

**What color was it?: A psychophysical paradigm for tracking subjective progress in  
continuous tasks**

**Anna Kosovicheva, Peter J. Bex**

Department of Psychology, Northeastern University

This is the final copy of the following article, accepted for publication:

**Kosovicheva, A., & Bex, P. J. (2020). What color was it?: A psychophysical paradigm for tracking subjective progress in continuous tasks. *Perception*. DOI: 10.1177/0301006619886247.**

## Abstract

When making a sequence of fixations, how does the timing of visual experience compare to the timing of fixation onsets? Previous studies have tracked shifts of attention or perceived gaze direction using self-report methods (e.g., Carlson, Hogendoorn, & Verstraten, 2006; Hunt & Cavanagh, 2009). We used a similar method, a dynamic color technique, to measure subjective timing in continuous tasks involving fixation sequences. Does the time that observers report reading a word coincide with their fixation on it, or is there an asynchrony, and does this relationship depend on the observer's task? Observers read sentences that continuously changed in hue and identified the color of a word at the time that they read it using a color palette. We compared responses to a non-reading condition, where observers reproduced their fixations, but viewed non-word stimuli. Results showed a delay between the color of stimuli at fixation onset and the reported color during perception. For non-word tasks, the delay was constant. However, in the reading task, the delay was larger for earlier compared to later words in the sentence. Our results offer a new method for measuring awareness or subjective progress within fixation sequences, that can be extended to other continuous tasks.

Keywords: eye movements, fixations, visual attention, awareness, presaccadic attention

Our daily lives are filled with complex visual tasks that require sequences of eye movements, from searching for your keys, to tracking the positions of cars when crossing a busy road, to reading the words of this sentence. At what time do we experience performing these different operations, relative to our pattern of fixations? For example, when reading a sentence, does your experience of reading a particular word coincide with fixating it, or does it precede or lag behind the time you fixate the word? In this task, the reader must plan and execute a series of saccades to bring the relevant information from peripheral vision into the fovea, recognize the individual letters and words, and combine word-level information in order to extract the meaning of the text, taking into account any contextual information. The steps involved in reading have been extensively studied over the last several decades, with work focusing on the underlying visual and cognitive mechanisms, including factors that influence letter and word recognition, fixation behavior, and reading speed (for reviews, see Legge, 2007; Rayner, 2009; Rayner, Pollatsek, Ashby, & Clifton, 2012). In contrast, much less is known about the time course of the observer's internal experience, or subjective progress in completing such complex visual tasks, in relation to the timing of fixations on the relevant items.

The literature on visual attention and eye movements would suggest that an observer's impression of the time course of reading, or other continuous tasks, may be distinct from the timing of fixations on the individual items. For instance, it is well-known that the focus of visual attention can be separated from the point of gaze, both spatially and temporally. This is evident from cuing experiments, which have shown that observers respond faster to pre-cued stimuli while fixating at a different location (e.g., Posner, 1980); and from eye tracking experiments, which have used performance in a secondary task to demonstrate that shifts of visual attention precede the onset of saccades (e.g., Deubel & Schneider, 1996; Kowler, Anderson, Doshier, & Blaser, 1995). Moreover, perception of both gaze direction and time are altered near the moment of a saccade. Observers' perceived direction of gaze leads ahead of the time of the saccade, with estimates of lead time ranging from approximately 40 to 250 ms. (Deubel, Irwin, & Schneider, 1999; Hunt & Cavanagh, 2009). This shift in perceived gaze may be the consequence of a more general process that supports the perception of continuity across saccades. In addition, observers experience a compression in perceived time for probes flashed during a saccade (Morrone, Ross, & Burr, 2005), and a lengthening of subjective duration immediately following saccade completion (Yarrow, Haggard, Heal, Brown, & Rothwell, 2001). This post-saccadic dilation of

duration may be part of a more general phenomenon that occurs with the introduction of a new stimulus (e.g., Hodinott-Hill, Thilo, Cowey, & Walsh, 2002; Rose & Summers, 1995).

While all of these factors together can contribute to dissociations between fixations and awareness, most often these phenomena have been examined with single saccades, and generally without the addition of a continuous task. In reading, these effects are complicated by the nature of the task itself. Reading is a highly skilled process, in which factors such as context, expectation, and difficulty all might influence an observer's subjective timing when reading a word. The contributions of these factors to word and sentence processing have been examined using behavioral, eye tracking, and electrophysiological measures (e.g., Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Kliegl, Nuthmann, & Engbert, 2006; Kutas & Federmeier, 2011; Rayner, 1998), which have offered valuable insights into language processing. However, much less is known about the observer's subjective pace in performing the task, and specifically the time that the observers report reading individual words.

To address this question, we developed and tested a novel psychophysical self-report procedure for measuring subjective progress during reading, relative to the time of fixations on individual words. In principle, it could be used to track awareness in any continuous task that involves sequences of eye movements, though we note that fixation and saccade characteristics in other types of tasks may be different, even within an individual observer (see e.g., Andrews & Coppola, 1999; Rayner, Li, Williams, Cave, & Well, 2007). In the specific case of reading, we were primarily interested in two questions. First, are observers able to reliably report the time that they experienced reading a word? If so, we would expect to see consistency in their responses across trials, as well as some relationship between word position, gaze position and the reported time that they read a word. Second, what is the relationship between the observers' self-reports and fixation time for a word? To examine the role of task, we compared the results for reading with a baseline task in which observers reproduced approximately the same sequence of saccades over a sequence of rectangles matched in size to the same words. Observers again reported the color of a rectangle, and the comparison of the asynchrony between the physical color during fixation and the apparent color upon perception quantifies general effects of eye movements and response biases.

The complexity of reading does not lend itself easily to self-report, as it is necessary to mark the passage of time in a way that minimally interferes with the reading process itself.

Previous studies examining other perceptual and cognitive processes have used an independent temporal reference, such as a clock, to measure time delays between when the moment a sound is played and the time that it is experienced. While the earliest use of this method dates back to Wilhelm Wundt (1883), more recent studies have used videos of a moving clock to measure the time it takes to shift attention from one location to another (Carlson et al., 2006), as well as observers' perceived gaze direction around the time of a saccade (Hunt & Cavanagh, 2009). These methods have demonstrated that observers can reliably report the time that items are attended based on an independently changing feature dimension. Instead of asking observers to attend to a clock while simultaneously reading a sentence, our alternative is to alter the text in a way that changes at a constant rate over time. In our experiments, we accomplished this by presenting sentences that gradually cycled through a set of hues (e.g., from red to yellow to green) over time. At the end of each trial, we asked observers to report the color of a particular word at the time that they read it. In addition, we recorded observers' gaze positions to compare the reported color of a word (e.g., green) to the color at the time of fixation onset (e.g., yellow). Comparison between the time at which these colors were present on screen allows us to answer: do observers report that they read a word at the same time that they fixated it, sometime before, or sometime after? Importantly, we should note that observers' reports of the time that they read a word are not necessarily a measure of the time that semantic information or other properties of the word were actually registered (just as subjective reports of gaze position can differ from the recorded gaze position; Hunt & Cavanagh, 2009). Instead, by asking observers to report the color of the word at the time that they read it, this technique allows us to measure the timing of observers' subjective progress through the task, relative to their fixations on individual words.

