stimulus in each fixation interval.

As in the Fixed-Gaze Experiment, a different 1/*f* background noise mask was independently generated for each interval. To avoid motion transients caused by the onset of each new noise mask, the masks were replaced during the saccade. Saccades were detected using a threshold eye velocity of 150 deg/s (Bahill, Clark, & Stark, 1975). If saccades did not occur within the specified time window, the trial was discarded. At the end of each trial, participants were presented with the same response display as in the Fixed-Gaze search task. Again, participants indicated the perceived target location by directing their gaze to the corresponding cue and pressing a key. Auditory feedback indicated whether the response was correct.

The Gaze-Shift search task consisted of the five unique combinations of the target permanence condition (Redundancy or Uncertainty), and either 1, 2, or 4 fixation intervals. Each condition was run in blocks of 50 trials. Practice trials and the selection of target contrast were the same as in the Fixed-Gaze search task. The Gaze-Shift experiment required approximately 7–8 sessions. Because of the difficult dual-task nature of the Gaze-Shift task, observers required 3–4 practice sessions, which were not analyzed. The analysis (see Fixed-Gaze Experiment) included only data from the last four sessions (40 blocks, consisting of 400 trials per condition).

An Ideal Observer Model of Memory-Limited Search

An ideal observer is a hypothetical device that performs a specified task optimally (as well as possible) given all the information available and any specified task constraints (Geisler, 2003, 2011). *Constrained* ideal observer models can be used to suggest or reject proposed constraints in models of human performance. In particular, if a proposed constraint results in an ideal observer that is outperformed, on average, by human observers, then the constraint cannot be valid. The main purpose of the ideal observer model in the current study is to set a lower bound on TSM capacity by rejecting any putative limits that fail to account for human performance.

Deriving the ideal observer model for a particular task requires making explicit all of the information available for the task as well and specifying any relevant constraints, such as limitations on sensory representation or encoding. The specification of these constraints makes it possible to determine the optimal strategies and the maximum attainable performance in a task, and to use these as a "gold standard" against which to compare human performance.

We modeled visual search in the Fixed-Gaze and Gaze-Shift search tasks using a TSM capacity-limited ideal observer. The ideal observer model was derived by extending the information-theoretic VSTM framework introduced by Sims et al. (2012) to allow for the optimal storage of information across multiple saccades (rather than a single memory interval) and combining it with the peripheral visibility and temporal integration components of the ideal searcher introduced by Najemnik and Geisler (2005).

The search display consisted of a field of spatial noise with the target signal appearing either in each fixation interval (Redundancy condition) or in only one of the fixation intervals (Uncertainty condition). The display for each interval was represented as a vector of scalar responses at each of the n potential target locations. The task in each trial was to accurately identify the actual target location J among the n potential locations (a table of the mathematical notation used in this section is provided in Appendix C).

The information available included the prior probabilities on target location and the visual information presented in the display. The extrinsic (task-imposed) constraints included spatial uncertainty regarding the position of the target and, in the Uncertainty condition, temporal uncertainty regarding the fixation interval within which the target would appear. The assumed intrinsic constraints included sensory response noise added by the visual system,[1] and memory encoding noise introduced as a result of using a limited memory capacity R. The magnitude of the sensory noise component was measured independently in the 2IFC detection task (see Detection Task and Visibility Maps), so that the only free parameter in the ideal observer was the total TSM capacity R.

The memory-limited ideal observer, illustrated schematically in Figure 4, can be thought of as comprising three interconnected components:

1. A **sensory representation component** (see The Sensory Representation Component) that encodes the physical signals displayed during each fixation interval into a set of sensory responses.

2. A **memory encoding/decoding component** (see The Memory Encoding Component) that optimally encodes the sensory responses from previous fixation intervals using the available memory capacity.

3. A **trans-saccadic integration component** (see The Trans-Saccadic Integration Component) that optimally combines the memory-encoded sensory responses across fixation intervals to infer the target location.

Figure 5 (left panel) shows expected performance for the TSM-limited ideal observer when the sensory signal-to-noise ratio (d') at all potential target locations is set to 1.5. The ideal observer's performance improves as the number of fixation intervals increases in the Redundancy condition because the observer gets more samples of the (noisy) target and, thus, more evidence about its true location. Note that across a range of putative TSM capacities, the capacity-limited ideal observer outperforms several classes of alternative models that do not integrate visual information across all intervals, shown in right panel of Figure 5. These include a "max rule" model that keeps track only of the strongest response across intervals and locations, and selects the location of this maximum response as the target location (Green & Swets, 1966; Swensson & Judy, 1981); as well as an unlimited capacity model that only remembers visual information from the previous fixation or the previous two fixations. This result indicates that integration across all intervals is essential to optimal performance. For the Uncertainty condition, where the target appears in only one interval and each additional interval adds temporal uncertainty, integration is less helpful and the ideal observer's performance worsens as the number of fixation intervals increases. This is expected, as previous work has shown that increasing temporal uncertainty diminishes performance (Lasley & Cohn, 1981).

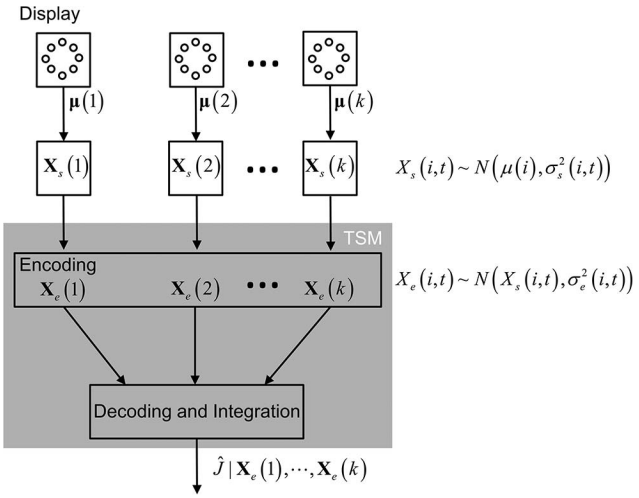

Figure 4. Schematic of the trans-saccadic memory (TSM) limited ideal observer model. $\boldsymbol{\mu}(t) = \mu(1, t), \ldots, \mu(n, t)$ is a vector representing target presence/absence at each location in the display during fixation t, $\mathbf{X}_s(t) = X_s(1, t), \ldots, X_s(n, t)$ represents the set of n noisy sensory responses obtained during fixation t, and $\mathbf{X}_e(t) = X_e(1, t), \ldots, X_e(n, t)$ represents the set of memory-encoded responses from fixation t. $\hat{J}$ represents the target location estimated from the encoded responses.

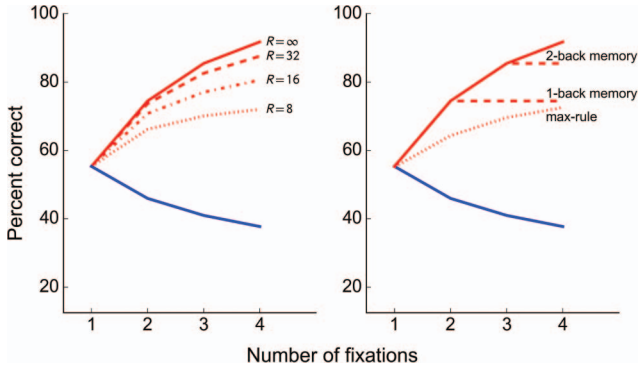

Figure 5. Expected performance of various theoretical observers for target contrasts corresponding to $d' = 1.5$ at each of the eight potential target locations. The solid curves represent the performance of the ideal observer in Redundancy (red, upper curve) and Uncertainty (blue, lower curve) conditions. The broken lines in the left panel show the performance of the trans-saccadic memory (TSM) limited ideal observer for various capacity limits R. The broken lines in the right panel show performance for several alternative theoretical observers: the dotted line shows performance for a max rule observer, while the two dashed lines show performance for an otherwise ideal observer that only stores information from one or two previous fixations. The computations underlying these curves are described in An Ideal Observer Model of Memory-Limited Search. See the online article for the color version of this figure.

