PLACING A LOWER BOUND ON TRANSSACCADIC MEMORYR CAPACITY USING VISUAL SEARCH

By

NICHOLAS KLEENE

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Melchi Michel

And approved by

_____

_____

_____

New Brunswick, New Jersey

May 2015

ABSTRACT OF THE THESIS

Placing a Lower Bound on Transsaccadic Memory Capacity Using Visual Search

By NICHOLAS KLEENE

Thesis Director:

Melchi Michel

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 transsaccadic memory (TSM), memory for information across saccades. TSM capacity may play a limiting role in tasks requiring efficient transsaccadic integration, such as multiple-fixation visual search tasks. Therefore, we sought to estimate TSM capacity and investigate its relationship to visual short-term memory (VSTM) using two visual search tasks, one where participants maintained fixation while saccades were simulated and another where participants made real saccades. We conducted an ideal observer analysis, which produced lower bound capacity estimates in both tasks that are in line with those previously found for VSTM. These estimates and the results of our ideal observer analysis indicate that TSM capacity may play a limiting role in multiple-fixation visual search tasks.

Acknowledgements

I would like to thank Dr. Melchi Michel, my committee members, Dr. Eileen Kowler and Dr. Manish Singh, and all the members of the Computational Vision Lab for their help with my thesis.

Table of Contents

List of Illustrations

Introduction

Saccades, rapid eye-movements that occur multiple times per second, are our primary means of gathering visual information. Each intervening fixation provides us with a snapshot of the visual world (Wurtz, 2008). However, due to our foveated visual system, high resolution information is only present in the central region of each of these samples. Despite this limitation, we are still able to combine information from previous fixations with the current one to generate a coherent percept of the world around us. For us to maintain such a representation there must be an integration process that combines information from previous fixations with the current one (Irwin, 1991). This integration process necessitates storing visual information from previous fixations in transsaccadic memory so that it may be retrieved later. Any limitations on transsaccadic memory capacity could hinder human performance in everyday tasks that require multiple fixations, such as driving, reading, or searching for a vehicle in the parking lot. Clearly, transsaccadic memory is a critical part of visual perception given how prevalent these types of tasks are in everyday life. Therefore, our goal in the current study was to place a lower bound on transsaccadic memory capacity in order to better understand how it limits performance in tasks that require multiple fixations.

Early hypotheses for transsaccadic memory postulated a visual integrative buffer that would store sensory information from previous fixations. During integration, this sensory information would be superimposed with the current fixation in spatial coordinates, resulting in a fused visual percept (Irwin, Yantis and Jonides, 1983). In this view, transsaccadic memory is similar to iconic memory in that it must be extremely high resolution and have a large capacity, at least when compared to visual short-term

memory. However, when participants were tasked with fusing two dot matrices presented on successive fixations, they proved unable to do so (Irwin et al., 1983). More recent theories posit that transsaccadic memory is mediated by the same system as visual short-term memory. In fact, additional work from Irwin found that transsaccadic memory operated on the same time course as visual short-term memory, has spatial resolution similar to visual-short-term memory, and that capacity estimates were similar to the commonly held "3-4 objects" for visual short-term memory (Irwin, 1991; Irwin, 1992). Although this is still an open question, current evidence supports the view that transsaccadic memory and visual short-term memory are mediated by the same system. As such, a secondary objective of this study was to evaluate this hypothesis by comparing capacity estimates obtained in our study with those obtained for visual short-term memory.

Although transsaccadic memory has rarely been examined explicitly, visual short-term memory is the topic of vast amount of research, particularly in regards to its capacity and the resolution of what is encoded (Luck, 2008). Much of the work on estimating visual short-term memory capacity has used change detection paradigms. In change detection tasks, participants view a stimulus array, then following an intervening blank or mask frame they view a new display and respond with whether one of the items changed or not (Luck, 2008). A major drawback in these approaches is that they do a poor job of representing the time course of saccades in natural scenes. It is extremely rare for natural scenes to exhibit significant changes over the span of a saccade, and when they do there are often transients that act as cues to the change. This makes change detection tasks ill-suited for our purposes. Since change detection tasks lack the

ecological validity to adequately examine transsaccadic memory, a better alternative is found in visual search tasks that require multiple fixations since these tasks require transsaccadic integration for efficient search performance. In contrast to change detection tasks, visual search tasks in the laboratory usually consist of static displays where participants are tasked with finding a target item (i.e. a specific letter). Critically, existing work on visual search has not been able to characterize how visual short-term memory might mediate transsaccadic memory (Eckstein, 2011). Instead, most models of visual search have focused on the influence of target-distractor similarity, display set size (Wolfe, 2006), and target position uncertainty (Michel and Geisler, 2011). Therefore, this represents a relatively novel paradigm for studying transsaccadic memory.

Another difficulty arises when deciding how to quantify transsaccadic memory capacity. A great deal of research has investigated this topic for visual short-term memory, resulting in two competing types of models for visual memory capacity. The first is a fixed-resolution, or "slots" model of visual short-term memory capacity. Slots models state that visual short-term memory can only hold a limited number of objects, with the general finding being that capacity is roughly 3-4 objects (Luck and Vogel, 1997). Objects that are held in memory are encoded with perfect resolution, but no information is stored for any additional objects. As such, slots models are an example of a high threshold model, one where items are either represented perfectly or not at all. The prevailing alternative to slots models are variable resolution hypotheses, otherwise known as "continuous resource" models. In continuous resource models, visual short-term memory capacity is defined by an arbitrary resource that can be allocated flexibly among all the objects in the display (Wilken and Ma, 2004). The drawback here is that as

the set size, the number of items in the display, increases so does the memory encoding noise associated with each item. This means that although each item can be represented in memory it will have imperfect resolution. This makes continuous resource models an example of a low threshold model, one where all items may be represented, albeit imperfectly. Apart from the number of items that are stored in memory, the other major difference between these two models is how much noise is present in each encoded representation. For slots models, each encoded item should have close to perfect resolution regardless of set size, while continuous resource models predict that encoding noise will increase as a function of set size. Unfortunately, this avenue of research has proven inconclusive with findings that support slots models (Zhang and Luck, 2008) and continuous resource models (Bays and Husain, 2008; Mazyar, Vandenberg and Ma, 2012). An alternative model of visual memory capacity may be required to resolve this deadlock.

A third issue comes into play when comparing capacity estimates for different stimuli. The main problem is in how an "item" is defined, whether an item is a single feature dimension or a set of integrated features. One study conducted by Xu (2002) found that participants performed better when asked to remember a set of integrated features (i.e. a colored, oriented line), as opposed to features on different parts of the same object (i.e. an oriented line attached to a colored square), seeming to indicate that an "item" should be a set of integrated features. However, Alvarez and Cavanagh (2004) found visual short-term memory capacity estimates to be extremely variable depending on the complexity of the stimuli used. In this study, capacity estimates ranged from 1.6 items for shaded cubes to 4.4 items for colored squares. This result would seem to

indicate that an “item” should be defined by a single feature dimension, contrasting the Xu (2002) result. Cleary it is uncertain what the correct definition of an “item” should be. Instead, what these studies really show is that that visual short-term memory capacity estimates are highly task dependent, and thus may not be viable for comparisons between different tasks or generalizations to the real world.

Given the limitations present in both fixed- and variable-resolution hypotheses, a new approach for quantifying memory capacity is needed. A solution can be found in rate-distortion theory. Rate-distortion theory 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 files that are more compressed use fewer bits, defined as binary units of information, to represent the original file. This relationship between the bit-rate and distortion of a system is defined by the rate-distortion function, which specifies the minimum number of bits 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. When considering visual short-term memory there is a clear link with rate-distortion theory, namely that visual short-term memory is concerned with efficiently storing visual signals so they may be retrieved and reconstructed at a later time (Sims, Jacobs and Knill, 2012). In this conception, the bit-rate of a visual memory system should be viewed as its capacity since this represents the amount of information that can be stored, while the distortion should be thought of as noise that is introduced into memory representations

due to its finite capacity. Using the optimal rate-distortion function we can then place a lower bound on transsaccadic memory capacity, provided that we are able to measure the noise that is present in the recovered signal. The optimal rate-distortion function for a Gaussian distributed information source and squared error distortion is presented in Figure 1. Additionally, rate-distortion theory allows us to quantify memory capacity in terms of bits, which has the beneficial feature of being a task independent unit of information, as opposed to "slots" or arbitrary resources.