### **Experiment 1: Simultaneous Presentation**

In Experiment 1, we performed a preliminary test of our method by asking observers, on each trial, to read randomly-generated four-word sentences that cycled through a continuous, circular set of hues, starting from a random hue and changing in a random direction (see Figure 1). At the end of each trial, observers were shown a response palette arranged in an annulus and were instructed to report the perceived color of a randomly selected (pre-cued) word at the time that they read it. In addition, they categorized each sentence as true or false, a task that required encoding all four words. To minimize the frequency of regressions to the beginning of the

sentence, we used observers' performance on this task to adaptively control the duration that each sentence was presented. Finally, we established a baseline for color responses in the absence of reading using a passive tracking task. In this condition, observers reproduced their eye movements from the reading task by tracking a crosshair, while viewing non-word stimuli (i.e., rectangles that covered the same area). Observers then reported the color of the rectangle at the time they fixated it. This comparison with a passive measure allows us to control for any general response strategies or biases, isolating the delays associated with performing the task itself.

## **Method**

**Participants.** Five observers (2 female, mean age: 33.2) participated in the experiment. All observers had normal or corrected-to-normal acuity and self-reported normal color vision as well as proficiency in English. Two observers were authors of the paper, and had behavioral responses comparable to those of the remaining observers, who were experienced in psychophysical experiments, but otherwise naïve to the purposes of the experiment (see individual data in Supplementary Figures S1-S4). Sample size was comparable to that of previous experiments using similar methods to examine the time course of shifts of attention (Carlson et al., 2006) and reading performance with the algorithm we used (Crossland, Legge, & Dakin, 2008). All observers gave informed consent prior to participating, in accordance with the Declaration of Helsinki. Procedures were approved by the Institutional Review Board at Northeastern University.

**Eye tracking.** Eye movements were recorded using an Eyelink 1000 desktop infrared eye tracker (SR Research Ltd., Mississauga, Ontario, Canada), used in conjunction with the Eyelink Toolbox for Matlab (Cornelissen, Peters, & Palmer, 2002). Gaze position was sampled binocularly at a rate of 1000 Hz. Observers completed a standard 9-point calibration procedure (Stampe, 1993) prior to each block of trials (mean error on validation = 0.43°). For each observer, the eye with the lowest mean error on validation was used for both the analysis and for the crosshair positions in the passive tracking task.

To reduce noise artifacts, a heuristic filtering algorithm was applied to the raw gaze position samples (see Stampe 1993, for details). Gaze information was then parsed into saccades and fixations, with the first time point at which the velocity exceeded 30°/s and the acceleration

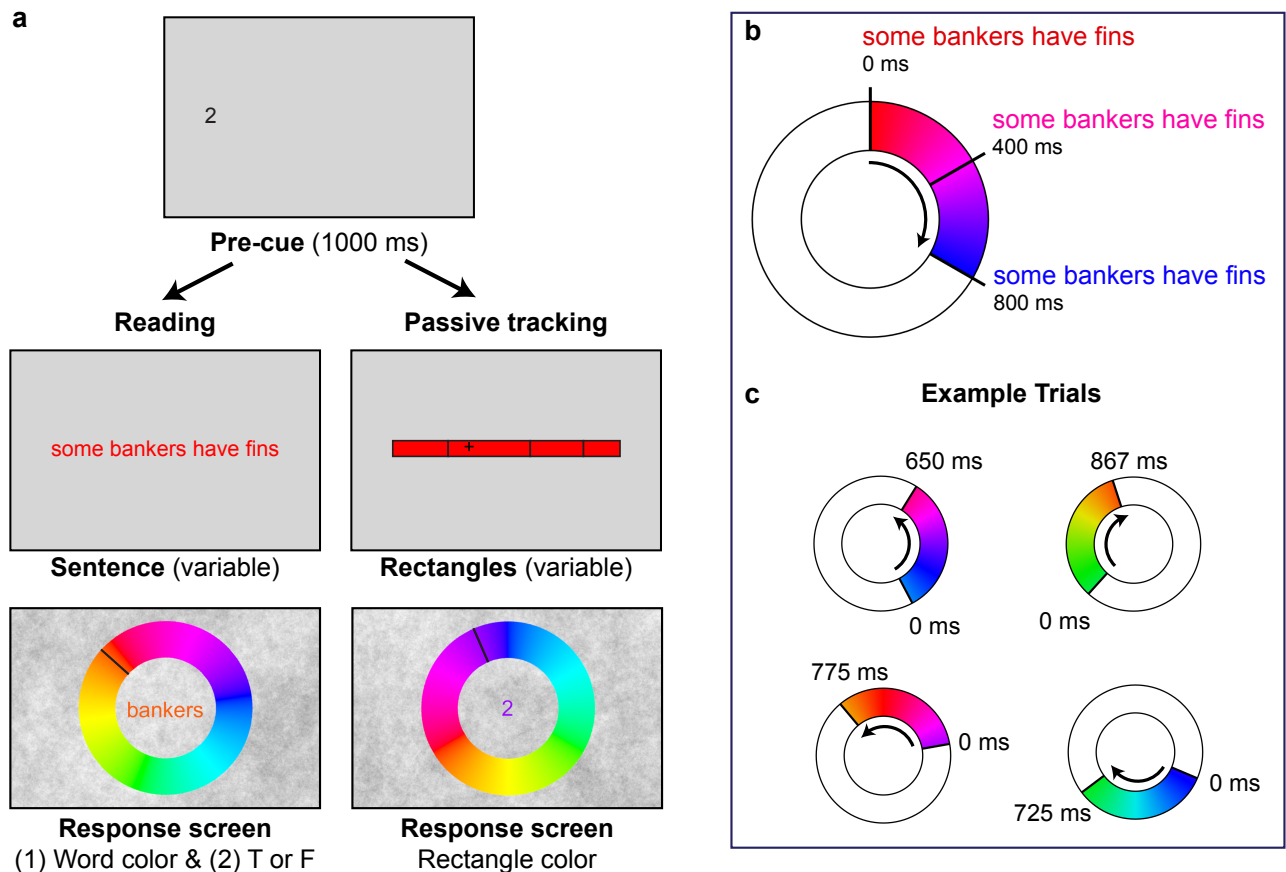
exceeded  $8000^{\circ}/s^2$  indicating the beginning of a saccade. Time points at which the velocity and acceleration fell below their respective thresholds defined the end of each saccade.