The Sensory Representation Component

During each fixation interval, the observer receives sensory signals from each of the n potential target locations. Consider the display during a single fixation interval t of a Redundancy trial. Let $X_s(i, t)$ represent the magnitude of a sensory response associated

[1] Target signal variability introduced by the $1/f$ noise mask, an extrinsic source of uncertainty, was modeled as part of the sensory noise component. See The Sensory Representation Component) for details.

with location $(\varepsilon_i, \theta_i)$, where ε_i is the eccentricity of region i, and θ_i is its angular direction with respect to the right horizontal meridian. We model the sensory response as a normally distributed random variable

$$X_s(i,t) \sim \mathcal{N}\big(\mu(i), \sigma_s^2(i,t)\big), \tag{1}$$

where $\mu(i)$ is the *display variable* that represents the expected value of the sensory response and takes on one of two values depending on whether or not the target is present at location i. For mathematical convenience and without loss of generality, we assume that $\mu(i)$ is 0.5 when the target is present at location i and -0.5 otherwise. The variance term $\sigma_s(i, t)$, which includes the combined effects of the $1/f$ noise mask added to the display and of sensory uncertainty, is characterized in terms of additive equivalent internal noise (Ahumada & Watson, 1985; Lu & Dosher, 1999), using observer sensitivity d' measurements obtained in the detection task

$$\sigma_s^2(i,t) = \frac{1}{d'^2(\varepsilon_i, \theta_i)}. \tag{2}$$

The measurement and fitting of the *visibility map*, $d'(\varepsilon_i, \theta_i)$, is described in Visibility Maps and Computing Visibility Maps (see Appendix A). More important, $\sigma_s^2(i, t)$ only represents the variance of the sensory response *noise* (i.e., the variance of the sensory response conditioned on whether or not the target is present in location i). When the true target location J is not specified, the overall variance of the sensory responses must be computed by marginalizing over possible values of $\mu(i)$.

Let σ_d^2 represent the variance of display variable $\mu(i)$. Because the target appears randomly in one of n locations, the display variable is a Bernoulli variable with proportion parameter $p = 1/n$, and its variance is

$$\sigma_d^2 = \frac{1}{n}\left(1 - \frac{1}{n}\right). \tag{3}$$

Finally, because the target location and the sensory noise are independent we can compute the marginal variance of the sensory response, $\sigma_{d+s}^2(i, t)$, as the sum of the display variance and the sensory variance

$$\sigma_{d+s}^2 = \sigma_d^2 + \sigma_s^2(i,t). \tag{4}$$

The Memory Encoding Component

When the task involves k fixation intervals, the observer must encode the sensory responses corresponding to each of n potential target locations in TSM for $k - 1$ of those intervals. Visual information obtained during the final (k^{th}) fixation can be used immediately and so no memory is needed for the final fixation. This is a conservative assumption in that relaxing it would increase the lower bound on TSM capacity required to achieve a given level of performance.

To achieve optimal performance in the search task, the observer must eventually reconstruct the encoded responses from memory and integrate them to compute the target location as described in The Trans-Saccadic Integration Component. If the sensory responses are encoded using fewer bits than required for decoding, then the reconstructed signal will be distorted. Theoretical lower bounds on this distortion for any capacity-limited system can be derived using rate-distortion theory (Berger, 1971; Cover & Thomas, 2006). We characterize distortion in terms of the squared error between the original sensory response $X_s(i, t)$ and the reconstruction of the associated encoded response $X_e(i, t) \sim \mathcal{N}(X_s(i, t), \sigma_e^2(i, t))$, where the encoding variance $\sigma_e^2(i, t)$ represents the minimum achievable expected distortion for a given TSM capacity. In addition to capacity, the encoding variance also depends critically on the variance of the source signals. In particular, Sims et al. (2012) showed that, for a channel with a capacity limit of $r(i, t)$ encoding a Gaussian information source with variance $\sigma_{d+s}^2(i, t)$, the minimum achievable encoding variance is

$$\sigma_e^2(i,t) = \frac{\sigma_{d+s}^2(i,t)}{2^{2r(i,t)} - 1}. \tag{5}$$

While the information source in the current model is not quite Gaussian (i.e., the marginal distribution of $X_s(i, t)$ is a mixture distribution), it is well approximated by a Gaussian source (in The Bernoulli Mixture of Gaussians and its Normal Approximation (see Appendix B) we compute the Shannon lower bound on its rate-distortion function and show that it is virtually indistinguishable from that of a Gaussian information source with equivalent variance). Equation 5 expresses two important relationships. First, the encoding error is directly related to the variance of the sensory response variable. This implies that for a fixed capacity $r(i, t)$, predictable, low-variance, sensory signals will be encoded with less error than unpredictable, high-variance, sensory signals. Second, and more importantly, encoding error is inversely related to memory capacity, such that decreasing the capacity, $r(i, t)$, increases the minimum encoding variance $\sigma_e^2(i, t)$. Moreover, for a given source variance $\sigma_{d+s}^2(i, t)$, this relationship is fixed such that any measured or imputed estimate of the error variance $\sigma_e^2(i, t)$ can be mapped to an equivalent lower bound on capacity. It is this critical relationship that we exploit to derive bounds on TSM capacity from human performance.

The memory capacity allocated to a particular location and fixation interval, $r(i, t)$, depends on the total TSM capacity of the observer, R, as well as the number of potential target locations n and fixation intervals $k - 1$ over which it must be divided. We assume that the sensory responses in the final (k^{th}) interval are available at the time of the decision and need not be encoded in memory. In the main analysis, we assume that available TSM capacity is divided evenly among target locations and fixation intervals; therefore, the capacity allocated to encoding the response from location i and fixation interval t is

$$r(i,t) = \frac{R}{n(k-1)}. \tag{6}$$

This equal allocation assumption is reasonable for our search task, which was designed to provide visual information of roughly constant quality across fixation intervals and potential target locations. However, we will consider the effect of more sophisticated allocation strategies in Dynamically Allocating TSM Capacity.

The Trans-Saccadic Integration Component

The observer's task is to determine the location, J, of the target based on the set of memory-encoded responses $\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)$ gathered over each of the k fixation intervals.

To accomplish this task, the observer must compute, for each possible target location i, the posterior probability, $p(i|\mathbf{X}_e(1), \ldots,$

$\mathbf{X}_e(k)$), that it contains the target. In the Redundancy condition, the expected response $\mu(i, t)$ is identical across fixation intervals and depends only on the target location. Therefore, we assume that the memory-encoded responses are spatially and temporally independent given the target location. This assumption allows us to factor the posterior probability as

$$\begin{aligned} p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) &\propto \pi(i)\prod_{j=1}^{n}\prod_{t=1}^{k} p(X_e(j,t)|i) \\ &= \pi(i)\prod_{t=1}^{k} p(X_e(j,t)|J=i)\prod_{j\neq i}^{n} p(X_e(j,t)|J\neq j) \\ &= \pi(i)\prod_{t=1}^{k}\frac{p(X_e(i,t)|J=i)}{p(X_e(i,t)|J\neq i)}\prod_{j=1}^{n} p(X_e(j,t)|J\neq j), \end{aligned} \tag{7}$$

where $\mathbf{X}_e(t) = X_e(1, t), \ldots, X_e(n, t)$ represents all the responses encoded during fixation interval t and $\pi(i)$ represents the prior probability that the target will appear in location i. Because we are conditioning on the absence of an actual target at location j (i.e., $J \neq j$), the final product term in Equation 7 no longer depends on the target location i. Additionally, because the prior over locations $\pi(i)$ is constant, we can simplify the posterior as

$$p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) \propto \prod_{t=1}^{k}\frac{p(X_e(i,t)|J=i)}{p(X_e(i,t)|J\neq i)}. \tag{8}$$

Substituting in the Gaussian Likelihoods and Simplifying, we find that

$$p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) \propto \exp\left[\sum_{t=1}^{k}\frac{X_e(i,t)}{\sigma^2(i,t)}\right], \tag{9}$$

where

$$\sigma^2(i,t) = \sigma_s^2(i,t) + \sigma_e^2(i,t). \tag{10}$$

is the variance of the noise added to the display signals by the combined effects of sensory representation and memory encoding.

The optimal decision rule for the Redundancy condition consists of selecting the maximum a posteriori location $\hat{J}$. That is,

$$\begin{aligned} \hat{J} &= \arg\max_i \left(\exp\left[\sum_{t=1}^{k}\frac{X_e(i,t)}{\sigma^2(i,t)}\right]\right) \\ &= \arg\max_i \sum_{t=1}^{k}\frac{X_e(i,t)}{\sigma^2(i,t)}. \end{aligned} \tag{11}$$

Equation 11 is essentially equivalent to the decision rule derived in Najemnik and Geisler (2005; and, more generally, in Green & Swets, 1966) for the independent noise case, except that it includes the effects of noise added by the memory encoding/decoding process. In particular, the integration weights ($1/\sigma^2(i, t)$) in Equation 11 are now computed as the inverse of the sum of sensory and memory encoding variances (Equation 10).