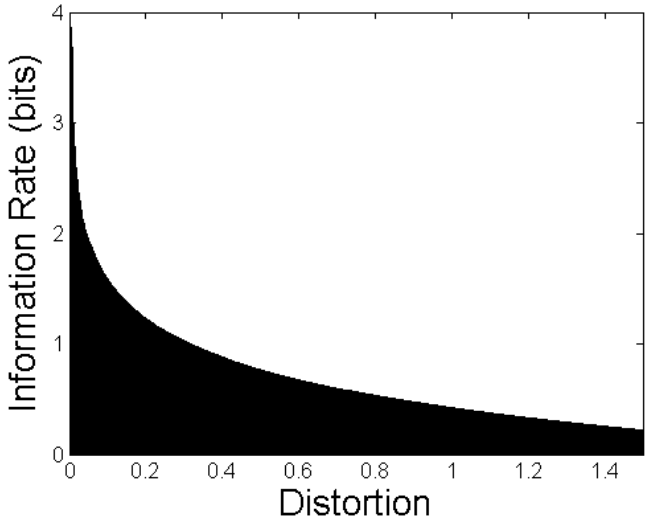

Figure 1. Rate-distortion function for a Gaussian distributed information source and squared error distortion function. The curve indicates the optimal rate-distortion bound, with the shaded region indicating physically realizable bit-rate and distortion pairs. The unshaded region is physically impossible.

This approach was introduced by Sims et al. (2012) to estimate visual short-term memory capacity using two change detection tasks. In one task participants had to remember line lengths, while in the other task they remembered orientations of arrows.

Two versions of each task were conducted, one in which the target feature was highly variable and one in which it was less variable. Using an ideal observer analysis, Sims et al. (2012) were able to estimate the distortion present in participants' memory encoded signals and subsequently place a lower bound on visual-short term memory capacity, finding estimates ranging from 4 to 8 bits across all conditions. Additionally, there were no significant differences found across tasks or variance conditions, providing evidence that this information theoretic approach works as a task independent way of quantifying memory capacity. Finally, participant data was well fit using a model that quantified visual short-term memory capacity in terms of bits when compared with slots and resource models, further validating this information theoretic approach to studying visual memory capacity.

In this paper, we will adapt the approach used by Sims et al. (2012) in order to estimate transsaccadic memory capacity for two different visual search tasks, one where we simulated saccades and one where participants made real saccades. In the simulated saccade task participants maintained fixation while we used a translating, blurred noise mask designed to mimic saccade induced transients, while in the real saccade task participants executed actual saccades. In both experiments, participants were presented with multiple samples of the stimulus, similar to how we receive snapshots of the visual world when performing visual search tasks in the real world. Given the framework introduced by Sims et al. (2012), we derived an ideal observer model to estimate transsaccadic memory capacity, as well as to determine whether transsaccadic memory is a limiting factor in visual search tasks. The ideal observer makes use of each participant's sensitivity to the target, as measured in a two-interval forced choice (2IFC) detection

task, in conjunction with rate-distortion theory in order to place a lower bound on transsaccadic memory capacity. We find that transsaccadic memory capacity is a limiting factor in multiple-fixation visual search tasks. Finally, we compare our capacity estimates with those found by Sims et al. (2012) to gain insight into whether transsaccadic memory and visual short-term memory truly are mediated by the same system.

Methods

In this section we describe two visual search tasks designed to evaluate the predictions of our ideal observer analysis. In both the Simulated Saccade and Real Saccade search tasks, the impact of memory was varied by providing observers with multiple samples of the target stimulus. The Real Saccade search task represents a straightforward test of transsaccadic memory capacity as observers were forced to integrate across a series of 5 degree saccades in order to localize a target signal in noise. For the Simulated Saccade search task observers were required to maintain fixation, therefore this experiment is a better controlled alternative that examines the role of memory capacity apart from saccades per se. The Simulated Saccade search task can also help control for the dual-task nature of the Real Saccade search task since observers need only localize the target, rather than executing a series of saccade in addition to target localization. Each visual search task was preceded and followed by a two-interval forced choice (2IFC) detection task. Data from the 2IFC detection task was used for selection of the target contrast in the corresponding search task, and as a sensitivity measure for use in our ideal observer model.

Participants

Data was collected from 5 participants in the Simulated Saccade experiment and from 4 participants in the Real Saccade experiment. All participants were naïve to the aims of this study, and all had normal or corrected-to-normal vision. Participants were paid $10 per hour.

Apparatus

Stimuli were presented on a Philips 202P4 CRT monitor at a resolution 1280x1024 pixels and a frame rate of 100 Hz. Participants were seated 70 cm from the display with their head position fixed in an adjustable headrest. Participants' right-eye was also tracked using an EyeLink 1000, sampling at 1000 Hz.

Stimuli

Target and Background

The target stimulus was a 2 cycle per degree Gabor, presented with 45° orientation counterclockwise from vertical. The target was embedded in a circular noise mask (the background), 20 degrees in diameter, which consisted of 1/f filtered noise with 10% RMS contrast. The area surrounding the background was set to the mean luminance of the display. The same target and background were used in the Real Saccade experiment, except the noise mask was 24 degrees in diameter, which allowed us to measure the sensitivity at each relevant eccentricity in the participants' visual field.

Simulated Saccade Transient

In order to simulate saccades we made use of a translating, blurred noise mask designed to replicate the saccade induced visual transients that are associated with a 10 degree saccade. This blurred noise mask was constructed by taking a sample of the noise background and convolving it with a horizontal boxcar filter, which simulated the horizontal blurring over a 10 ms. interval. When presented, each 10 ms. frame of the blurred noise mask was translated 70 pixels to the right or left from the previous frame to account for the retinal slip that would occur had the participant made a real saccade. The simulated saccade transient was used for both 2IFC detection tasks and in the Simulated Saccade search task. It was not used in the Real Saccade search task since there was no

need to simulate saccades. Figure 2 presents the target and background, as well as the simulated saccade transient for the Simulated Saccade experiment.

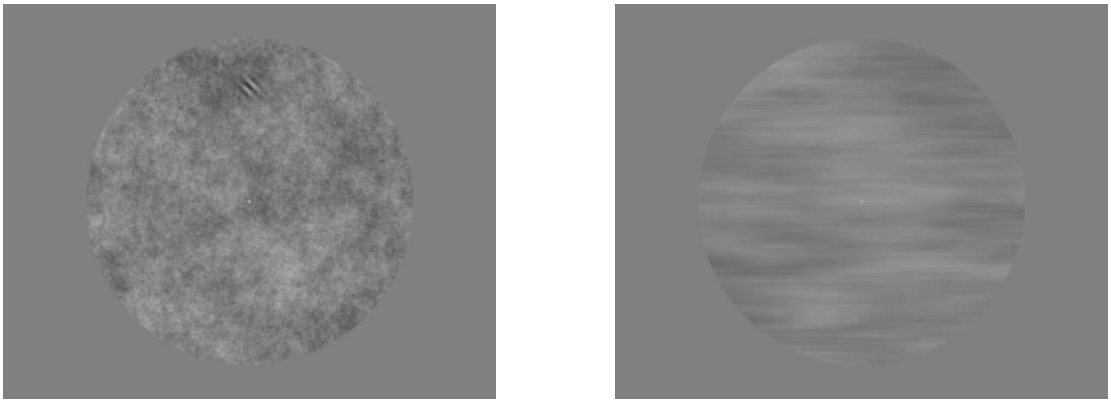

Figure 2. Left: The stimulus, with target present, for the Simulated Saccade experiment. Right: Blurred noise mask used to simulate saccade-induced visual transients in the Simulated Saccade search task.

Procedure

Each participant in the Simulated Saccade experiment completed approximately 9 one-hour long sessions, while each participant in the Real Saccade experiment completed 13 one-hour long sessions. Each session was conducted on a separate day. In both experiments, participants were first trained in a 2IFC detection task until their performance stabilized (~1-2 sessions). Participants' sensitivity to the target was measured using the 2IFC detection task (pre-test). Next, participants completed either the Simulated or Real Saccade search task. Due to the dual-task nature of the Real Saccade search task it was necessary to first train participants until the performance stabilized (~1 session). In order to control for potential changes in sensitivity over the course of the experiment, we re-measured participant sensitivity using the 2IFC detection task in both the Simulated and Real saccade experiments (post-test). Each session consisted of 10 blocks of 50 trials each, for a total of 500 trails per session.