**Stimuli.** Stimuli were presented on a 27" gamma-corrected ASUS VG278HE LCD monitor and run on a Dell XPS 8300 computer with a Quadro FX 4600 graphics card. The experiment was programmed using the Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) in Matlab (The MathWorks, Inc., Natick, MA). Display resolution was set to  $1920 \times 1080$  and the refresh rate to 120 Hz. Observers viewed the display binocularly at a distance of 50 cm, using a chinrest to minimize head motion. Unless otherwise noted, stimuli were presented on a uniform gray background ( $36.26 \text{ cd/m}^2$ ) and all text was generated in 42 pt. Arial font, corresponding to a capital letter height of  $1.49^{\circ}$  (lowercase =  $1.07^{\circ}$ ) and a size at which reading speed are maximal in normally-sighted foveating observers (Mansfield, Legge, & Bane, 1996).

Sentences were four word statements that were randomly generated in Matlab (see Crossland, Legge, & Dakin, 2008 for a full description of the algorithm). Briefly, each sentence was constructed by first randomly selecting a noun (e.g., “bankers”, “cheese” “fish”, “hats”) and then a two-word description from a list that matched its category (e.g., humans, inanimate objects). For instance, if the noun was “teachers” selected, a random true (e.g., “eat food”) or false (e.g., “are venomous”) descriptor was chosen. To construct a sentence, one of three random quantifiers (“no”, “some”, or “all”) was selected and added to the beginning of the sentence (e.g., “no geese live underwater”). The sentences were generated such that half were true and half were false. Sentences were presented as a single line of text, and a new sentence was generated on each trial for each participant. With the exception of proper nouns, which were capitalized, sentences were displayed in all lowercase letters with two character spaces between consecutive words. The mean length of each sentence, including spaces, was  $26.2 \pm 3.1$  characters ( $23.14^{\circ} \pm 3.12^{\circ}$ ). Figure S5 in the Supplementary Materials shows additional data on individual word statistics (frequency, word length) as a function of word position.

Text colors were manipulated in HSV color space and varied only in hue, from  $0^{\circ}$  to  $360^{\circ}$ , while saturation and value were maintained at 100% (minimum luminance:  $11.36 \text{ cd/m}^2$ ; maximum luminance  $92.75 \text{ cd/m}^2$ , mean contrast: 19.23%). As shown in Figure 1b, on each trial, the initial hue was randomly selected between  $0^{\circ}$  and  $360^{\circ}$ , and traversed one-third of the range of hues ( $120^{\circ}$ ) at a uniform rate for the duration of the trial (see Procedure for duration

information). To prevent observers from anticipating the direction of the color change, the direction of change (clockwise or counterclockwise) was randomly selected on each trial. The average rate across all trials was 174.8 %/s (SD across participants = 39.1 %/s). In pilot testing, we found that this rate of change was optimal for observers to generally experience a single color per fixation, and easiest for them to report a single color at the end of the trial.



**Figure 1.** Stimuli used in the reading and passive tracking tasks in Experiment 1. As shown in (a), each trial began with a numeric pre-cue (1 through 4), which indicated which word (or rectangle) that observers would be responding to. They were then shown a four-word sentence (or a rectangle with four subregions) that gradually changed hue over time, as illustrated in (b). The set of hues always spanned 120° (one-third) of the full set of possible hues, starting at a random hue and changing in a random direction, as shown in the examples in (c). Duration on each trial was determined by a staircase procedure (see Method for details). At the end of the trial, observers were shown a response screen (a), in which they matched the color of the target word (or rectangle) to the color it was when they read it. In the reading trials, observers also indicated whether the sentence was true or false.

## Procedure.



**Reading task.** As shown in Figure 1, at the beginning of each trial, observers were presented with a number (1, 2, 3, 4) for a period of 1000 ms. Observers were told that this number served as a pre-cue (100% validity) that indicated which word (the first, second, third, or fourth) they would be prompted to respond to at the end of the trial. The number was displayed in black ( $0.21 \text{ cd/m}^2$ ) and positioned such that the left edge of the number was in the same location as the left edge of the sentence on the next screen.

Next, observers were shown a randomly selected four-word sentence (see Stimuli for description) at the center of the display. Trial duration was determined by two randomly interleaved staircases controlled by the QUEST algorithm (Watson & Pelli, 1983). Each staircase started was set to converge to 85% accuracy and consisted of 64 trials each. Following an initial duration estimate of 750 ms, the duration on each subsequent trial was calculated from the mean of the posterior probability density function (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994). The purpose of the staircases was to allow observers sufficient time to read the sentence in its entirety, while minimizing the frequency of regressions. Importantly, accurate true/false classification of each sentence requires the reader to process all four words, and the removal of any one word produces performance below the threshold level of 85% correct (Crossland et al., 2008). Observers were instructed to read the sentence from left to right, as they would normally and to determine whether it was true or false.

Finally, observers were shown a response screen consisting of a response palette on a noise background ( $36.26^\circ$  square field of  $1/f$  noise at 50% contrast), which served as a mask. A new noise background was generated on each trial. The response palette was a screen-centered annulus ( $9.09^\circ$  outer radius and  $5.26^\circ$  inner radius) that varied in hue from  $0^\circ$  to  $360^\circ$  as a function of spatial angle in 512 equal steps across the full wheel. The mapping between hue angle and spatial angle was randomized on each trial (i.e.  $0^\circ$  in hue space was randomly assigned to a spatial angle between  $0^\circ$  and  $360^\circ$ ).

On each trial, the target word was displayed centrally within the annulus. Observers were instructed to use the response palette to select the color of the target word at the time that they read it. Observers moved the mouse which controlled an angular “cursor” (a black line,  $3.84^\circ$  long and  $0.07^\circ$  wide) positioned inside the annulus. The initial angle of the cursor was randomly selected on each trial. As observers moved the cursor, the color of the central target word changed to match the corresponding color on the annulus. Observers were instructed to match

the color of the target word to the color it was at the time that they read it. In addition, observers indicated whether the sentence was true or false. Observers made their responses by first using the mouse to adjust the cursor position to the appropriate color and then pressing the left mouse button if the sentence was false, and the right mouse button if the sentence was true. Following the mouse click, the program advanced to the next trial.

Observers completed 256 trials, divided into two blocks of 128 trials each. Each block consisted of four trials for every unique combination of target word position (1, 2, 3, or 4) sentence type (true or false), color change direction (clockwise or counterclockwise) and staircase (one of two), presented in a random order. Observers completed at least 32 practice trials before starting the experiment, which were not included in the final analysis.

***Passive tracking task.*** Following completion of the reading task, observers performed a passive tracking task, in which the stimuli and procedure were designed to match each trial in the reading task as closely as possible. For each observer, every trial in the reading task had a corresponding, “matched” trial in the passive tracking task, that was identical in duration, target number (1 through 4), and the colors presented. Each observer viewed the same trials again (in two blocks of 128 trials each), in the same order, with the exceptions described below.

Following the numerical pre-cue, observers were shown a rectangle, divided into four subregions, filled the same color as the corresponding sentence in the reading condition, and outlined in black ( $0.07^\circ$  width). The rectangle was the same height and length as the corresponding sentence in the reading task, and each subregion corresponded to the position of each word in the reading task (lines were drawn in between the two spaces that divided consecutive words).