In the Uncertainty condition, the target only appears in one fixation interval. This introduces a dependency between fixation intervals (i.e., if the target appears in fixation t, then it cannot appear in any other fixation). We can remove this dependency by conditioning on the target interval u, then marginalizing. The posterior probability is then

$$p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) \propto \pi(i)\sum_{u=1}^{k}\pi(u)\prod_{j=1}^{n}\prod_{t=1}^{k} p(X_e(j,t)|i,u), \tag{12}$$

where $\pi(u)$ represents the prior probability that the target will appear during fixation interval u. Because the prior probabilities on both target location (i) and target interval (u) are both flat and are independent of each other, the prior terms are constants; thus,

$$\begin{aligned} p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) &\propto \sum_{u=1}^{k}\prod_{j=1}^{n}\prod_{t=1}^{k} p(X_e(j,t)|i,u) \\ &= \prod_{j=1}^{n}\prod_{t=1}^{k} p(X_e(j,t)|i,T\neq t)\sum_{u=1}^{k}\frac{p(X_e(j,u)|i,T=u)}{p(X_e(j,u)|i,T\neq u)}, \end{aligned} \tag{13}$$

where T represents the actual target interval. Note that because we are conditioning on target absence ($T \neq t$), the first term on the right-hand side of Equation 13 no longer depends on the target location i. Substituting in the Gaussian likelihoods and simplifying gives

$$\begin{aligned} p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(k)) &\propto \sum_{t=1}^{k}\frac{p(X_e(j,t)|i,T=t)}{p(X_e(j,t)|i,T\neq t)} \\ &= \sum_{t=1}^{k}\exp\left[\frac{X_e(i,t)}{\sigma^2(i,t)}\right]. \end{aligned} \tag{14}$$

Thus, for Uncertainty trials the optimal decision rule consists of selecting the maximum a posteriori location $\hat{J}$ such that

$$\hat{J} = \arg\max_J \sum_{t=1}^{k}\exp\left[\frac{X_e(i,t)}{\sigma^2(i,t)}\right]. \tag{15}$$

In contrast to the decision rule for the Redundancy condition (Equation 11), the sum in Equation 15 occurs outside of the exponentiation (an expansive nonlinearity) that means that the posterior distribution will be dominated by the largest (variance-normalized) response. The result is that imperfect integration and limited memory have a minimal effect on observer performance. In simulating the Uncertainty condition, we assume that ideal observer has unlimited memory for integration. Within the Uncertainty condition of our search experiments, the optimal decision rule (Equation 15) produces behavior indistinguishable from that of a max rule (Nolte & Jaarsma, 1967; Pelli, 1985) that simply selects the target location with the largest overall response. An observer using the max rule requires minimal memory capacity, only enough to store the location and value of the maximum response for comparison with subsequent fixation intervals, and does not do any integration. Therefore, the expected decrement in performance in the Uncertainty condition is entirely because of increasing temporal uncertainty (Lasley & Cohn, 1981), and trans-saccadic integration is not expected to noticeably improve performance. The expected performance of our ideal observer (for a target contrast corresponding to $d' = 1.5$), is shown in Figure 5.

Results

Visibility Maps

Results from the posttest detection experiment were used to estimate target contrast sensitivities for each observer (see Computing Visibility Maps (see Appendix A) for details of the psychophysical model). For the Fixed-Gaze experiment, this consisted simply of computing the sensitivity at each of the eight possible target locations. For the Gaze-Shift experiment, however, contin-

uous visibility maps were computed to allow the effect of arbitrary gaze shifts to be computed. The resulting visibility maps (one for each human observer in the Gaze-Shift experiment) are shown in Figure 6.

Across all participants, the falloff in sensitivity with increasing eccentricity is relatively shallow, especially when compared with results obtained for targets with higher spatial frequencies (e.g., Geisler, Perry, & Najemnik, 2006; Michel & Geisler, 2011; Najemnik & Geisler, 2005). This is by design. We wanted to minimize the impact of saccadic targeting strategies per se and focus instead on the role of trans-saccadic memory, so we chose a low spatial-frequency target whose detectability varies minimally with changes in visual field position.

Search Results

Fixed-Gaze experiment. We simulated TSM capacity-limited ideal observers for each human observer in the Fixed-Gaze experiment using the visibility maps (see Visibility Maps) measured in the posttest detection task. Optimal performance curves for each TSM capacity and condition were computed as the expected performance of the corresponding capacity-limited ideal observer model, averaged over 100,000 simulated trials. Measured performance for individual participants, along with optimal performance curves for four different TSM capacities R, are plotted in Figure 7.

We computed lower bounds on TSM capacity by simulating ideal observers in the Redundancy condition with TSM capacity limits ranging from 0.1 to 32 bits, in increments of 0.1 bits, and selecting the minimum capacity that yielded expected performance equal to or greater than that measured for the corresponding human participant. The estimated capacities (Figure 7, inset values) seem to cluster into two groups: around 8 bits and around 4 bits. An interesting feature of the data are that participants whose performance in the (capacity-insensitive) Uncertainty condition approximates the performance of their peripheral-sensitivity-matched ideal observers (i.e., AS, RGJ, and SEK) fall into the higher 8-bit cluster, while those whose performance in the Uncertainty condition falls significantly short of ideal fall into the lower 4-bit cluster. This suggests that the small lower bounds computed for the latter group may underestimate their actual memory capacity.

Individual estimates for both memory-intensive redundant conditions (two-interval and four-interval) are listed in Table 1. A Wilcoxon's rank-sums test showed that the capacity estimates did not vary significantly across these two conditions $W = 25, p = .69$.

The range of estimated capacities in Table 1 is consistent with the range of approximately 3–8 bit capacities reported by Sims et al. (2012) for VSTM. What these capacities mean in terms of an

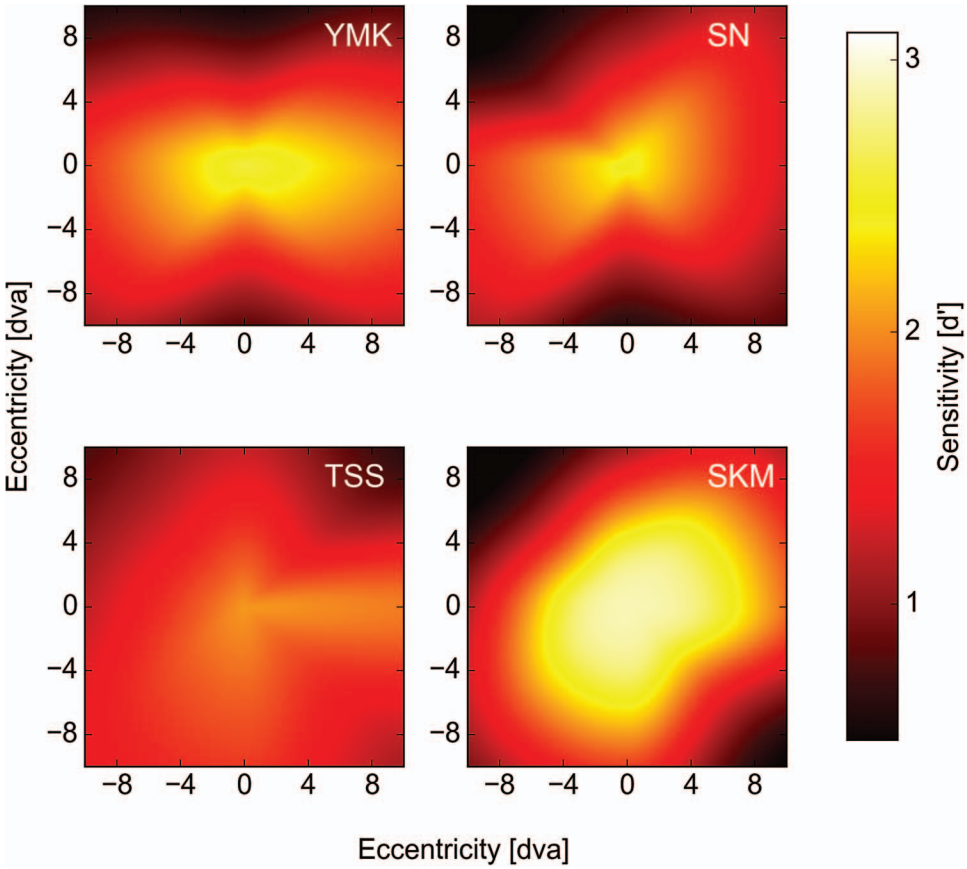

Figure 6. Individual visibility maps for participants in the Gaze-Shift experiment. Brighter colors indicate greater sensitivity. See the online article for the color version of this figure.