2IFC Detection Tasks

Each trial in the 2IFC detection task began with the participant fixating a circular fixation marker at the center of the display. A circle cued the participant to one of the eight possible locations in which the target could appear. In the Simulated Saccade experiment, the cue always appeared 7 degrees from fixation, while in the Real Saccade experiment the cue could appear at the fovea, 2.5, 5 or 10 degrees from fixation, with trials in this experiment blocked by eccentricity. The target was always presented in the cued location. Participants initiated each trial with a button press. After a stimulus onset asynchrony (SOA) of 100-400 ms., participants maintained fixation at the center of the display and viewed two intervals of 1/f filtered noise, one with the target and one without. The order of the interval containing the target, as well as the location of the target/cue, was randomized on each trial. Each interval was displayed for 250 ms., followed by 50 ms. of the simulated saccade transient, and separated by a 500 ms. blank frame of mean luminance. The direction of the simulated saccade transient was randomly determined for each interval. After viewing both intervals, the participant indicated which interval the target was in with a button press. Participants then received auditory feedback. Finally, if at any point during a trial the participant broke fixation they were notified and the trial was aborted. A new trial was then generated and the experiment proceeded as before.

Prior to the start of each block, participants completed a five point calibration routine to ensure precise tracking of eye position. This was followed by five practice trials where the target contrast was fixed at 50% RMS 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). In the Simulated Saccade experiment a separate adaptive procedure was used for each target location, while in the Real Saccade experiment a separate adaptive procedure was used for each eccentricity. For most of the detection sessions completed as posttest measures (after completing the search task) the target contrast was fixed at the RMS contrast used in the search task. The one exception was the final detection session in the Real Saccade experiment, where target contrast was once again selected using an adaptive procedure.

Simulated Saccade Search Task

Similar to the detection task, each trial in the Simulated Saccade task began with the participant fixating a circular fixation marker at the center of the screen. Eight evenly spaced circular cues located 7 degrees from fixation indicated the possible locations of the target. Subjects initiated each trial with a button press, and following an SOA of 100-400 ms. they viewed one, two, or four intervals of the target and background (to be referred to as fixation intervals), each presented for 250 ms. and followed by 50 ms. of the simulated saccade transient. In addition to the three different fixation interval conditions, there were two target permanence conditions, a Redundancy condition and an Uncertainty condition. In Redundancy trials the target appeared in the same location in each fixation interval, but in Uncertainty trials the target appeared in only one fixation interval. Therefore, as the number of fixation intervals increases, the memory load for Redundancy trials should increase since observers must integrate across fixation intervals to achieve optimal performance. In contrast, the memory load should not change as a function of the number of fixation intervals for Uncertainty trials since the target is only presented once. At the end of each trial, participants were once again presented with the

eight circular cues. They then directed their gaze towards their indicated target location and finalized their response with a button press, receiving auditory feedback.

Trials were blocked by the target permanence condition (Redundancy or Uncertainty) and the three fixation interval conditions (one, two or four). Since a one fixation interval Redundancy condition is functionally equivalent to a one fixation interval Uncertainty condition, this produced a total of 5 conditions. Participants were always informed of the condition prior to beginning a block. At the beginning of each block participants completed five practice trials in which the target contrast was fixed at 50% RMS contrast. Data from these practice trials were not recorded. For the rest of the block, target contrast was selected by finding the RMS contrast necessary to produce 1.5 d' performance at each of the 8 locations, determined using data collected in the detection task. Finally, if at any point during a trial the participants broke fixation they were notified and the trial was aborted. A new trial was then generated and the experiment proceeded as before.

Real Saccade Search Experiment

Participants began each trial fixating at a circular fixation point 2.5 degrees to the left or right of the display center. Eight circular cues evenly spaced 7.5 degrees from the display center cued participants to the potential target locations. When ready, participants initiated the trial with a button press. They then viewed one, two, or four fixation intervals, each followed by a 5 degree saccade. Each fixation interval was presented for 100 ms., at which point participants were cued to saccade 5 degrees to the left or right, depending on the displacement of the initial fixation mark. In order to cue participants to make a saccade, the current fixation mark was extinguished, a new fixation mark

appeared, and a tone was played. Once cued, participants were required to wait 50 ms. before making a saccade in an attempt to prevent anticipatory saccades. Participants were required to initiate the saccade no later than 250 ms. after cue onset. Therefore, participants were required to make a saccade between 150 and 400 ms. after stimulus presentation. The threshold for detecting a saccade was 200 degrees per second. If the participants did not initiate their saccade within the specified time window or their saccade was too slow they were notified, and the trial was aborted. Once a saccade was detected, the display was blanked, and the next fixation interval was presented upon landing.

The Real Saccade search task had the same 5 conditions as the Simulated Saccade Search task, except in this case each fixation interval was followed by a 5 degree saccade rather than a simulated saccade transient. Once again, trials were blocked by the target permanence condition (Redundancy or Uncertainty) and the fixation interval condition (one, two, or four). At the beginning of each block participants completed five practice trials in which the target contrast was fixed at 50% RMS contrast. Data from these practice trials was not recorded. For the rest of the block target contrast was selected by finding the RMS contrast necessary to produce 1.25 d' performance at 5 degrees from fixation, determined using data collected in the detection task. Finally, if the participants ever broke fixation they were notified, and the trial was aborted, and a new one was generated. The only time during a trial where participants were allowed to break fixation was during the 250 ms. window of saccade initiation.

Transsaccadic Memory-limited Ideal Observer

The first step in developing our ideal observer model for the visual search tasks was to determine how human performance should be limited by a variety of factors known to have an influence on tasks such as these so we may incorporate them into our model. For the visual search tasks described above one limiting factor is the participants' sensitivity to the target. Humans are known to have reduced visual acuity in the periphery, and these tasks require detection of a target signal in the periphery. Theoretically, transsaccadic memory capacity should also be a limiting factor since it can constrain observer's ability to integrate across fixation intervals. Additionally, since the target could appear in any of 8 locations participants will also have spatial uncertainty about which location it would appear in. Finally, the Uncertainty condition adds temporal uncertainty since in this condition the target will appear in one fixation interval but not the others. A schematic for the ideal observer model is presented below:

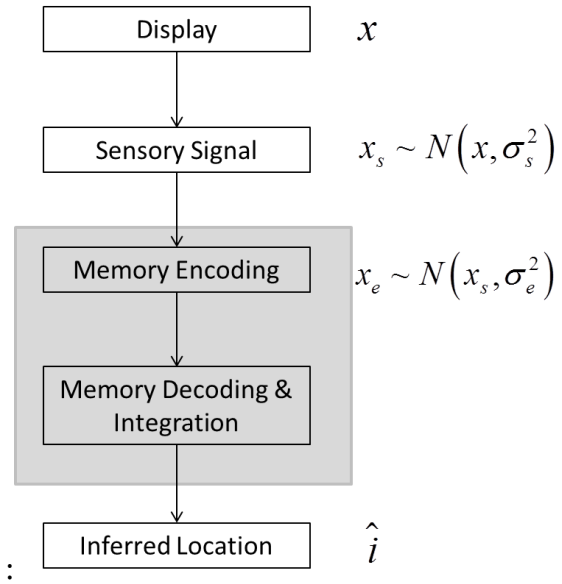

Figure 3. Schematic of the transsaccadic memory limited ideal observer model. x, x_s and x_e represent the mean sensory response, the actual sensory response generated by the

observer, and the noise-corrupted memory encoded signal at each location and fixation interval, respectively. $\hat{i}$ indicates the inferred location of the target.

For simplicity, let us first consider a single fixation interval. The first step in this schematic is for the display to be created, with a mean sensory response, x, generated at each of the eight potential target locations. For mathematical convenience and without loss of generality we define target presence with a mean response of 1 and target absence with a mean response of 0. These values represent the expected sensory response at locations with the target and without, respectively. The importance here is not on the actual values themselves, but rather to maintain a difference of 1 between the means of the signal distribution (target present location) and the noise distribution (target absent locations). This allows us to represent any changes in signal-to-noise ratio (SNR) using only the variance of the signal and noise distributions. The sensory response, x_s, is then generated by drawing from a normal distribution centered on the corresponding x, with variance, σ_s^2, defined by:

$$\sigma_s^2 = \frac{1}{d'^2} \quad (1)$$

Here, d' corresponds to the psychophysical SNR at each location, determined using the data from our 2IFC detection task, and defined as:

$$d' = \frac{\mu_s - \mu_n}{\sqrt{\frac{1}{2}\left(\sigma_s^2 + \sigma_n^2\right)}} \quad (2)$$

Where the subscripts *s* and *n* correspond to signal and noise distributions, respectively, while μ and σ^2 are the mean and variance of the corresponding distributions. Equation 2

makes it clear that any changes in observer SNR can be purely represented in terms of the variance. Since d' is a measure of the observer's sensitivity to the target Equation 1 defines the observers uncertainty associated with the sensory response at a particular location. Greater values of d' indicate greater sensitivity, and therefore less uncertainty, resulting in reduced sensory noise.