Instead of reading, observers were instructed to track a black crosshair ( $0.5^\circ$  in length with a stroke width of  $0.07^\circ$ ) by following it with their eyes as closely as possible. The position of the crosshair was based on the ending x- and y- gaze positions of each fixation previously recorded from the same observer for the matching in the reading task. Positions were taken every fixation that temporally overlapped with the presentation of the sentence (i.e., every fixation that ended after the onset of the sentence and started before the onset of the response screen). In order to match the timing of observers’ gaze positions between the reading and passive tracking tasks as closely as possible, the cursor was placed at the corresponding positions of each fixation, but shifted 200 ms earlier in time. In other words, if a given fixation in the reading task began 450

ms following the onset of the sentence, a crosshair at its location in the passive tracking task was presented 250 ms following the onset of the rectangles (rounded to the nearest frame). This leading interval was introduced to allow additional time for observers to respond to changes in its position (i.e., time to make a saccade), and its duration was determined from both pilot testing and known saccade latencies (Carpenter, 1988). For fixations that started before the onset of the sentence (i.e., during the pre-cue), the cross was presented on the pre-cue display as well. The crosshair representing the last fixation on every trial was presented up until the onset of the response screen, after which it was removed.

To confirm that this chosen lead time for presenting the cross (200 ms) resulted in temporally similar gaze position traces between the reading and passive tracking conditions, we separately calculated the root-mean-square (RMS) error of the gaze position between matched pairs of trials in the two conditions, at different temporal offsets relative to one another. Then, for each trial, we determined the lag that produced the lowest RMS value. Across observers, we observed no consistent temporal biases in either direction; the RMS error between the recorded traces was lowest when the passive tracking condition lagged behind the reading condition by 10.5 ms (95% CI [3.2, 19.1]; see Figure S7 for details and additional data on tracking accuracy).

The response screen in the passive tracking task was identical to the screen used in the reading task, with a number (1, 2, 3, or 4) presented centrally indicating the target rectangle. Observers were instructed to adjust the cursor to match the color of the target rectangle at the time that they fixated it.

**Data analysis.** For each trial, we calculated the time point closest to the observer's color selection, relative to both the onset of the trial and the time of the observer's first fixation on the target. To analyze each observer's responses relative to the onset of the trial, we first found the time point in the trial at which the sentence was the same hue selected by the observer. Because only one third of a full wheel was presented on any given trial (and observers were given the full response palette to choose from), observers' responses would occasionally fall outside the set of colors presented. For those trials, we extrapolated the remaining time points based on the rate of color change for the trial analyzed. We then calculated the proportion of one full cycle corresponded to the observer's response, where 0 represents the beginning of the trial, and 1/3 corresponds to the end of the trial. Due to the circular nature of the data (where 1 represents the

end of a full cycle), we fit the distribution of proportions to a two-parameter von Mises distribution (mean,  $\mu$ , and concentration,  $\kappa$ ) using maximum likelihood estimation.

To analyze observers' responses relative to their fixation on the target word (or rectangle, in the passive tracking task), we first determined the first fixation that spatially overlapped with the target, based on whether it was inside a rectangular box that matched the height and length of the word (spaces between words were divided evenly). On trials in which there were multiple fixations on the target word, only the first fixation was analyzed. Trials in which there was no fixation on the target word were excluded from this analysis (15.3% of all trials).

Confidence intervals for parameter estimates were estimated using a bootstrapping procedure (Efron & Tibshirani, 1993). For each observer, individual trials were resampled with replacement and the corresponding responses were fit to a von Mises distribution. This procedure was repeated for 1000 iterations to produce a 95% confidence interval around the parameter estimate. Unless otherwise specified, statistical comparisons were performed using permutation tests, by first shuffling responses from individual trials between pairs of conditions (e.g., between the first and second target word), and then calculating the mean difference between conditions. This procedure was repeated for 1000 iterations to obtain a null distribution of differences between a given pair of conditions.

The full set of data and materials are available on the Open Science Framework online: <https://osf.io/fhwat/>.

## **Results and Discussion**

Observers' true-false classification accuracy was close to the expected threshold level of 85% (mean observed performance: 85.9%), produced by presenting sentences for an average duration of 736 ms (*SD* within each observer: 155 ms).

In analyzing observers' color responses, we first determined whether observers were able to perform the task accurately. If observers are accurate in their color selection, we would expect a systematic relationship between the target word number (1 through 4) and the time point corresponding to the color selected by the observer. In other words, when asked to report the color of the first word, they should, on average, report colors presented early in the trial. When asked to report the color of the fourth word, observers should report colors presented late in the trial. In contrast, if observers guess randomly, or always report the color presented at the beginning (or the end) of the trial, we should see no relationship between target word number

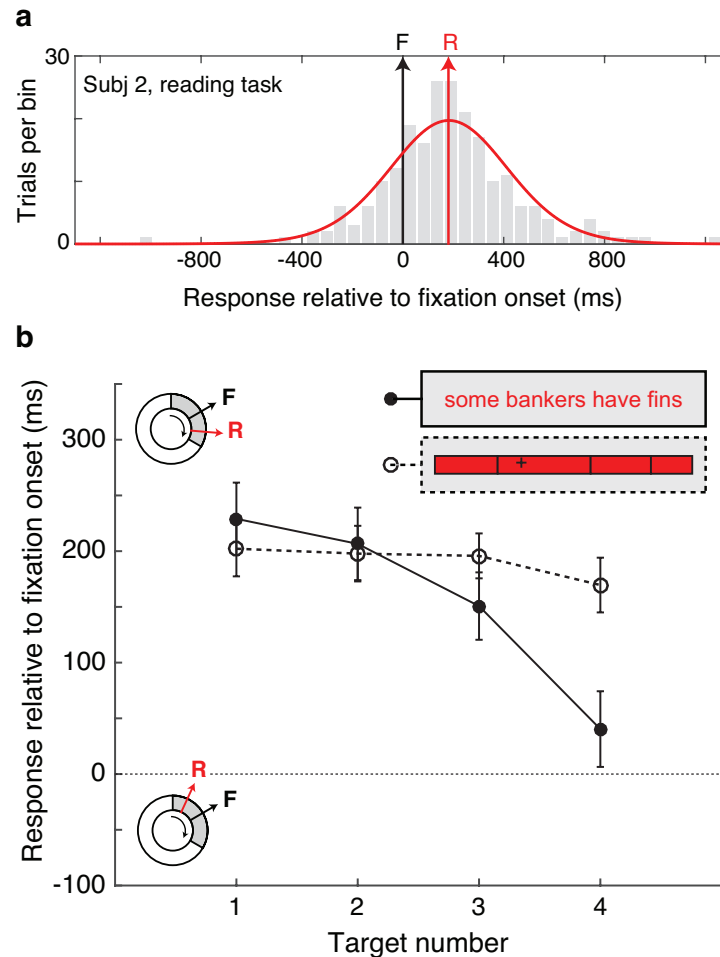
and color selection. Figure S1 in the Supplemental Material shows the mean responses, relative to trial onset, for each target word in the reading task, based on fitting the response distribution to a von Mises distribution. As expected, when asked to report the color of the fourth word, observers selected colors presented later in the trial (0.28 of the way through a cycle, where 1/3 represents the end of the trial, 95% CI [0.26, 0.29]) than when asked to report the color of the first word (0.11 of one cycle, 95% CI [0.10, 0.13]),  $p < .001$  (permutation test). We observed a similar relationship in the passive tracking task—when asked to report the color of the fourth rectangle, observers selected colors that were presented later in the trial (0.29 of one full cycle, 95% CI [0.28, 0.30]) than when asked to report the color of the first rectangle (0.11 of one cycle, 95% CI [0.10, 0.12], see Figure S1),  $p < .001$ .