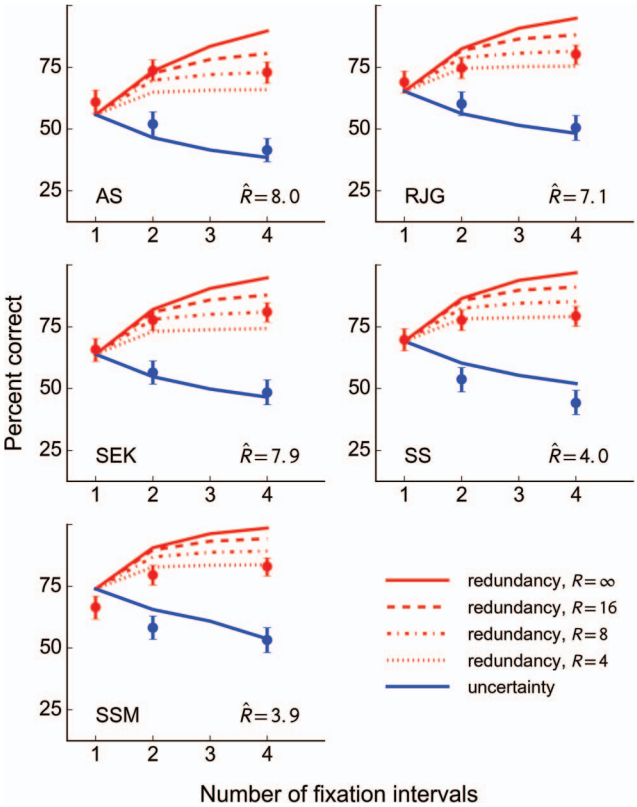

Figure 7. Performance in the Fixed-Gaze search task. Each panel represents a different participant. Markers represent human data, with 95% confidence interval's. Curves represent simulation results for Uncertainty (blue, lower curve) and Redundancy (red, upper curve) trials. Broken red curves indicate performance for ideal observers with various trans-saccadic memory (TSM) capacity limits. Inset values ($\hat{R}$) in each panel indicate estimated capacities for the four-interval condition. See the online article for the color version of this figure.

item-based capacity limit depends in part on the structure of the task. For example, in a task that required participants to recall which of two independently selected and highly discriminable symbols was presented at each of n locations, an observer with a memory capacity of four bits could recall four of these symbols without error, an observer with a capacity of eight bits could recall eight of these symbols without error. However, if each location could contain one of four symbols (rather than one of two), then an observer with four bits of memory could only recall two locations without error and an observer with eight bits of memory could only remember four symbols. Thus, the range of capacities in Table 1 is also broadly consistent with the 4 ± 1 item limit traditionally reported for VSTM.

Table 1
Individual Capacity Estimates for the Fixed-Gaze Search Experiment

	Redundancy condition	
Participant	Two intervals	Four intervals
AS	20.2	8.0
RJG	4.3	7.1
SEK	7.7	7.9
SS	3.6	4.0
SSM	2.1	3.9

Gaze-Shift experiment. Optimal performance curves for the Gaze-Shift experiment were estimated in a similar manner to those in the Fixed-Gaze experiment. A critical difference was that we had to account for the effects of the gaze shifts. Thus, when simulating individual trials, the ideal observer model used the interpolated visibility map for the corresponding participant to determine $d'(c, \varepsilon_t, \theta_t)$ based on the eye-position measured in fixation interval t. We computed the expected performance of the ideal observers by simulating 250 repetitions of each trial for a total of 100,000 simulations per condition. All other aspects of the analysis were identical to those used for the Fixed-Gaze experiment. Optimal performance curves and measured performance for individual participants are plotted in Figure 8.

Individual lower bounds on TSM capacity were computed as in the Fixed-Gaze experiment (Fixed-Gaze Experiment), and the resulting estimates are listed in Table 2. These estimates differ from those obtained in the Fixed-Gaze experiment in two respects. First, the capacities estimated for the four-interval condition (M = 21.5) are significantly larger than those estimated for two-interval condition (M = 4.6), Wilcoxon's W = 10, p = .028. This suggests that human observers make more efficient use of TSM memory when they must store information from more than a single fixation. Second, the larger capacities estimated for the four-interval condition in the Gaze-Shift experiment (M = 21.5) are significantly larger than those estimated in the Fixed-Gaze experiment (M = 6.2) W = 10, p = .016. This suggests that visual information is

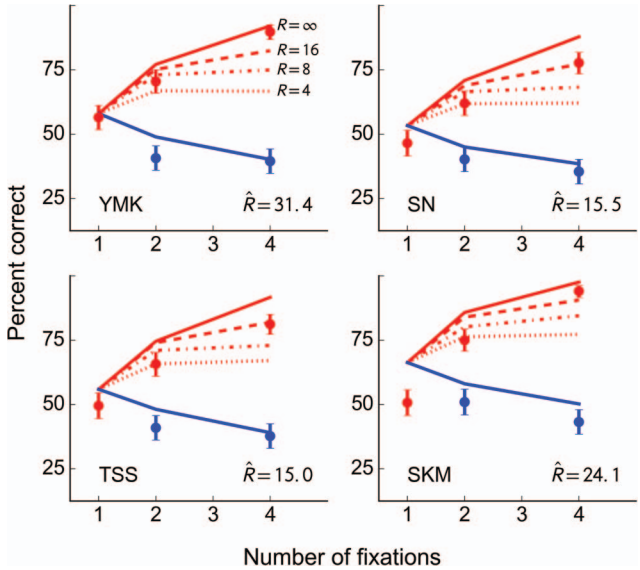

Figure 8. Performance in the Gaze-Shift search task. Each panel represents a different participant. Markers represent human accuracy, with 95% confidence intervals. Curves represent optimal performance curves for Uncertainty (blue, lower curve) and Redundancy (red, upper curve) trials. Broken red curves indicate performance for ideal observers with various trans-saccadic memory (TSM) capacity limits. See the online article for the color version of this figure.

Table 2
Individual Capacity Estimates for the Gaze-Shift Search Experiment

	Redundancy condition	
Participant	Two intervals	Four intervals
YMK	6.1	31.4
SN	3.9	15.5
TSS	4.0	15.0
SKM	4.3	24.1

encoded more efficiently into memory or that memory capacity is greater when the task involves actual eye movements than when it does not.

Dynamically allocating TSM capacity. The large capacities measured for the four-interval condition of the Gaze-Shift experiment are surprising and suggest that the capacity of TSM (measured in the Gaze-Shift experiment) may be larger than the capacity of traditional VSTM (measured in the Fixed-Gaze experiment). In evaluating this possibility, we must consider and account for factors that might inflate the lower-bound estimates obtained for the Gaze-Shift experiment. One such factor is the potential use of a dynamic memory allocation strategy. The ideal observer formulation presented in The Memory Encoding Component relies on the simplifying assumption that TSM capacity is divided evenly among potential target locations and fixation intervals, an assumption analogous to the equal-precision assumption typical in continuous-resource models of VSTM (van den Berg, Shin, Chou, George, & Ma, 2012). Because the search experiments were designed so that each interval and location contained information of approximately equal quality, this is a sensible encoding strategy. However, this is not the only plausible encoding strategy. An alternative class of encoding strategies involves adjusting the capacity allocated to each target location dynamically, to prioritize information at a particular location or subset of locations (Bays & Husain, 2008; van den Berg et al., 2012). In particular, observers encoding a search display into memory might prioritize information from locations that are more likely to contain the target. For example, if an observer is confident after gathering information from an initial fixation that the target is located in one of three locations, then the observer might choose to allocate the bulk of the available TSM capacity toward encoding information from those three locations. This strategy helps reduce the effective encoding noise at the most likely target locations (at the cost of increasing noise at the unlikely locations). To characterize the impact of this type of dynamic memory allocation strategy on search performance, we devised a simple probability-weighted allocation model in which the fraction of memory capacity allocated to target location i during fixation t is related to the posterior target probability by

$$r(i,t) \propto p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(t))^{\gamma}, \quad (16)$$

where $p(i|\mathbf{X}_e(1), \ldots, \mathbf{X}_e(t))$ (see Equation 9) represents the posterior probability associated with location i computed up to the t^{th} fixation, and γ is a parameter controlling the *posterior weight*, the degree to which the posterior probability determines the dynamic allocation of TSM capacity. For example, when $\gamma = 0$ TSM capacity is allocated evenly among all potential target locations; when $\gamma = 1$ then TSM capacity is allocated in direct proportion to the normalized posterior in Equation 9; and as γ approaches infinity, TSM capacity is devoted entirely to encoding information from the most likely target location (a strategy similar to the "maximum-of-outputs" rule).