The sensory signal is then encoded in memory, generating a memory encoded signal, x_e, at each location with memory encoding noise, which we model as equivalent noise (Lu & Dosher, 1999). This added noise represents the minimum achievable distortion for a given bit rate as defined by the rate-distortion function. Therefore, the memory encoding noise, σ_e^2, in our ideal observer analysis is chosen so that it will achieve this theoretical bound for a given bit rate, with distortion defined as the squared difference between the memory encoded signal and sensory signal. This means that information rate and distortion for our ideal observers will always reside on the curve defining the rate-distortion bound in Figure 1. The memory encoded response is generated by drawing from a normal distribution centered on the sensory response, with variance obtained from:

$$\sigma_e^2 = \frac{\sigma_d^2 + \sigma_s^2}{e^{2r} - 1} \quad (3)$$

In Equation 3, σ_d^2 corresponds to the display variance, while r defines the total memory capacity allocated to a particular location and fixation interval. If memory capacity is high then the memory encoded signal should closely resemble the sensory signal as memory encoding noise will be minimum. Alternatively, if memory capacity is low this will increase the variability of the memory encoded signals, making them less reliable

indicators of the sensory response. The display variance and the total memory capacity allocated to a particular location and fixation interval are given by:

$$\sigma_d^2 = p(1-p) \quad (4)$$

$$r = \frac{R}{n(k-1)} \quad (5)$$

In addition to sensory noise, the amount of noise in each memory encoded response is in part related to the predictability of the target location. Since the target is equally probable to appear in any of the eight locations, the probability that a particular location will have a mean response of 1 (target present) is $\frac{1}{8}$. Therefore, for a particular location the target will be present roughly one out of every eight times. This means that the variance of the target location must be equal to that of a Bernoulli distribution with $p = \frac{1}{8}$.

In Equation 5, the R term corresponds to the total transsaccadic memory capacity of the ideal observer. This is the one free parameter in the model that will be fit to participant data, giving us a lower bound estimate for transsaccadic memory capacity. We make the simplifying assumption that the ideal observer divides memory capacity evenly among n potential target locations and $k-1$ fixation intervals. The $k-1$ term reflects the reasonable assumption that transsaccadic memory will not be required for the final fixation interval as observers should be able to simply integrate the sensory responses of the final fixation interval with the memory encoded responses from the previous intervals It is possible that observers will actually employ more sophisticated encoding strategies, such as one where transsaccadic memory capacity allocation is adjusted from one fixation interval to the next. An encoding strategy such as this could indicate that the capacity

estimates obtained here are too high. This is because observers could ignore locations that are unlikely to contain the target, and instead allocate all of their capacity to locations with a high likelihood of containing the target.

In the final stage of our ideal observer model, sensory and memory encoded likelihoods are integrated and the ideal observer makes a decision about which location the target appeared in. The decision rule for Redundancy trials is then given by:

$$\hat{i} = \arg\max\left(\exp\left(\sum_{t=1}^{k} \frac{x_{e_{it}}}{\sigma_{total}^2} \right), i \right) \quad (6)$$

With:

$$\sigma_{total}^2 = \sigma_{s_{it}}^2 + \sigma_{e_{it}}^2 \quad (7)$$

This is essentially the decision rule derived in Najemnik & Geisler (2005) for independent noise samples, with the added component of memory encoding variance. In Najemnik & Geisler (2005), observers were performing a visual search task with a similar display to the experiments described here, with the exception that their gaze was unconstrained. For their task, integration was performed by maintaining a sum of weighted sensory responses at each potential target location across all fixations, with each response weighted by the inverse variance, which is known to correspond to precision. The posterior probability distribution can then be computed from this weighted sum, and the observer can find the maximum log posterior by selecting the location with the maximum posterior probability. We extend this decision rule to our task, where transsaccadic memory should be a limiting factor in performance. Therefore, the weights, σ_{total}^2, in Equation 6 are now given by the inverse of the sum of sensory and memory encoding variances, as shown in Equation 7. It is also important to note that the sensory

responses are used for the final fixation interval instead of the memory encoded responses since $\sigma_e^2 = 0$ for this interval. This is because observers should presumably be able to integrate the sensory likelihoods generated on the final fixation interval with those previously encoded in memory. Since the sensory signal need not be encoded in memory there should be no additional memory encoding noise.

For Uncertainty trials, the decision rule is given by:

$$\hat{i} = \arg\max \left(\sum_{t=1}^{k} \exp\left(\frac{x_{s_{it}}}{\sigma_{s_{it}}^2} \right) \right) \quad (8)$$

Which is derived from the case for 1 of *N* unknown signals presented in Peterson, Birdsall & Fox (1954). Here, observers must integrate sensory likelihoods at each location across fixation intervals. The observer then selects the location with the greatest posterior probability. The result of this computation is that the posterior distribution will be dominated by the location in the fixation interval with the greatest response. Therefore, imperfect integration will have a minimal influence on observer performance. As such, we assume that observers have unlimited memory for integration in this condition, with $\sigma_e^2 = 0$. Under the conditions of our experiment, this decision rule produces behavior indistinguishable from that of an observer that uses the max rule (Pelli, 1985; Notle & Jaarsma, 1967), which would just select the target location in the fixation interval with the greatest response. An observer using the max rule would require minimal memory capacity, only enough to store the location and value of the maximum response for comparison with subsequent fixation intervals. In this case observers wouldn't actually perform integration, meaning transsaccadic integration is not expected

to noticeably improve performance in this condition. The predictions of our ideal observer, with a d' of 1.5, can be seen in Figure 4.

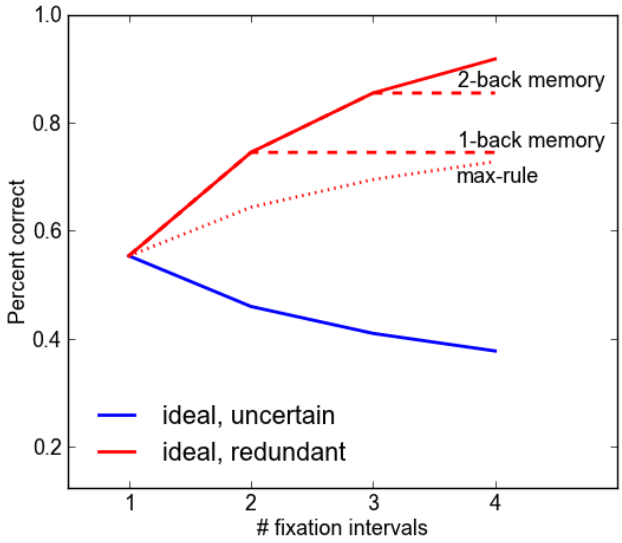

Figure 4. Predictions of the transsaccadic memory-limited ideal observer model with a d' of 1.5 for Redundancy trials (red) and Uncertainty trials (blue). The dotted line shows predictions for a max-rule observer with no integration. Additionally, dashed lines show predictions if transsaccadic memory is 1- or 2-back.

As can be seen from in Figure 4, the ideal observer performs better as the number of fixation intervals increase in Redundancy trials since the observer gets more samples of the target and therefore more evidence about its true location. It is important to note that the transsaccadic memory limited ideal observer outperforms a model that selects its response based simply on a max rule, making integration an essential part of optimal performance in this condition. Additionally, predictions for an observer with memory for only one or two fixation intervals are also shown. In these cases, observers would only be using one or two previous samples since they are unable to store anything beyond that.

For Uncertainty trials the ideal observer performs worse as the number of fixation intervals increases because each additional fixation interval adds temporal uncertainty.

Results

Threshold Fits and Visibility Maps

One critical aspect of our ideal observer model is the generation of sensory signals. In the Simulated Saccade experiment we used maximum likelihood to estimate participant sensitivity at each potential target location. We began by fitting a single slope across all target locations using maximum likelihood, assuming that it stays relatively flat for a low spatial frequency target, a result found in previous work (Ackermann & Landy, 2013). Additionally, the number of trials required to obtain separate, reliable slope estimates for each potential target location (~200 trials per location) would prove intractable. We then used maximum likelihood to fit a threshold at *each* target location, conditioned on the previously estimated slooe. Finally, we computed the target contrast required to produce a 1.5 d' performance at each potential target location for use in the Simulated Saccade search task. This meant that a different target contrast was used at each location. For our ideal observer analysis of the Simulated Saccade search task we then converted the proportion correct scores at each target location from our post-test to obtain d' measurements at the particular contrasts tested. This conversion was carried out using the well-known formula:

$$d' = \sqrt{2}\Phi^{-1}(p) \quad (9)$$

Where p is the proportion correct and Φ^{-1} is the inverse cumulative normal distribution.