In addition, to verify that observers were responding reliably, we performed a permutation test to compare the observed widths of the response distributions to those expected by random guessing. These were estimated by randomly shuffling, across trials, the mapping between observers' responses and the color corresponding to the onset of the trial, and then calculating the proportion of one cycle corresponding to the observer's response. We then calculated the resulting concentration parameter ( $\kappa$ ) from the von Mises distribution and repeated this procedure for 1000 iterations to obtain a null distribution of concentration parameters produced by chance performance.

Figure S2 shows the observed distribution widths in both the reading and passive tracking tasks, with the mean of the null distribution for comparison. For each target word number (1 through 4) in both types of tasks (reading and passive tracking), observers' response distributions were narrower than those produced by chance (all  $p$  values  $< .001$ ). The distribution widths, together with the observed relationship between target word number and observers' responses, demonstrate that observers were reliably reporting the colors presented at different time points in the trial, and were not simply reporting the first or last color presented. Although observers were accurately able to report the colors presented, we also note that color responses were generally biased toward category centers (e.g., red, yellow, green; Figure S3), consistent with centering or categorical biases in many perceptual judgments (e.g., Hollingworth, 1910; Huttenlocher, Hedges, & Duncan, 1991). However, since the color starting point was randomly selected on each trial, the distance between the category center and the color at trial onset (or at

fixation onset) is random on each trial. We would therefore expect this to introduce additional variance, but not systematic directional biases, when averaged over trials.

To analyze observers' color responses relative to the first fixation on the target word, we subtracted the time point following the onset of the trial at which observers first fixated the target word from the time point corresponding to the color they selected. Positive values represent responses that lagged behind the onset of the observer's fixation on the target word, and negative values represent responses that were ahead of the observers' fixation on the target word. A value of zero would indicate that the observer selected the color presented at exactly the time of the onset of their fixation on the target word. Figure 2A shows a representative distribution of one observer's responses relative to fixation onset and the corresponding von Mises fit. Figure 2B shows the average delays as a function of target word number for both the reading and the passive tracking tasks (see Figure S4 for individual observer data). In the reading task, the lag relative to fixation onset decreased from 229.5 ms (95% CI [196.9, 261.2]) to 40.4 ms (95% CI [5.8, 74.9]). All pairwise comparisons between lags at each target word, aside from the first and second target words ( $p = .306$ ), were significant at a Bonferroni-corrected alpha of .008, all  $p$ -values  $< .003$ . In contrast, in the passive tracking task, the lag relative to fixation onset decreased from 202.2 ms (95% CI [178.5, 225.9]) to 169.6 ms (95% CI [146.2, 193.0]), and none of the pairwise comparisons were significant ( $\alpha = .008$ ), all  $p$ -values  $> .015$ .



**Figure 2.** (a) Distribution of responses relative to the onset of the first fixation on the target word for a representative observer. Fixations were aligned across trials such that 0 ms represents fixation onset (F). Positive values represent responses (R) corresponding to time points in the trial following fixation onset, and negative values represent time points prior to fixation onset. The curve shows the fitted von Mises distribution. (b) Data were fitted separately for each target word number (1 through 4) in both the reading (solid line) and passive tracking tasks (dashed line), and averaged across observers. Error bars represent bootstrapped 95% confidence intervals.

In Experiment 1, we developed and tested a new procedure to measure subjective progress in a reading task, in which we instructed observers to report colors that corresponded to the time when they experienced reading particular words in a sentence (which may be distinct from the time that semantic information or other properties of the word were actually registered). We compared color reports to a passive tracking task, in which subjects reproduced the same set of fixations, reporting the color of a rectangle at the time that they fixated it. Observers' responses indicated that they could reliably report previously seen colors in both tasks. Color responses, measured relative to the start of the trial, systematically changed as a function of

position, and response variability was well below chance levels of performance. In addition, in measuring the delay between fixation onset and reported color, we observed distinct patterns of lags in the two tasks. When required to read a sentence, the delays in observers' responses were longer for earlier words in the sentence than for later words in the sentence. In contrast, when viewing non-word stimuli—but reproducing the same eye movements—observers' responses were relatively constant as a function of target position.

The observed pattern of results in the reading condition is unlikely to be explained by biases in observers' responses, for instance, due to poor memory for the presented color, or a general tendency to report colors presented towards the end of a trial, or simple strategies based on memorizing the range of colors presented. Observers were also similarly precise in their color reports as a function of word position, with similar distribution widths ( $\sqrt{1/\kappa}$ ) of 0.102 (95% CI [0.091, 0.110]) and 0.085 (95% CI [0.072, 0.094]) for the first and fourth words in the reading condition, respectively (Figure S2a). Importantly, in our passive tracking condition, observers viewed the same colors and made the same pattern of eye movements, but were shown non-word stimuli. Therefore, if observers are using a general response strategy across the two conditions, these biases should be reflected in the passive tracking task as well (and we additionally rule out the possibility of an interaction between task and response strategy in Experiment 2; see General Discussion). Similarly, the presence of a pre-cue cannot account for the difference in results, as it was present in the passive tracking task as well as the reading task. Another interpretation is that observers' responses are simply a reflection of the total duration that they fixated a word. However, the total gaze duration on a target word was similar in the two tasks, with both conditions showing a similar decline in gaze duration with later words in the sentence (see Figure S6).

## **Experiment 2: Sequential Presentation**

In the absence of explanations based on response bias, these results point to a difference in the subjective pacing between the reading and passive tracking conditions. As they read a sentence, observers reported reading individual words with progressively shorter delays following fixation onset, indicating a perceived speeding up of their progress towards the end of the sentence. By the time they reach the final word in the sentence, observers' reported reading time is earlier than their reported fixation time in the rectangle condition. What accounts for this



difference between the reading and passive tracking conditions? As there are a number of possible explanations for this pattern of results, we divide them into two general categories. One possible class of explanations is related to the contents of the sentence itself. For example, the sentence generation algorithm might produce consistent variation in the difficulty or predictability of the words as a function of their position in the sentence, and observers might report reading words with a shorter delay, relative to fixation onset, if they are shorter, easier, or more predictable, and therefore processed faster. We note that neither word length or frequency showed the same pattern of gradual decrease (or increase) as a function of position in the sentence (see Figure S5). However, these statistics do not fully capture information about the predictability of words in a sentence, and it is possible that later words in these sentences might be more predictable than earlier words.