The dynamic encoding strategy described in Equation 16 will differ most from the equal-allocation strategy when the posterior distribution is narrow and informative (i.e., when some potential locations can be ruled out). This might occur in the Gaze-Shift task, where the quality of visual information provided in a fixation interval might vary significantly as a function of the eye position during the fixation. Moreover, any potential benefit should be most evident for the four-fixation interval Redundancy condition that requires encoding and decoding visual information from three fixations into and from TSM. Therefore, we evaluated the dynamic allocation strategy in the four-interval redundancy condition as follows. First, we computed optimal performance curves as a function of both the total available TSM capacity, R, and the probability-weight exponent, γ. As in Gaze-Shift Experiment, we computed these curves by simulating 250 repetitions of each trial (100,000 total trials) for each combination of parameters sampled along a grid, and computing the average accuracy of the capacity-limited ideal observer across these repetitions. The results for the Gaze-Shift experiment are plotted in Figure 9.

These results show that the effect of the dynamic allocation strategy varies systematically with TSM capacity. For large TSM capacities (e.g., 16 or 32 bits) that allow search accuracy to approach the capacity-unlimited performance ceiling (see the

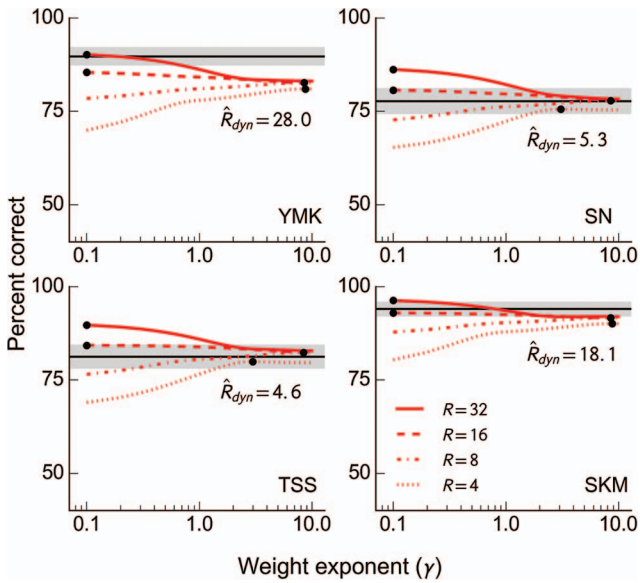

Figure 9. Effect of dynamic allocation on search performance. Proportion correct as a function of posterior weighting function exponent (γ) for individual observers in the Redundancy condition of the Gaze-Shift search task. Different linestyles represent performance for ideal observers with different capacity limits (32, 16, 8, or 4 bits). The black markers indicate the optimal weight exponent γ for each of the plotted capacity limits. The black lines and gray shaded areas indicate measured human accuracy and 95% confidence intervals. See the online article for the color version of this figure.

curves for $R = \infty$ in Figure 8), the capacity-limited observer receives little or no benefit from dynamically allocating TSM capacity. The optimal weighting occurs near $\gamma = 0$, which corresponds to an equitable allocation. By contrast, as TSM capacity decreases, peak performance is achieved at progressively larger values of γ, which corresponds to a progression toward the "max"-like strategy of encoding only information from the most likely target location. This is a novel finding: in a demanding search task, observers with small memory capacities should benefit from a strategy that redistributes TSM capacity dynamically according to the posterior probability at the time of encoding.

Finally, we evaluated the effect of dynamic allocation on our estimates of TSM-capacity by selecting the smallest capacity $\widehat{R}_{dyn}$, computed across all probability weight exponents γ, that allowed the capacity-limited ideal observer to equal or exceed the measured human performance. The resulting values for each participant are shown in Figure 9. For two of the participants (YMK and SKM), the capacity estimates ($M = 23.1$) computed after accounting for the dynamic allocation strategy remain essentially unchanged. However, for the other two participants (SN and TSS) the capacity estimates ($M = 5.0$) drop to within the 4–8 bit range estimated for the Fixed-Gaze task. Overall, this result indicates that observers can potentially benefit from a dynamic posterior-weighting strategy. However, the dynamic strategy cannot account completely for the difference in capacities computed for the Fixed-Gaze and Gaze-Shift experiments.

Discussion

We developed a normative framework, the capacity-limited ideal observer model, for evaluating the effect of TSM capacity on visual search performance. This framework extends a recent model of capacity-limited memory in change-detection (Sims et al., 2012) to deal with the integration of information across gaze shifts in a visual search task.

Our framework is not the first ideal observer analysis to focus on the transsaccadic integration of information. Wolf and Schutz (2015) and Ganmor, Landy, and Simoncelli (2015) for example, used ideal observer models to determine how efficiently observers combine information from foveal and peripheral retinal patches across a single saccade, and the ideal searcher analyses in Najemnik and Geisler (2005) and Geisler et al. (2006) examine the efficiency of search across many saccades. Other work on transsaccadic integration has focused on the timecourse of information integration across saccades in search (e.g., Caspi, Beutter, & Eckstein, 2004; Ludwig, Davies, & Eckstein, 2014), showing that multiple-fixation search does make use of memory across several fixations. The capacity-limited ideal observer model expands upon these by providing a principled way to examine the impact of memory capacity on transsaccadic integration.

A critical feature of the capacity-limited ideal observer model is that the effect of capacity limits is derived a priori from information-theoretic principles. For any specified TSM capacity, the ideal observer model makes parameter-free predictions of search performance. Conversely, for any specified search performance, the ideal observer model allows us to determine the TSM capacity required to achieve this performance. We used the capacity-limited ideal observer model in this way to estimate and compare memory capacity in a Fixed-Gaze experiment that required storage and integration of information in VSTM and a Gaze-Shift experiment that required storage and integration of information across saccades in TSM.

Most capacity estimates in the Fixed-Gaze experiment were between 4–8 bits, which is broadly consistent with the values reported by Sims et al. (2012), who found capacity estimates ranging from about 3–6 bits for orientation and 4–8 bits for line length in a change-detection task. Because the Sims et al. (2012) estimates are for VSTM capacity, the fact that our estimates are consistent provides support for the hypothesis that VSTM measured conventionally, and VSTM measured via a sequence of simulated fixations, are mediated by common or comparable systems.

Analysis of the Gaze-Shift experiment yielded capacity estimates (15–32 bits) significantly greater than those found in the Fixed-Gaze task. The most notable difference between the two experiments was the substitution of a simulated saccadic transient in place of an actual saccade in the Fixed-Gaze task. Therefore the discrepancy in estimates obtained across the two tasks hints at a fundamental difference between the effect of a simulated saccadic transient and the effect of executing an actual saccade. Since capacity estimates were significantly higher in the Gaze-Shift experiment than in both the Fixed-Gaze experiment and the study conducted by Sims et al. (2012), the difference in estimated capacities suggests that the Fixed-Gaze experiment was tapping convential VSTM, and that the TSM used in the Gaze-Shift experiment might be mediated by a different memory system with a greater capacity than VSTM.

However, there are other potential explanations for the difference in estimated capacities. One possibility is that the masking produced by the simulated saccadic transient exceeded that produced by the retinal slip of the stimulus during an actual saccade. For example, if the simulated saccadic transients produced significant backward masking, but the similar retinal stimuli in the Gaze-Shift experiment were suppressed by the saccade (e.g., Ross, Morrone, Goldberg, & Burr, 2001), then the estimated memory capacity would be artificially low for the Fixed-Gaze experiment but not for the Gaze-Shift experiment. However, two findings appear to rule out this explanation. First, a control version of the detection task (data not shown) that did not include the simulated saccadic transient yielded detection thresholds indistinguishable from those in the original task. Second, the capacities measured for the Fixed-Gaze task agreed closely with those measured for traditional VSTM (Sims et al., 2012).