For the Real Saccade experiment it was necessary to fit full visibility maps for each participant. We began by fitting a single slope across all target locations using maximum likelihood, assuming that the slope would vary minimally across all locations. This was done by pooling all of the detection data collected in the pre-test, regardless of

angular location and eccentricity. Next, we again used maximum likelihood to estimate each participant's threshold at the fovea, keeping the slope fixed. For all other possible target locations we modeled threshold using the following:

$$\alpha_{\theta\rho} = \alpha_0 e^{\tau_\theta} \quad (10)$$

Which has been widely used as a model for the falloff in visual acuity as a function of eccentricity (Peli, Yang & Goldstein, 1991; Geisler, Perry, & Najemnik, 2006; Michel & Geisler, 2011; Ackermann & Landy, 2013). Here α corresponds to the threshold and τ_θ is a free parameter. Visibility maps are represented in polar coordinates, so θ denotes the angular coordinate, while ρ denotes the radial coordinate, therefore α_0 corresponds to the participants' threshold at the fovea. To fit each τ_θ we used maximum likelihood to determine the most probable τ_θ parameter for each angular direction given our data at 2.5, 5 and 10 degrees and keeping the slope fixed. Since we tested in 8 angular directions this provided us with a total of 8 τ_θ parameters. With our τ_θ parameters in hand we computed visibility maps by interpolating between the angular and radial coordinates for which we had data. The resulting visibility map for one representative subject can be seen below in Figure 5. Visibility maps for all five participants in the Real Saccade experiment are presented in the appendix.

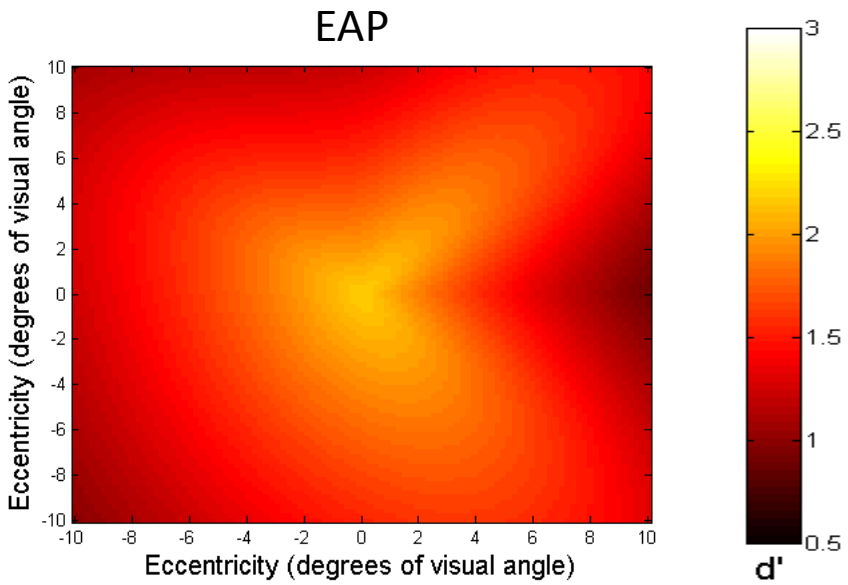

Figure 5. Visibility map for participant EAP in the real saccade experiment. Brighter colors indicate greater sensitivity. Note that as absolute eccentricity increases the fall off in sensitivity is relatively minimal.

As can be seen in the visibility map for participant EAP, the reductions in sensitivity are relatively minimal as eccentricity increases. Most participants' visibility maps fit this pattern. This was by design, as we chose a low spatial frequency target specifically because it would be almost as detectable at 5 degrees as it would be at 10 degrees.

Human Data and Model Predictions

Simulated Saccade Experiment

For each participant in the Simulated Saccade experiment we calculated the proportion correct for each of the five conditions. Ideal observer predictions were generated using Monte Carlo sampling with 10,000 trials per simulation. For each human observer we constructed different ideal observers by using their sensitivity measurements

obtained in our 2IFC detection posttest. Additionally, for each participant we simulated ideal observers with unlimited capacity and capacity limits of 4, 8 and 16 bits. Observers with unlimited capacity were simulated to evaluate the possibility that transsaccadic memory capacity was not a limiting factor for this task. The results for a representative participant are displayed below:

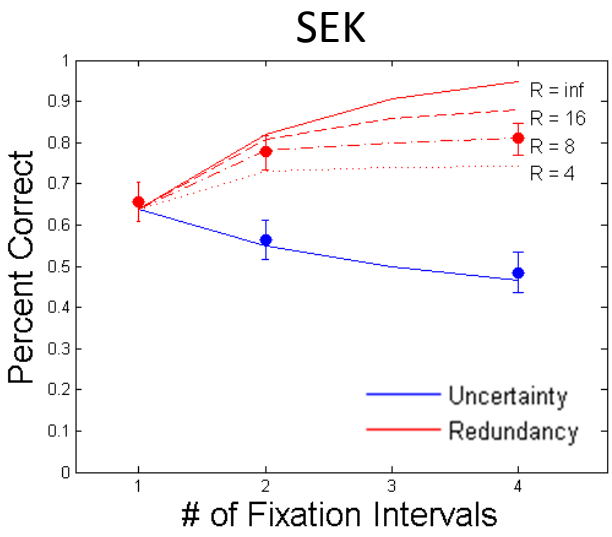

Figure 6. Model predictions and human data for participant SEK in the Simulated Saccade search task. Data points represent human data, with 95% CI's. Curves represent model predictions in Uncertainty (blue) and Redundancy (red) trials. Different red curves indicate ideal observers with different transsaccadic memory capacity limits, which are displayed to the right of the curve.

To determine each individual's transsaccadic memory capacity we next used maximum likelihood to estimate the minimum number of bits that the ideal observer would require to match the participants' performance in each condition. For the 2-fixation interval Redundancy condition we found estimates ranging from 2.1 to 9.9 bits,

with a median estimate of 5.5 bits, while for the 4-fixation interval Redundancy condition we found capacity estimates ranging from 3.9 to 8.0 bits, with a median estimate of 6.2 bits.

Real Saccade Experiment

Ideal observers for the Real Saccade experiment were implemented in a similar manner to the Simulated Saccade experiment. The one exception was that we now used each participant's actual eye-position in conjunction with the interpolated visibility map to determine the ideal observer's sensitivity at each target location and fixation interval. All other analyses were conducted exactly the same as in the Simulated Saccade experiment. The results for a representative participant are displayed below:

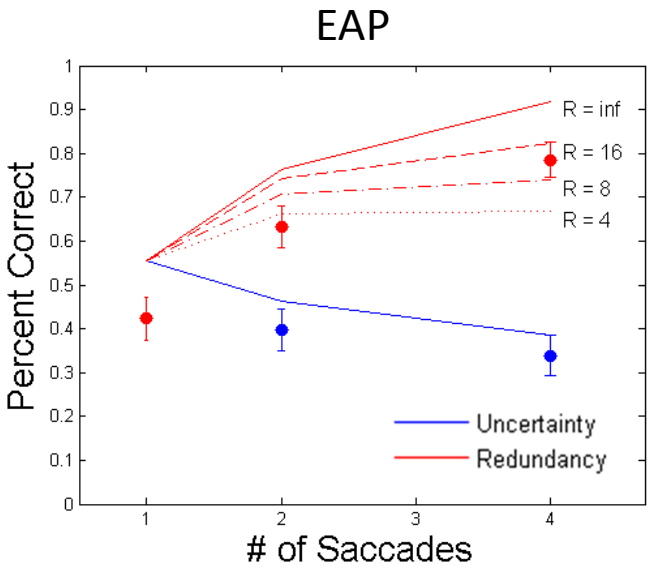

Figure 7. Model predictions and human data for participant EAP in the Real Saccade search task. Data points represent human data, with 95% CI's. Curves represent model predictions in Uncertainty (blue) and Redundancy (red) trials. Different red curves indicate ideal observers with different transsaccadic memory capacity limits, which are displayed to the right of the curve.

It should be noted that one participant was outperformed by all memory-limited ideal observers, making it impossible to generate capacity estimates. Another subject outperformed a memory-unlimited observer for the 2-fixation interval Redundancy condition, indicating that transsaccadic memory capacity was not a limiting factor in this condition. For the rest of the participants, we once again used maximum likelihood to determine the minimum number of bits that the ideal observer required to approximate each individual's performance. For the 2-fixation interval Redundancy condition we found estimates ranging from 3.3 to 4.0 bits, with a median estimate of 3.7 bits, while for the 4-fixation interval Redundancy condition we found capacity estimates ranging from 7.5 to 34.0 bits, with a median estimate of 9.5 bits.