Another possible set of explanations is related to more general strategies involved in extracting information across eye movements during reading, that are not specific to the contents of the sentence. Although subjects are making the same pattern of fixations in the reading and passive tracking conditions, observers may allocate attention differently between the two conditions, for example, by shifting attention ahead of fixation at progressively earlier time points. These pre-saccadic shifts of attention might be reflected in observers' color reports. For example, previous work on transsaccadic feature integration has shown that when a target changes color during a saccade, observers' percepts reflect a mixture of the pre and post-saccadic colors (Oostwoud Wijdenes, Marshall, & Bays, 2015; Schut, Van der Stoep, Fabius, & Van der Stigchel, 2018). This phenomenon might contribute to the pattern of results we observed in the reading condition in Experiment 1, if pre-saccadic shifts of attention occur progressively earlier (relative to saccade onset) as observers make a sequence of saccades to words in the sentence. In the next experiment, we measured perceived reading time in a procedure that eliminated eye movements between successive words, presenting them sequentially at the fovea, rather than simultaneously, in an effort to differentiate between these two possible classes of explanations.

We tested this by repeating Experiment 1, but instead presenting the sentences one word at a time in a rapid serial visual presentation (RSVP) stream (see Figure 3). If the decrease in lag for later words in the sentence is a result of factors related to the content of the sentences alone, we would expect the same decrease in delay for later words in the sentence that we observed in Experiment 1. However, if the decrease in lag for later words is a result of oculomotor and

attentional strategies that are specific to reading a continuous line of text, we would expect the response lags to remain relatively constant as a function of word position, as observers are fixating a single location. In this instance, they are presented with one word at a time and have no other stimulus to which they can direct their attention or saccades.

## **Method**

**Participants.** Four observers (1 female, mean age: 33.0), who participated in Experiment 1 also participated in Experiment 2. As before, observers reported normal or corrected-to-normal vision and gave informed consent prior to participating.

**Stimuli and procedure.** The stimuli and procedure used in Experiment 2 were identical to those in Experiment 1, with the following exceptions.

In the reading task, following the numerical pre-cue, observers viewed each sentence one word at a time (see Figure 3B). Both the pre-cue and the words were presented centrally on the screen (mean word length =  $5.03 \pm 0.78$  characters) and the total duration of each sentence was selected to match the total sentence durations used in Experiment 1. For each observer, a random sentence duration was selected (without replacement) from their data in Experiment 1, and each word was presented for one-fourth of the total time, rounded to the nearest frame. For instance, if the randomly drawn duration from Experiment 1 was 700 ms, each word was presented for 175 ms in Experiment 2. Each word was immediately replaced by the next word, and the last word was immediately replaced by the response screen. Although the durations were drawn from each observer's previous data, a new sentence was re-generated, and a random starting hue and direction (clockwise vs counterclockwise) was selected on each trial. The hue of each word changed continuously (in equal steps) throughout the trial, irrespective of the abrupt changes in the text itself.

As in Experiment 1, observers completed a passive viewing task, in which they were shown a set of four rectangles that occupied the same regions as the words in the reading task. Each rectangle was presented centrally for the same duration as the corresponding word in the reading task. However, as each word was presented foveally, there was no crosshair on the display. The response screens for the reading and passive viewing tasks were identical to those used in Experiment 1, and observers completed a total of 512 trials across four blocks (two sets of 128 trials for both the reading and passive viewing task).

**Data Analysis.** The data analysis procedure for Experiment 2 was the same as that used in Experiment 1. However, since observers were viewing the words foveally, the time of the onset of each word (rather than the onset of the nearest fixation) was subtracted from observers' responses.

## Results and Discussion

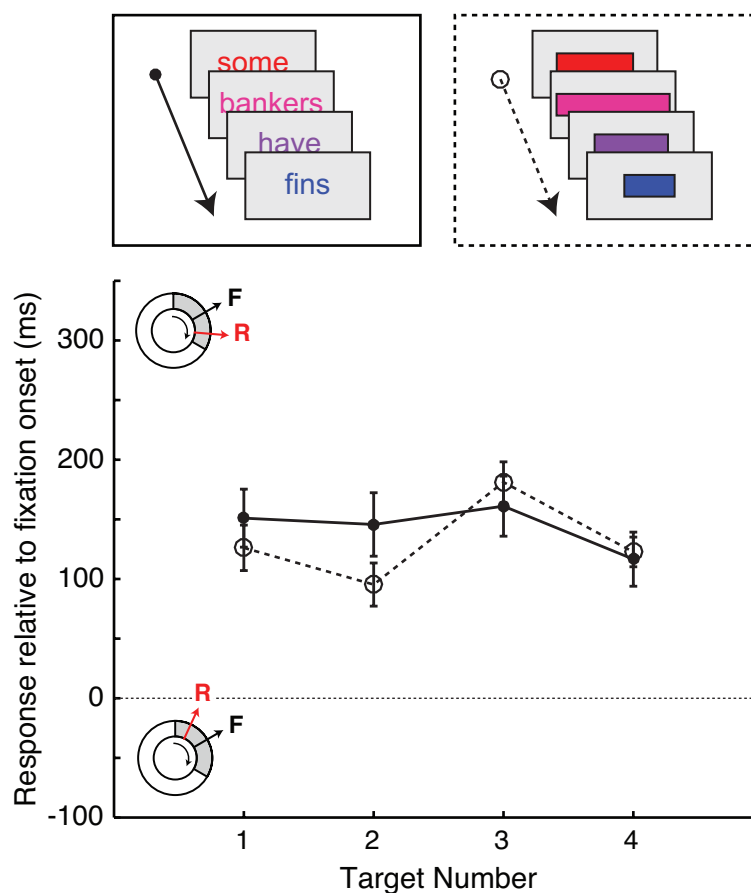
Following our analyses from Experiment 1, we first determined whether observers were able to perform the task reliably. As in Experiment 1, when reporting the color of the fourth word, observers selected colors presented later in the trial than when reporting the color of the first word, in both the reading task (0.30 of one cycle, 95% CI [0.29, 0.31] vs. 0.07 of one cycle, 95% CI [0.06, 0.08],  $p < .001$ ) and in the passive viewing task (0.30 of one cycle, 95% CI [0.29, 0.31] vs 0.06 of one cycle, 95% CI [0.05, 0.07],  $p < .001$ ; see Figure S1). In addition, for each of the four target words, observers' response distributions were significantly narrower than those produced by chance (see Figure S2, all  $p$  values  $< .001$ ).

Figure 3 shows mean response delays relative to fixation onset (in this case, fixation onset and word onset are equivalent) for both the reading and passive viewing conditions (see Figure S4 for individual observer data). As in Experiment 1, positive values represent color responses corresponding to time points following fixation onset, and negative values represent time points prior to fixation onset.