A second possibility is that participants in the Gaze-Shift experiment used a flexible encoding strategy that dynamically adjusted the proportion of available memory capacity allocated to each potential target location based on the information received from previous fixation intervals. Such a strategy could potentially provide a more efficient memory encoding, lowering the capacity required to account for the measured search accuracy. In this case, the capacity estimates obtained in the Gaze-Shift experiment would be artificially high. We evaluated this possibility in a secondary analysis (see Dynamically Allocating TSM Capacity) and found that accounting for a dynamic memory allocation strategy could indeed lower TSM capacity estimates. Simulations of the memory-limited ideal observer showed that the performance of two human participants in the Gaze-Shift experiment could be explained by dynamic allocation of a small ($M = 5.0$) capacity.

However, for the other two human participants in the Gaze-Shift experiment, the optimal dynamic allocation strategy had only a negligible effect on estimated capacity ($M = 23.1$). Whether or not human observers actually use a dynamic allocation strategy, this result suggests that such a strategy cannot completely account for the observed difference in our observed measurements of TSM and VSTM capacities.

A third possibility is that the retinal shift of target locations across fixation intervals imposed by saccades in the Gaze-Shift experiment helps to decorrelate noise in the sensory input and, thus, to lower the effective signal-to-noise ratio of the integrated sensory responses (a similar role has been suggested for microsaccades by Kuang, Poletti, Victor, & Rucci, 2012, but for temporal correlations in external rather than internal noise). The advantage of integrating across noisy samples of a common stimulus is greatest when the noise is uncorrelated (independent) across the samples (e.g., perfectly correlated samples add no additional information, while independent samples reduce the noise by a factor of $1/\sqrt{\text{nr. samples}}$). Though the $1/f$ noise masks used in the search tasks were statistically independent across fixation intervals, visual sensitivity measurements indicate that the sensory response noise was dominated overwhelmingly by internal factors (i.e., average absolute detection efficiency was less than 10%), which might give rise to correlations.

Measurements in a number of visual areas, including retina (Pillow et al., 2008) and primary visual cortex (Smith & Kohn, 2008), show that neural population responses exhibit substantial temporal correlations. Moreover, these correlations are spatially localized and fall off dramatically with increasing retinotopic distance. In the Fixed-Gaze task, the visual information falls on approximately the same retinal locations across intervals, which means that the sensory response noise is likely correlated across intervals. In the Gaze-Shift task, however, visual information from each potential target location falls on a different part of the retina in each fixation interval. This means that the sensory responses associated with a potential target location are less likely to be correlated in time (across fixations), and more likely to be uncorrelated and informative. It follows then that sensory responses in the Gaze-Shift search task would be less correlated in time than the sensory responses in the Fixed-Gaze search task, resulting in higher memory capacity estimates in the Gaze-Shift search task. Thus, if local temporal correlations do play a large role in determining the sensory information available for decoding, then the *effective* capacity of TSM might be larger than that of VSTM even if they are mediated by a common system. Teasing apart the effects of sensory correlations from those of limited memory capacity will require further research.

More generally, because the goal of our ideal observer analysis was to determine how transsaccadic memory limitations *should* impact integration and search performance, we did not explicitly consider other limits on human temporal integration. It is possible (and even likely) that some of the inefficiencies that contribute to apparent memory capacity limitations are in fact because of fundamental visual temporal integration mechanisms unrelated to memory (e.g., Eckstein, Whiting, & Thomas, 1996b, 1996a). To the extent that such inefficiencies contribute to a decline in search performance, they reduce the tightness of our estimated lower bounds on memory capacity, which means that our reported bounds may underestimate the capacity of human transsaccadic memory.

Conclusion

By combining measurement of human search performance in gaze-fixed and multiple-fixation search tasks with an ideal observer analysis of memory capacity-limited performance, we have shown that the effective capacity of trans-saccadic memory (measured in bits) is at least as large as that of conventional visual STM and may potentially be larger. The high effective capacity of TSM relative to that of conventional VSTM measured in this study could be because of factors that come into play during natural saccades. These include sensory factors, namely the decorrelation of visual noise across glances because of the changes in location of the retinal image, and strategic factors, namely the dynamic allocation of memory across glances. These factors can operate to enhance effective memory in the case of natural saccades and to depress effective memory in artificial, fixed-gaze laboratory tasks. Future work on visual memory must continue using natural saccades along with the computational modeling approaches such as those developed here across a wide variety of tasks.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 529–543. http://dx.doi.org/10.1037/0096-1523.15.3.529

Ackermann, J. F., & Landy, M. S. (2013). Choice of saccade endpoint under risk. *Journal of Vision, 13,* 27. http://dx.doi.org/10.1167/13.3.27

Ahumada, A. J., & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 1133–1139.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15,* 106–111. http://dx.doi.org/10.1111/j.0963-7214.2004.01502006.x

Bahill, A. T., Brockenbrough, A., & Troost, B. T. (1981). Variability and development of a normative data base for saccadic eye movements. *Investigative Ophthalmology & Visual Science, 21,* 116–125.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences, 24,* 191–204. http://dx.doi.org/10.1016/0025-5564(75)90075-9

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854. http://dx.doi.org/10.1126/science.1158023

Berger, T. (1971). *Rate distortion theory: A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice Hall.

Bienaymé, I.-J. (1867). Considérations à l'appui de la dècouverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés [Considerations in Support of Laplace's Finding on Probability in the Method of Least Squares]. *Journal de Mathématiques Pure et Appliquées, 12,* 158–176.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11,* 4. http://dx.doi.org/10.1167/11.5.4

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of National Academy of Sciences, 101,* 13086–13090.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/047174882X

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24,* 87–114; discussion 114–185.

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science, 9,* 111–118. http://dx.doi.org/10.1111/1467-9280.00020

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11,* 14. http://dx.doi.org/10.1167/11.5.14

Eckstein, M. P., Whiting, J. S., & Thomas, J. P. (1996a). Detection and contrast discrimination of moving signals in uncorrelated Gaussian noise. *Proc SPIE, 2712,* 9–25.

Eckstein, M. P., Whiting, J. S., & Thomas, J. P. (1996b). Role of knowledge in human visual temporal integration in spatiotemporal noise. *Journal of the Optical Society of America, 13,* 1960–1968.

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,* 3. http://dx.doi.org/10.1167/11.12.3

Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis, 17,* 238–246. http://dx.doi.org/10.1137/0717021

Ganmor, E., Landy, M. S., & Simoncelli, E. P. (2015). Near optimal integration of orientation information across saccades. *Journal of Vision, 15,* 1–12.

Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825–837). Boston, MA: MIT Press.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51,* 771–781. http://dx.doi.org/10.1016/j.visres.2010.09.027

Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision, 6,* 858–873. http://dx.doi.org/10.1167/6.9.1

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23,* 420–456. http://dx.doi.org/10.1016/0010-0285(91)90015-G

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 307–317. http://dx.doi.org/10.1037//0278-7393.18.2.307

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception and Psychophysics, 34,* 49–57. http://dx.doi.org/10.3758/BF03205895

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39,* 2729–2737. http://dx.doi.org/10.1016/S0042-6989(98)00285-5

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology, 22,* 510–514.

Lasley, D. J., & Cohn, T. (1981). Detection of luminance increment: Effect of temporal uncertainty. *Journal of the Optical Society of America, 71,* 845–850. http://dx.doi.org/10.1364/JOSA.71.000845

Lu, Z.-L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 16,* 764–778.

Luck, S. J. (2008). Visual short term memory. In S. J. Luck & A. Hollingworth (Eds.), *Visual memory* (chap. 3, pp. 247–289). New York, NY: Oxford University Press. http://dx.doi.org/10.4249/scholarpedia.3328

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of National Academy of Sciences, 111,* E291–E299.

Michel, M. M., & Geisler, W. S. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision, 11,* 1–18. http://dx.doi.org/10.1167/11.1.18.Introduction

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391. http://dx.doi.org/10.1038/nature03390

Nolte, L., & Jaarsma, D. (1967). More on the detection of one of M orthogonal signals. *The Journal of the Acoustical Society of America, 41,* 497.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34,* 1703–1721.