Discussion

The results for the Simulated Saccade task indicate that transsaccadic memory capacity most likely plays a limiting role in multiple-fixation search tasks. This can be seen both from the decrements in performance seen in ideal observers with lower capacity limits, as well as the apparent capacity estimates measured for each participant. It is important to note that capacity estimates for the 4-fixation interval Redundancy trials represent tight lower bounds on transsaccadic memory capacity. It is possible that each participant truly has greater transsaccadic memory capacity than measured here, but given their performance they can't possibly have less. The fact that most capacity estimates also ranged from 4 to 8 bits indicates that there may be a degree of individual differences in terms of transsaccadic memory capacity. However, it is possible that each participant actually has a similar number of bits since our estimates simply represent lower bounds on capacity. Additionally, the capacity estimates obtained here are broadly consistent with those reported by Sims et al. (2012). Sims et al. (2012) found capacity estimates ranging from about 4 to 6 bits for orientation and 4 to 8 bits for line length. Since these estimates are for visual short-term memory capacity, the fact that our estimates are consistent provides support for the hypothesis that visual short-term memory and transsaccadic memory are actually mediated by the same system. Finally, bounds increased as a function of the number of fixation intervals for all participants except one. This provides evidence that transsaccadic memory is not just 1- or 2-back since participants must have been using information from each fixation interval to achieve their measured performance.

Results obtained from the Real Saccade task are somewhat less clear. Once again, we observed decrements in performance for ideal observers with lower transsaccadic memory capacity limits, as well as apparent capacity estimates for participants, again supporting the view that transsaccadic memory capacity is a limiting factor in multiple-fixation visual search tasks. Our range of capacity estimates was somewhat larger in this experiment, with estimates ranging from 7.5 to 34.0 bits. However, this is likely due to some limitations, which will be discussed shortly. We again observed that transsaccadic memory capacity estimates increased as a function of the number of fixation intervals, once again providing evidence that transsaccadic memory isn't just 1- or 2-back.

Analysis of two participants' data proved difficult as one underperformed all ideal observer model predictions (XH), while the other outperformed most model predictions (SKM). In the case of XH, underperformance is most likely due to difficulty completing both goals of a trial simultaneously; that is the ability to make alternating saccades while also monitoring for target appearance. This can be seen when looking at the proportion of trials that were aborted during the search task. After performance stabilized, XH always had at least 40% of the trials aborted in each session. In addition to creating uncertainty about which trials would actually be completed, running so many extra trials could have induced fatigue. By contrast, no other participant had over 36% of the trials aborted after their performance stabilized, with the number of aborted trials falling to 20% or lower by the end of the experiment. The end result is that XH was forced to initiate 100 extra trials per session, on average, when compared to other participants.

On the other hand, SKM's over performance is likely due to one of two possibilities. One limitation in the Real Saccade experiment is the blank frame that was

presented following saccade initiation and continuing until saccade termination. The inclusion of this blank frame could have allowed participants to use transients to aid them in locating the target, artificially inflating performance. To eliminate this possibility another version of the experiment must be conducted, this time without a blank frame included. This experiment would be conducted in the same manner as what is described here, however when a saccade is detected we would immediately present the next fixation interval instead of a blank frame. The other possible explanation is that we underestimated SKM's threshold at the potential target locations. Future work will explore different models for computing visibility maps in the Real Saccade task to determine whether this explanation is valid.

Given the results of our Simulated and Real Saccade experiments, it can be concluded that transsaccadic memory capacity does play a limiting role in multiple-fixation visual search tasks. In addition, capacity estimates are in line with those reported for visual short-term memory by Sims et al. (2012), providing evidence that transsaccadic memory and visual short-term memory are mediated by the same system. In fact, it is extremely likely that transsaccadic memory and visual short-term memory are truly the same thing. Further investigation will be required to verify that this claim holds true in the Real Saccade experiment.

It is important to also discuss the assumptions of our ideal observer analysis. The ideal observer described in this paper will give the best possible performance, given that the only limitations on performance are noise, memory capacity, and that memory capacity is actually used optimally. Our model also assumes perfect integration, which is likely not true for humans. If observers were to integrate imperfectly this would mean

that the bounds generated in this paper are too loose to be useful. In order to minimize the effect of imperfect integration noise samples presented in each fixation interval were independent, meaning integration would provide observers with maximal integration. Additionally, given the bounds obtained in the Simulated and Real Saccade experiments (for all but XH) we can conclude that this assumption did not prove problematic.

Finally, it should be noted that what is presented here is a somewhat novel approach to examining visual memory, both in the actual tasks participants completed and the way in which memory capacity is quantified. Future research projects should strongly consider using visual search tasks instead of change detection tasks to study visual memory capacity since search tasks more accurately represent the temporal dynamics of the real world. A shift towards quantifying memory capacity in terms of bits would also greatly aid the study of visual short-term memory. Since bits are a task independent unit of information, using this approach can remove much of the variability in capacity estimates found for different stimuli, allowing for better comparisons across experimental paradigms. These are important considerations that should be taken into account for any project designed to estimate visual memory capacity.

Appendices

A1. 2IFC Detection task

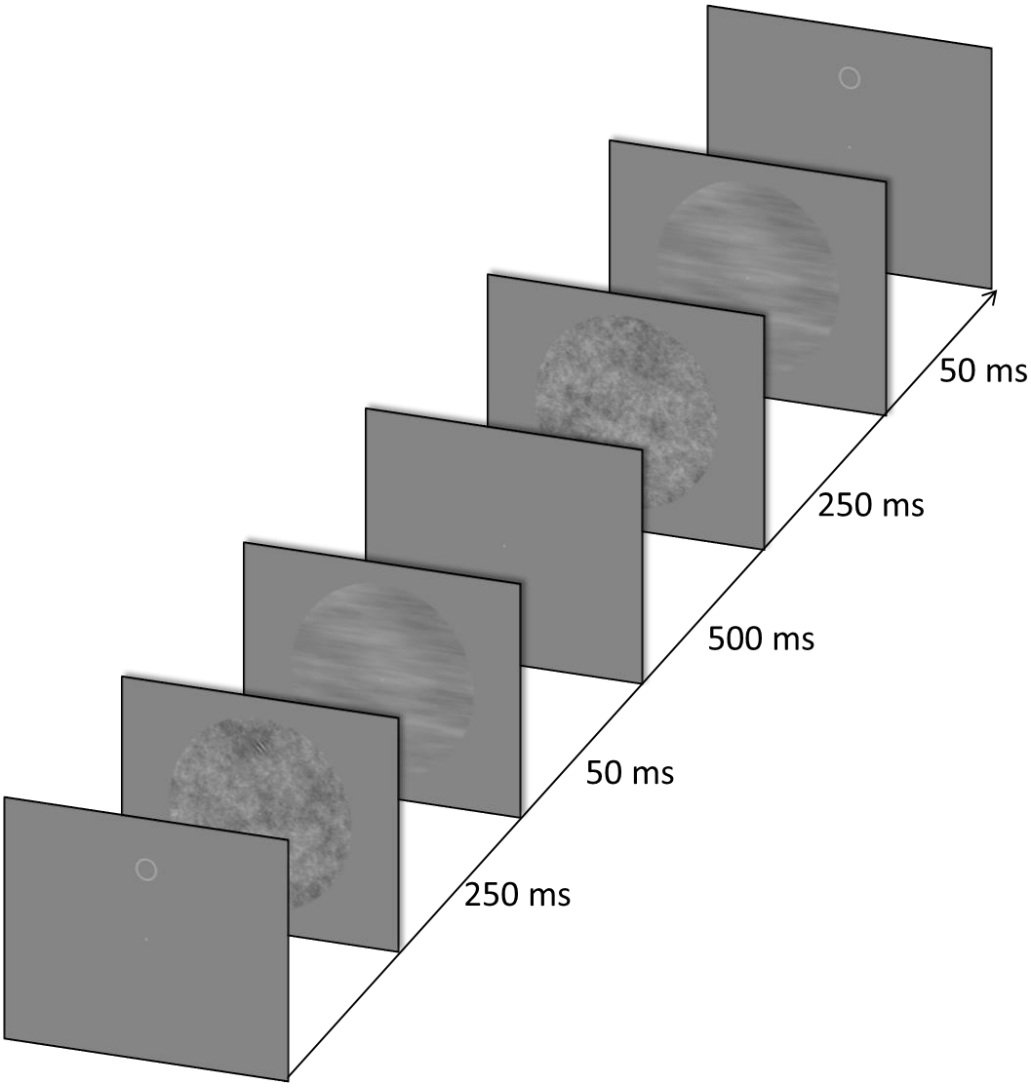

Figure A1. Stimulus sequence for the 2IFC detection tasks. Here, the target is present in the first interval.