In contrast to the results in Experiment 1, the lag (relative to fixation onset) remained relatively constant between the first and last words in the reading task, decreasing slightly from 151.0 ms (95% CI [126.8, 175.2]) to 116.6 ms (95% CI [94.7, 138.5]). Aside from the comparison between the third and fourth target word, ( $p = .008$ ), none of the pairwise comparisons between lags at each target word reached significance ( $\alpha = .008$ ), all  $p$ -values  $> .02$ . Similarly, in the passive viewing task, the lag relative to fixation onset was similar between the first (126.1 ms; 95% CI [107.7, 144.5]) and fourth (122.7 ms; 95% CI [109.8, 135.6]) targets. On average, observers showed later responses relative to fixation onset for the third target compared to each of the three remaining targets (all  $p$ -values  $< .001$ ). However, none of the remaining pairwise comparisons reached significance ( $\alpha = .008$ ; all  $p$ -values  $> .02$ ).

Therefore, in Experiment 2, we can conclude that presenting sentences sequentially, eliminating readers' ability to plan saccades or shift attention outside the fovea, results in an approximately constant reported reading time following fixation onset. The decrease in reported

reading time observed in Experiment 1 is unlikely a consequence of sentence content alone, as sentences were produced with the same algorithm, and followed a similar logical structure. Although we observed relatively constant delays between fixation onset and the reported color in Experiment 2, one remaining possibility is that this pattern of ‘flattening’ is an artifact of having the full sentence duration divided equally among the four words. This even distribution is different from the distribution in Experiment 1, which showed a general decrease in gaze duration with later words in the sentence (see Figure S6). To verify that distributing word duration evenly alone does not produce this constant pattern of delays, we re-analyzed the results of Experiment 1, analyzing a subset of trials which had an even or nearly-even distribution of fixation durations among the four words (see Figure S8 for details). This subset of the results in Experiment 1 still showed a comparable decrease in lag for later words in the sentence, and a similar difference between the reading and passive tracking conditions. Therefore, it is unlikely that the pattern of flattening in Experiment 2 is primarily due to the even distribution of word durations.



**Figure 3.** Observers' response delays relative to fixation onset (equivalent to word onset) in the sequential presentation experiment (Experiment 2). Delays are plotted separately for the reading and passive viewing conditions (solid and dashed lines, respectively; see legend). Positive values represent responses corresponding to time points in the trial following fixation onset, and negative values represent time points prior to fixation onset. Error bars represent bootstrapped 95% confidence intervals.

### General Discussion

How do an observer's self-reports of performing a continuous task compare to the timing of their fixations? These experiments provide a basic demonstration of a novel dynamic color technique, in which we measured the time that observers reported reading words within short sentences, tracking their subjective progress. On average, across word positions, observers typically reported reading a word 150-160 ms following fixation onset. In addition, when showing observers a continuous line of text, we detected a robust position-based effect, in which the lag between fixation onset and reported color was lower for later words (40-100 ms) compared to earlier words (200-220 ms) in the sentence, indicating a subjective increase in the speed with which they performed the task. By presenting each word sequentially, we showed that this position-based change in delay is eliminated when observers are shown words individually, and cannot gain additional information by either making a saccade or shifting attention outside the fovea. Again, these timings correspond to observers' self-reports in completing this task, which may be distinct from the time that words were actually processed.

Furthermore, we used a baseline passive tracking condition to exclude a number of alternative explanations for this position-dependent effect. Any general response bias or strategy in reporting previously seen colors should be present in both tasks, allowing us to isolate the delays associated with observers' experience of performing the task. However, one remaining concern is that there may be an interaction between task and response strategy. In other words, the reading task necessarily requires an additional cognitive processing load, and the demands from having a second task might produce differential patterns of response bias, or perhaps poorer memory of the target's color than they would in a passive condition. The results from the reading task in Experiment 2 directly address this issue. In Experiment 2, observers performed a reading task with the same cognitive demands (reading a sentence and classifying it as true or false) as those in Experiment 1. Here, the pattern of response delays remains relatively constant as a

function of word position. Therefore, the decreasing pattern of lags in the first experiment is unlikely due to task difficulty alone.

As discussed previously, observers' self-reports using similar methods have provided a rich source of information about attentional processes and perceived continuity across saccades (e.g., Carlson et al., 2006; Hunt & Cavanagh, 2009; Yarrow et al., 2001). What can we learn from an observer's self-reported time of reading a word? Our results in Experiment 1 are broadly consistent with an increase in the subjective speed of the task, relative to fixations, near the end of a sentence. The absence of this effect in Experiment 2, despite similar content (i.e., sentences produced by the same algorithm) suggests that observers' reports may be tied more closely to lower-level attentional and oculomotor processes involved in reading, including pre-saccadic shifts of attention, rather than higher-level cognitive and linguistic processes. This is further supported by the observations of similar patterns of lags between the reading and passive tracking conditions in Experiment 2. If subjective pace were a product of sentence content alone, we would expect very different lags between the reading and passive tracking conditions. Together, these results suggest that there may be an important dissociation between observers' self-reports and higher-level cognitive processing during reading.

Nevertheless, these results point to processes that are specific to reading, as we observe differences between reading and passive tracking when observers are presented with the entire stimulus at once. One possible account for this difference is that attention shifts ahead of fixation, at progressively earlier time points during reading, resulting in shorter delays in reported reading time relative to fixation. The literature on attention and eye movements during reading generally supports the idea that attention can shift outside of the currently fixated word, though there has long been debate as to how attention is allocated during reading, specifically whether attention shifts serially over individual words or whether attention is distributed along a gradient (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Pollatsek, Fisher, & Rayner, 1998). This distinction has important implications for language processing and eye movements during reading, particularly in relation to the question of how much information is processed outside of the fovea (see Rayner, 2009 for a review). Other work measuring shifts of attention over saccadic sequences in non-reading tasks has shown that observers generally allocate attention to the current point of fixation and one target ahead (Gersch, Kowler, & Doshier, 2004), though attention can be distributed over multiple future saccade targets in some

situations (Baldauf & Deubel, 2008; Gersch, Kowler, Schnitzer, & Doshier, 2009; Godijn & Theeuwes, 2003). We note that these differences in the allocation of attention between the reading and passive tracking conditions may also be partly attributable to differences in saccadic programming (i.e., voluntary vs reflexive) between the two conditions. In the reading task, saccadic programming is self-paced, while in the passive tracking task, saccades are programmed in response to a change in the crosshair's position. Although the results of the present study cannot definitively address the question of attentional allocation during reading, further work measuring observers' subjective progress may offer useful insights into attention within saccadic sequences. For example, future work might examine whether the observed pattern of results is specific to sentences with a regular structure, such as the ones used here, and the degree to which this structure might facilitate saccadic pre-planning.