Peli, E., Yang, J., & Goldstein, R. B. (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America A, 8,* 1762–1774. http://dx.doi.org/10.1364/JOSAA.8.001762

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 1508–1532.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290. http://dx.doi.org/10.3758/BF03203943

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., . . . Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature, 454,* 995–999. http://dx.doi.org/10.1038/nature07140

Prime, S., Tsotsos, L., Keith, G., & Crawford, J. (2007). Visual memory capacity in transsaccadic integration. *Experimental Brain Research, 180,* 609–628. http://dx.doi.org/10.1007/s00221-007-0885-4

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 261–282. http://dx.doi.org/10.1037/0096-1523.24.1.261

Prinzmetal, W., Nwachuku, I., Bodanski, L., Blumenfeld, L., & Shimizu, N. (1997). The phenomenology of attention: 2. Brightness and contrast. *Consciousness and Cognition, 6,* 372–412. http://dx.doi.org/10.1006/ccog.1997.0313

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24,* 113–121. http://dx.doi.org/10.1016/S0166-2236(00)01685-4

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27,* 623–656. http://dx.doi.org/10.1145/584091.584093

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. In *Institute of Radio Engineers, International Convention Record* (Vol. 7, pp. 142–163). http://dx.doi.org/10.1234/12345678

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119,* 807–830. http://dx.doi.org/10.1037/a0029856

Smith, M. A., & Kohn, A. (2008). Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience, 28,* 12591–12603. http://dx.doi.org/10.1523/JNEUROSCI.2929-08.2008

Swensson, R. G., & Judy, P. F. (1981). Detection of noisy visual targets: Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics, 29,* 521–534.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review, 121,* 124–149. http://dx.doi.org/10.1037/a0035234

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory

limitations. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 8780–8785. http://dx.doi.org/10.1073/pnas.1117465109

Verghese, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research, 34,* 2453–2461.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4,* 11. http://dx.doi.org/10.1167/4.12.11

Wolf, C., & Schutz, A. C. (2015). Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *Journal of Vision, 15,* 1–18.

Wolfe, J. M. (2007). Guided search 4. 0 current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–120). Oxford, United Kingdom: Oxford University Press. http://dx.doi.org/10.1167/1.3.349

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12,* 219–224. http://dx.doi.org/10.1111/1467-9280.00339

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 458–468.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–U13. http://dx.doi.org/10.1038/Nature06860

Appendix A

Computing Visibility Maps

Visual sensitivity for each human observer was characterized in terms of a *visibility map* that describes the effective signal-to-noise ratio as a function of target contrast c and retinal position (ε, θ), where ε represents eccentricity and θ represents angular direction in the visual field. The visibility map was computed by taking the inverse of the standard normal integral of a function describing accuracy in the posttest detection data

$$d'(c, \varepsilon, \theta) = \sqrt{2}\Phi^{-1}[\mathrm{PC}(c, \varepsilon, \theta)] \quad \text{(A1)}$$

where $\Phi(\cdot)$ represents the standard normal integral, and $\mathrm{PC}(c, \varepsilon, \theta)$ is a psychometric function representing the proportion correct in the 2IFC detection task. As in Najemnik and Geisler (2005), the factor $\sqrt{2}$ accounts for the fact that there were two intervals in the detection task, but only a single interval per fixation of the search task. We modeled the detection accuracy at a particular location (ε, θ) as a cumulative Weibull function of target contrast c

$$\mathrm{PC}(c, \varepsilon, \theta) = 1 - 0.5\exp\left[-\left(\frac{c}{\alpha(\varepsilon, \theta)}\right)^{\beta}\right], \quad \text{(A2)}$$

where β is a parameter controlling the steepness of the psychometric function and α is a contrast threshold parameter that varies with the retinal position of the target. Individual estimates of the steepness parameter computed for different target locations did not vary significantly from each other, so we fit the visibility maps assuming a common steepness parameter β. This is consistent with detection measurements reported recently (Ackermann & Landy, 2013) for targets of similar spatial frequency in $1/f$ noise.

In the Fixed-Gaze experiment the target could only appear at one of eight retinal locations, so the visibility map consisted of a set of eight discrete thresholds $\alpha(\varepsilon_i, \theta_i)$, $i \in \{1, \ldots, 8\}$. These thresholds were obtained via maximum likelihood by fitting Equation A1 directly to the proportion correct obtained at each of the possible target locations.

In the Gaze-Shift experiment, participants made saccades and the retinal position of the target shifted continuously as a function of gaze angle. Therefore, this experiment required a more continuous representation of the visibility map. To compute the visibility map, we modeled contrast thresholds $\alpha(\varepsilon, \theta)$ as a log-linear function of retinal position

$$\alpha(\varepsilon, \theta) = \alpha_0 \exp(\tau_\theta \varepsilon), \quad \text{(A3)}$$

where α_0 represents the foveal contrast threshold and τ_θ is a log slope parameter controlling the rise in contrast thresholds as a function of eccentricity.

This function has been shown to accurately describe the rise in contrast thresholds with increasing eccentricity across a variety of tasks (Peli, Yang, & Goldstein, 1991). The resulting psychometric model has 10 parameters: $\alpha(0)$, β, and a separate log-slope parameter τ_{θ_i} for each of the eight directions along which we measured detection performance $\theta_i \in \{0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°\}$, where $i = \theta/45°$ represents the direction index. We used maximum likelihood to fit this model to each observer's detection data. For intermediate values of θ not measured in the detection experiment (i.e., for nonintegral i), we computed τ_θ by using a monontonic piecewise cubic function (Fritsch & Carlson, 1980) to interpolate smoothly between the nearest measured values.

(*Appendices continue*)

Appendix B

The Bernoulli Mixture of Gaussians and Its Normal Approximation

The marginal sensory response distribution $p(S)$ corresponding to a particular location and fixation interval in the Redundancy condition (its conditioned counterpart is labeled $p(X_s(i, t))$ in the text) is a Bernoulli mixture of two Gaussian distributions with different means, that is,

$$p(S) = p\phi\left[\frac{S-0.5}{\sigma_s}\right] + (1-p)\phi\left[\frac{S+0.5}{\sigma_s}\right], \quad \text{(B1)}$$

where $\phi(\cdot)$ represents the standard normal density, $p = 1/n$ represents the proportion of locations in which the target signal is present, and the first and second terms in the sum represent the contributions of the target-present and target-absent response densities, respectively.

Alternatively, the (marginal) sensory response variable can be formulated as the sum of independent Bernoulli and Gaussian variables, that is,

$$S = B + N, \quad \text{(B2)}$$

where $B \sim Bernoulli(p) - 0.5$ is a (shifted) Bernoulli variable representing the mean of the target-present or target-absent response and $N \sim \mathcal{N}(0, \sigma_s^2)$ represents the equivalent sensory noise. These two interpretations are mathematically equivalent, but in the exposition below we focus on the latter interpretation because the expressions for the variance and rate-distortion function of S are more easily motivated under this interpretation.

Equation 3 depends on two distinct assumptions entailed in the model described above: that the target appears at a particular location i with probability $p = 1/n$, and that the difference between the variable B under the target present and target absent conditions is 1. As long as these assumptions are met, then the variance of the display variable will be that of a Bernoulli variable with $p = 1/n$. Moreover, because the Bernoulli (B) and Gaussian (N) components are independent, and each has finite variance, Bienaymé's formula for the variance of the sum of uncorrelated variables (Bienaymé, 1867) applies and, thus,

$$\mathrm{Var}[B + N] = \mathrm{Var}[B] + \mathrm{Var}[N]. \quad \text{(B3)}$$

This is the result shown in Equation 4.

As described above, the marginal distribution of the sensory response S in our model is a Gaussian mixture distribution. Nonetheless, Equation 5 implicitly models this marginal distribution as a normal distribution with variance equal to that of the mixture distribution. How accurate is this approximation?

If the means of the two Gaussian components were far apart and if each component contributed nearly equally to the mixture, then the (unimodal) normal approximation to the (bimodal) mixture would be quite poor (these conditions would occur, for example, if the target were highly detectable and there were only two potential target locations). In our experiments, however, the target-absent component dominates the mixture and the distance two means is small relative to the component variance σ_s^2. As a result, a normal distribution with variance computed as in Equation 4 of the manuscript approximates the mixture distribution very well. In Figure B1 we demonstrate this by directly comparing the two distributions when $d' = 1.5$, which was the average value measured for the target contrasts and locations used in the search experiments.