A2. Simulated saccade search task

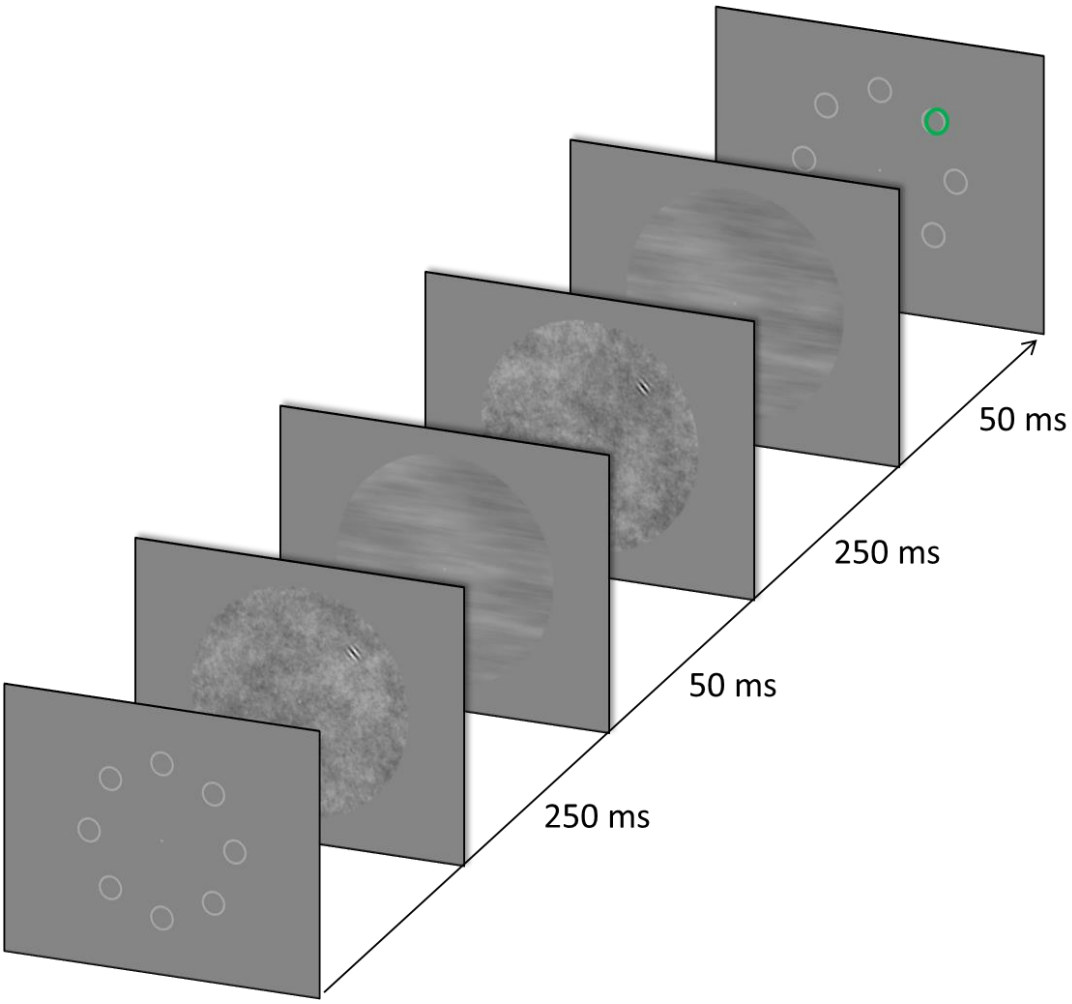

Figure A2. Stimulus sequence for a two-fixation interval, redundancy trial in the simulated saccade search task. In uncertainty trials, the target would only be present in the first or second fixation interval, rather than in each. Green circle in the final frame indicates response location.

A3. Visibility maps

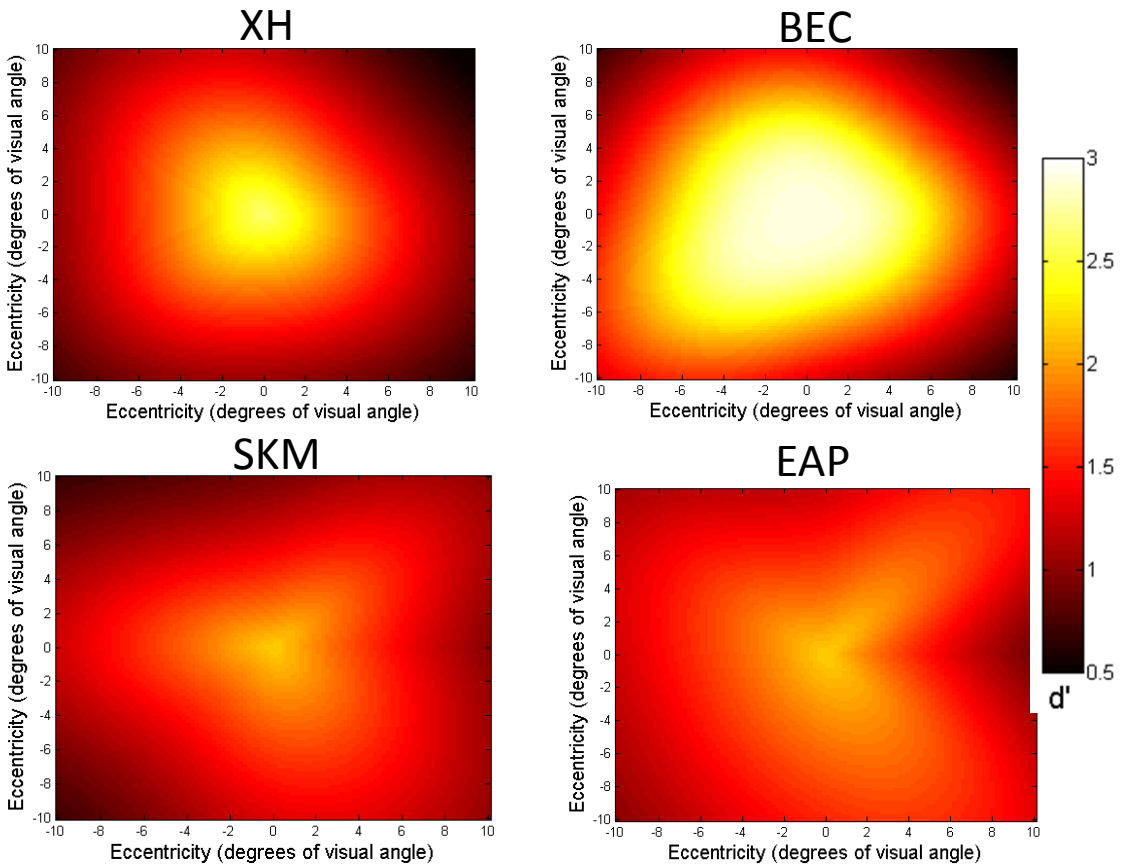

Figure A3. Visibility maps for each participant in the real saccade task. Bright regions indicate areas with greater sensitivity. Note how the fall off in sensitivity is minimal with increasing eccentricity.