It is of course possible that these small apparent differences in the distributions will nonetheless have large consequences for the resulting rate-distortion functions. In the next section, we derive the Shannon lower-bound on the rate-distortion function for the mixture variable. We then compare this lower bound with the rate-distortion function for its Gaussian-source approximation and show that the two functions are virtually indistinguishable.

Rate-Distortion Lower Bound for a Bernoulli Mixture of Gaussians

In this section, we show that the Shannon lower-bound (SLB) on the rate-distortion function for the sensory response variable S, whose actual distribution is a Bernoulli Mixture of Gaussians, closely matches the rate-distortion function for a Gaussian variable with equivalent variance used to approximate S in the memory-limited searcher model. Because the goal of the current study is to estimate lower bounds on memory capacity, it is not necessary to show that the SLB corresponds to an achievable rate-distortion function (that is difficult), but only to show that the SLB for the actual distribution of S is not significantly lower than that of the approximating distribution used in the model (Figure B2).

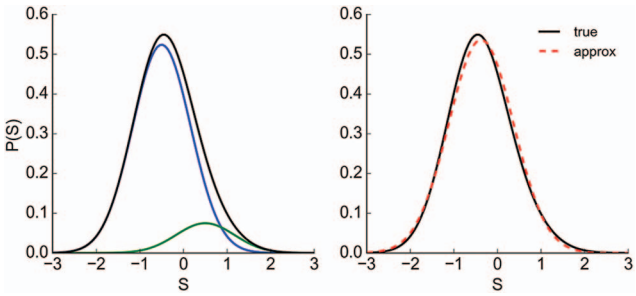

Figure B1. Comparison of the Gaussian mixture density and its normal approximation for parameters similar to those used in the ideal observer model. Left panel: the target-absent (blue) and target-present (green) components of the mixture density (black). Right panel: the true (mixture) target density (solid black curve) and its normal approximation (dashed red curve). See the online article for the color version of this figure.

(*Appendices continue*)

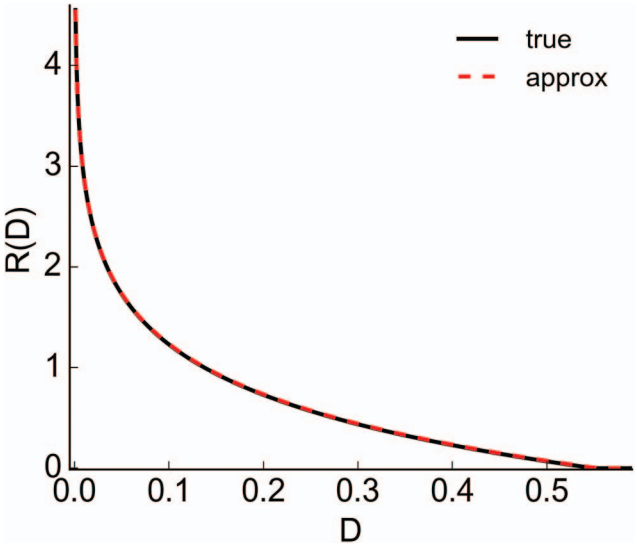

Figure B2. Comparison of rate-distortion Shannon lower bounds (SLBs) for the Bernoulli mixture of Gaussians variable and for its normal approximation when $d' = 1.5$ (both shown in Figure B2). Note that for the normal approximation, the rate-distortion function is equal to its corresponding SLB (Cover & Thomas, 2006, Theorem 10.3.2). See the online article for the color version of this figure.

Assuming a squared error distortion measure, the rate-distortion SLB (Berger, 1971; Equation 4.6.14) for a source S is

$$R_S(D) = h(S) - \frac{1}{2}\log(2\pi eD). \tag{B4}$$

We can use the *chain rule* for conditional entropy to break up the $h(S)$ term. Thus,

$$\begin{aligned} R_S(D) &= h(S) - \frac{1}{2}\log(2\pi eD) \\ &= H(B) - H(B|S) + h(S|B) - \frac{1}{2}\log(2\pi eD), \end{aligned} \tag{B5}$$

where $h(\cdot)$ represents *differential entropy*. Note that because $p(S|B)$ is a normal distribution with the *SD* of N,

$$h(S|B) = h(N) = \frac{1}{2}\log(2\pi e\sigma_s^2).$$

Plugging this into Equation b5, we find that

$$\begin{aligned} R_S(D) &= H(B) - H(B|S) + \frac{1}{2}\log(2\pi e\sigma_s^2) - \frac{1}{2}\log(2\pi eD) \\ &= H(B) - H(B|S) + \frac{1}{2}\log\left(\frac{\sigma_s^2}{D}\right) \\ &= H_b(p) + \frac{1}{2}\log\left(\frac{\sigma_s^2}{D}\right) - H(B|S), \end{aligned} \tag{B6}$$

where $H_b(p)$ is the binary entropy function

$$H_b(p) = -p\log(p) - (1-p)\log(1-p).$$

The only remaining entropy term to compute is $H(B|S)$, the conditional entropy of the class B given S:

$$\begin{aligned} H(B|S) &= \int_{s\in S} p(s)H(B|S=s)ds \\ &= -\int_{s\in S} p(s)\sum_{b\in B} p(b|s)\log\left[p(b|s)\right]ds \\ &= \int_{s\in S} p(s,b)\sum_{b\in B}\log\frac{p(s)}{p(s,b)}ds \\ &= \sum_{b\in B} p(b)\int_{s\in S} p(s|b)\log\frac{p(s)}{p(s,b)}ds \\ &= p(b_0)\int_{s\in S} p(s|b_0)\log\frac{p(s)}{p(s,b_0)}ds + p(b_1)\int_{s\in S} p(s|b_1)\log\frac{p(s)}{p(s,b_1)}ds. \end{aligned} \tag{B7}$$

Finally, substituting in the Gaussian densities yields

$$\begin{aligned} H(B|S) &= p\int_{-\infty}^{\infty}\phi\left[\frac{s-b_0}{\sigma_s}\right]\log\left(p + (1-p)\frac{\phi\left[\frac{s-b_1}{\sigma_s}\right]}{\phi\left[\frac{s-b_0}{\sigma_s}\right]}\right)ds \\ &\quad + (1-p)\int_{-\infty}^{\infty}\phi\left[\frac{s-b_1}{\sigma_s}\right]\log\left((1-p) + p\frac{\phi\left[\frac{s-b_0}{\sigma_s}\right]}{\phi\left[\frac{s-b_1}{\sigma_s}\right]}\right)ds, \end{aligned} \tag{B8}$$

where $\phi[\cdot]$ represents the standard normal pdf.

In computing the rate-distortion SLB for the mixture model, we evaluated the integrals in Equation B8 numerically (using Gauss-Hermite quadrature) and with values based on the model parameters described in the text. That is, with $p = 1/8$, $B = \{-0.5, 0.5\}$, and $\sigma_s \in [1/2, 1]$.

(*Appendices continue*)

Appendix C

Mathematical Notation

	Ideal Observer Model
$i, j \in \{1, \ldots, n\}$	an index over potential target locations
$t, u \in \{1, \ldots, k\}$	an index over fixation intervals
J	the true target location
$\mu(i, t)$	the expected sensory response at location i and interval t
$X_s(i, t)$	the sensory response at location i and interval t
$X_e(i, t)$	the memory-encoded response for location i and interval t
$\boldsymbol{\mu}(t)$, $\mathbf{X}_s(t)$, $\mathbf{X}_e(t)$	vectors consisting of all signals/responses from interval t
$\sigma_s^2(i, t)$	the variance of the sensory noise associated with location i during interval t
$\sigma_{d+s}^2(i, t)$	the marginal (unconditioned) variance of the sensory response associated with location i and interval t
$\sigma_e^2(i, t)$	the variance of the memory encoding noise associated with location i and interval t
R	total available memory capacity (in bits)
$r(i, t)$	allocated capacity for location i and interval t
γ	an exponent on the posterior probability distribution that controls the dynamic allocation of memory
	Psychophysical Model
$\Phi^{-1}(\cdot)$	the inverse of the standard normal integral
(ε, θ)	a visual field location, specified in polar coordinates
$\alpha(\varepsilon, \theta)$	the target contrast threshold at (ε, θ)
α_0	the foveal contrast threshold
β	the steepness parameter for a Weibull function

Received September 19, 2016
Revision received November 20, 2017
Accepted December 23, 2017 ■