A4. Simulated saccade results & model fits

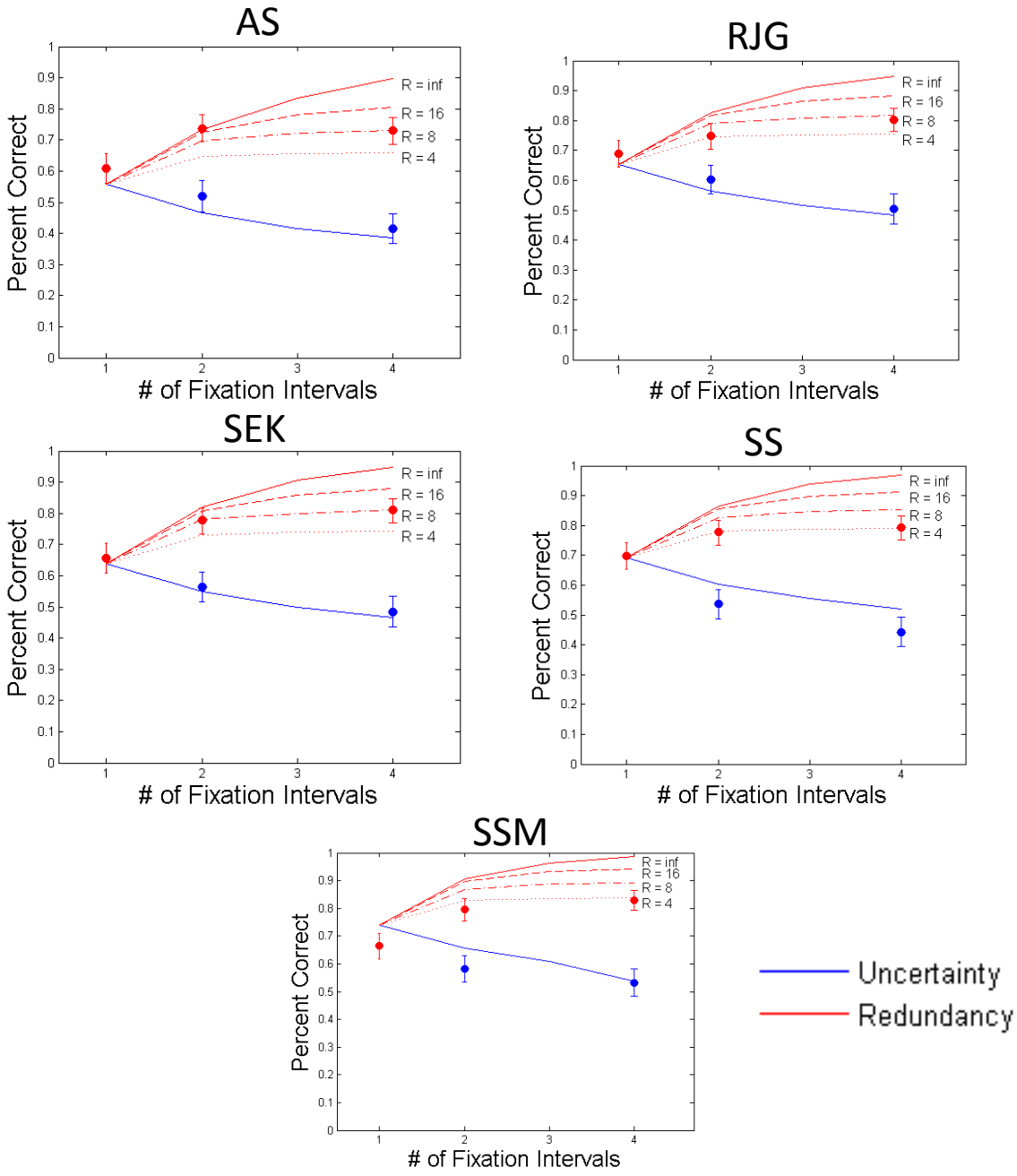

Figure A4. Model predictions and human data for each participant in the simulated saccade search task. Data points represent human data, with 95% CI's. Curves represent model predictions in uncertainty (blue) and redundancy (red) trials. Different red curves indicate ideal observers with different transsaccadic memory capacity limits, which are displayed to the right of the curve.

A5. Real saccade results & model fits

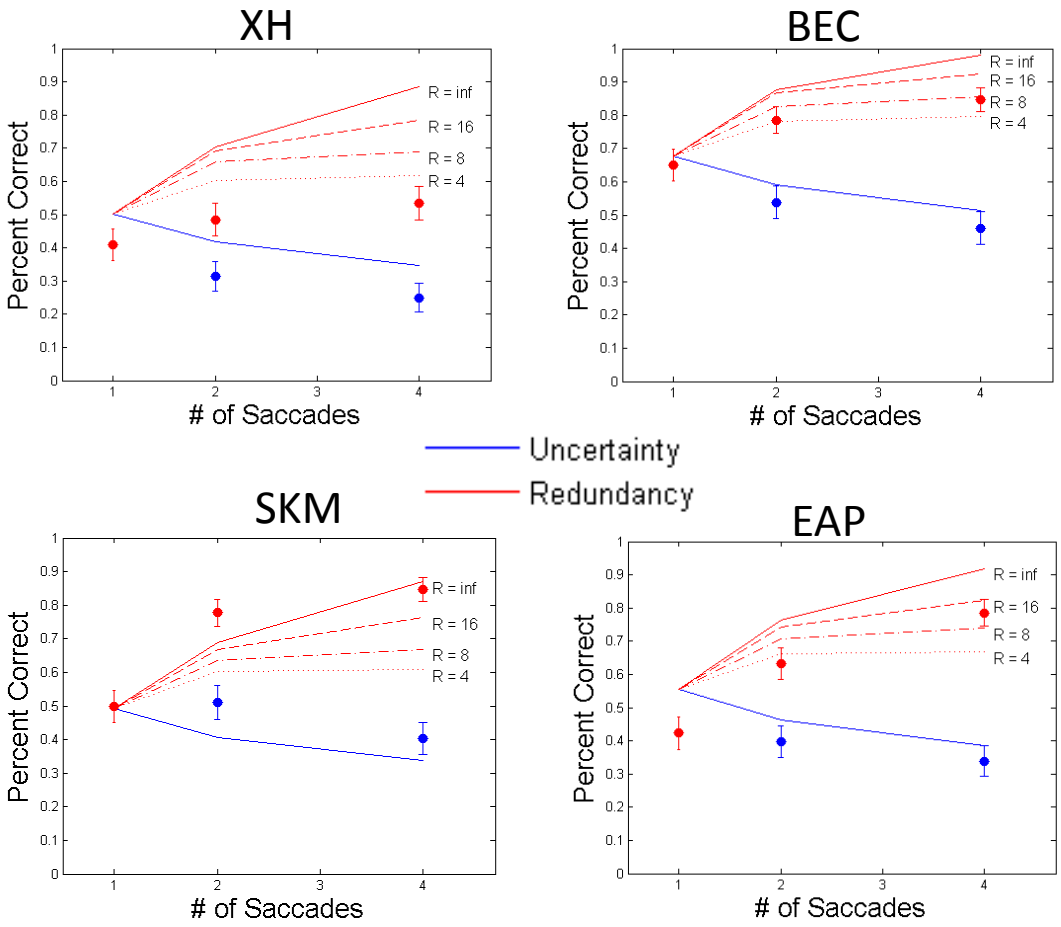

Figure A5. Model predictions and human data for each participant in the real saccade search task. Data points represent human data, with 95% CI's. Curves represent model predictions in uncertainty (blue) and redundancy (red) trials. Different red curves indicate ideal observers with different transsaccadic memory capacity limits, which are displayed to the right of the curve.

References

Ackermann, J.F., Landy, M.S. (2013). Choice of saccade endpoint under risk. *Journal of Vision, 13*, 1-20.

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*(2), 106-111.

Bays, P., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science, 321*, 851-854.

Eckstein, M., (2011). Visual Search: A retrospective. *Journal of Vision, 11*(5), 1-36.

Geisler, W.S., Perry, J.S., Najemnik, J. Visual search: the role of peripheral information measured using gaze-contingent displays. *Journal of Vision, 6*, 858-873.

Hollingworth, A., Richard, A., & Luck, S. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General, 137,* 163-181.

Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420-456.

Irwin, D., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34*, 49-57.

Irwin, D. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 307-317.

Kontsevich, L. L., and Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*, 2729-2737.

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

Luck, S. (2008). Visual Short-term Memory. In *Visual Memory* (pp. 43-85). New York: Oxford University Press.

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

Mazyar, H., Berg, R., & Ma, W. (2012). Does precision decrease with set size? *Journal of Vision, 12*, 1-16.

Michel, M.M., & Geisler, W.S., (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision, 11*(1) 1-18.

Najemnik, J. & Geisler, W.S., (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387-391.

Nolte, L.W. & Jaarsma, D., (1967). More on the detection of one of *M* orthogonal signals. *Journal of the Acoustical Society of America, 41*(2), 497-505.

Peli, E., Yang, J. & Goldstein, (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America A, 8*, 1762-1774.

Pelli, D.G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*(9), 1508-1532.

Peterson, W.W., Birdsall, T.G. & Fox, W. The theory of signal detection. *Information Theory, Transactions of the IRE Professional Group, 4*, 171-212.

Sims, C., Jacobs, R., & Knill, D. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119*(4), 807-830.

Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision, 4*, 1120-1135.

Wolfe, J.M., (2006). Guided Search 4.0: Current Progress with a Model of Visual Search.

Wurtz, R. (2008). Neuronal mechanisms of visual stability. *Vision Research, 48*, 2070-2089.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 458-468.

Zhang, W., & Luck, S. (2008). Discrete Fixed-resolution Representations in Visual Working Memory. *Nature, 453*, 233-235.