Together, the procedure and preliminary results presented here have demonstrated a new psychophysical method for measuring subjective progress in continuous visual tasks. In using a reading task as a test case, we have measured observers' self-reports of the time that they experienced reading individual words, which to our knowledge, have not been systematically studied before. These self-reports provide a novel metric that could be used in conjunction with existing tools to study strategies in a large number of visual and cognitive tasks that are otherwise difficult to probe. Importantly, the inclusion of a baseline condition allows us to measure the relative timing between a reading task and a passive tracking control. Future work might examine whether task demands (i.e., the addition of the color report itself, present in both conditions) might influence the subjective timing of the task for participants. Furthermore, while the demonstration here is relatively simple, we expect that variants of this paradigm could be utilized for more complicated sentences and passages, or extended to other types of continuous tasks (e.g., multiple object tracking, visual search; Kosovicheva, Feffer, Alaoui Soce, Cain, & Wolfe, 2017), or integrated with continuous report techniques (Bonnen, Burge, Yates, Pillow, & Cormack, 2015). In addition, estimates of the subjective pace of reading tasks in other populations (e.g., individuals with reading impairments, different age groups) may provide insights into the differences in reading and attentional processes in these populations.

**Acknowledgments:** This work was supported by funding from the National Institutes of Health (R01 EY029713 to PJB and F32 EY028814 to AK).

## References

- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39(17), 2947–2953. [http://doi.org/10.1016/S0042-6989\(99\)00019-X](http://doi.org/10.1016/S0042-6989(99)00019-X)
- Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research*, 184(3), 411–425. <http://doi.org/10.1007/s00221-007-1114-x>
- Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, 15(3), 1–16. <http://doi.org/10.1167/15.3.14>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <http://doi.org/10.1163/156856897X00357>
- Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, 6(2006), 1406–1411. <http://doi.org/10.1167/6.12.6>
- Carpenter, R. H. S. (1988). *Movements of the Eyes* (2nd ed.). London: Pion.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. <http://doi.org/10.3758/BF03195489>
- Crossland, M. D., Legge, G. E., & Dakin, S. C. (2008). The development of an automated sentence generator for the assessment of reading speed. *Behavioral and Brain Functions*, 4(1), 1. <http://doi.org/10.1186/1744-9081-4-14>
- Deubel, H., Irwin, D. E., & Schneider, W. X. (1999). The subjective direction of gaze shifts long before the saccade. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current Oculomotor Research* (pp. 65–70). New York: Plenum Press.
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827–1837. [http://doi.org/10.1016/0042-6989\(95\)00294-4](http://doi.org/10.1016/0042-6989(95)00294-4)
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of Eye Movements and EEG in Natural Reading : Analyses and Review. *Journal of Experimental Psychology: General*, 140(4), 552–572. <http://doi.org/10.1037/a0023885>
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. London: Chapman & Hall.
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, 112(4), 777–813. <http://doi.org/10.1037/0033-295X.112.4.777>
- Gersch, T. M., Kowler, E., & Doshier, B. (2004). Dynamic allocation of visual attention during the execution of sequences of saccades. *Vision Research*, 44(12), 1469–1483. <http://doi.org/10.1016/j.visres.2003.12.014>



- Gersch, T. M., Kowler, E., Schnitzer, B. S., & Doshier, B. A. (2009). Attention during sequences of saccades along marked and memorized paths. *Vision Research*, 49(10), 1256–1266. <http://doi.org/10.1016/j.visres.2007.10.030>
- Godijn, R., & Theeuwes, J. (2003). Parallel Allocation of Attention Prior to the Execution of Saccade Sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 882–896. <http://doi.org/10.1037/0096-1523.29.5.882>
- Hodinott-Hill, I., Thilo, K. V, Cowey, A., & Walsh, V. (2003). Auditory Chronostasis. *Current Biology*, 12(20), 1779–1781. [http://doi.org/10.1016/s0960-9822\(02\)01219-8](http://doi.org/10.1016/s0960-9822(02)01219-8)
- Hollingworth, H. L. (1910). The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461. <http://doi.org/10.2307/2012819>
- Hunt, A. R., & Cavanagh, P. (2009). Looking ahead: The perceived direction of gaze shifts before the eyes move. *Journal of Vision*, 9(9), 1–7. <http://doi.org/10.1167/9.9.1>
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and Particulars: Prototype Effects in Estimating Spatial Location. *Psychological Review*, 98(3), 352–376. <http://doi.org/10.1037/0033-295X.98.3.352>
- King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, 34(7), 885–912. [http://doi.org/10.1016/0042-6989\(94\)90039-6](http://doi.org/10.1016/0042-6989(94)90039-6)
- Kleiner, M., Brainard, D., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVF Abstract Supplement*, 36(14), 1.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12–35. <http://doi.org/10.1037/0096-3445.135.1.12>
- Kosovicheva, A., Feffer, J., Alaoui Soce, A., Cain, M., & Wolfe, J. M. (2017). When does visual search move on?: Using the color wheel to measure the dynamics of foraging search [Abstract]. *Journal of Vision*, 17(10). <http://doi.org/10.1167/17.10.86>.
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The Role of Attention in the Programming of Saccades. *Vision Research*, 35(13), 1897–1916.
- Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review Of Psychology*, 62, 621–647. <http://doi.org/10.1146/annurev.psych.093008.131123>
- Legge, G. E. (2007). *Psychophysics of Reading in Normal and Low Vision*. Boca Raton: CRC Press.
- Mansfield, J. S., Legge, G. E., & Bane, M. C. (1996). Psychophysics of reading in normal and low vision. *Investigative Ophthalmology & Visual Science*, 37(8), 1492–1501.
- Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience*, 8(7), 950–954. <http://doi.org/https://doi.org/10.1038/nn1488>

- Oostwoud Wijdenes, L., Marshall, L., & Bays, P. M. (2015). Evidence for optimal integration of visual feature representations across saccades. *Journal of Neuroscience*, 35(28), 10146–10153. <http://doi.org/10.1523/JNEUROSCI.1040-15.2015>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <http://doi.org/10.1163/156856897X00366>
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <http://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. <http://doi.org/10.1080/17470210902816461>
- Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, 47, 2714–2726. <http://doi.org/10.1016/j.visres.2007.05.007>
- Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading* (2nd ed.). New York: Psychology Press.
- Reichle, E. D., Pollatsek, a, Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. <http://doi.org/10.1037/0033-295X.105.1.125>
- Rose, D., & Summers, J. (1995). Duration illusions in a train of visual stimuli. *Perception*, 24(10), 1177–1187. <http://doi.org/10.1068/p241177>
- Schut, M. J., Van der Stoep, N., Fabius, J. H., & Van der Stigchel, S. (2018). Feature integration is unaffected by saccade landing point, even when saccades land outside of the range of regular oculomotor variance. *Journal of Vision*, 18(7), 1–17. <http://doi.org/10.1167/18.7.6>
- Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2), 137–142. <http://doi.org/10.3758/BF03204486>
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–20. <http://doi.org/10.3758/BF03202828>
- Wundt, W. (1883). *Philosophische Studien*. Leipzig: Wilhelm Engelmann.
- Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, 414(6861), 302–305.

## Supplementary Figures

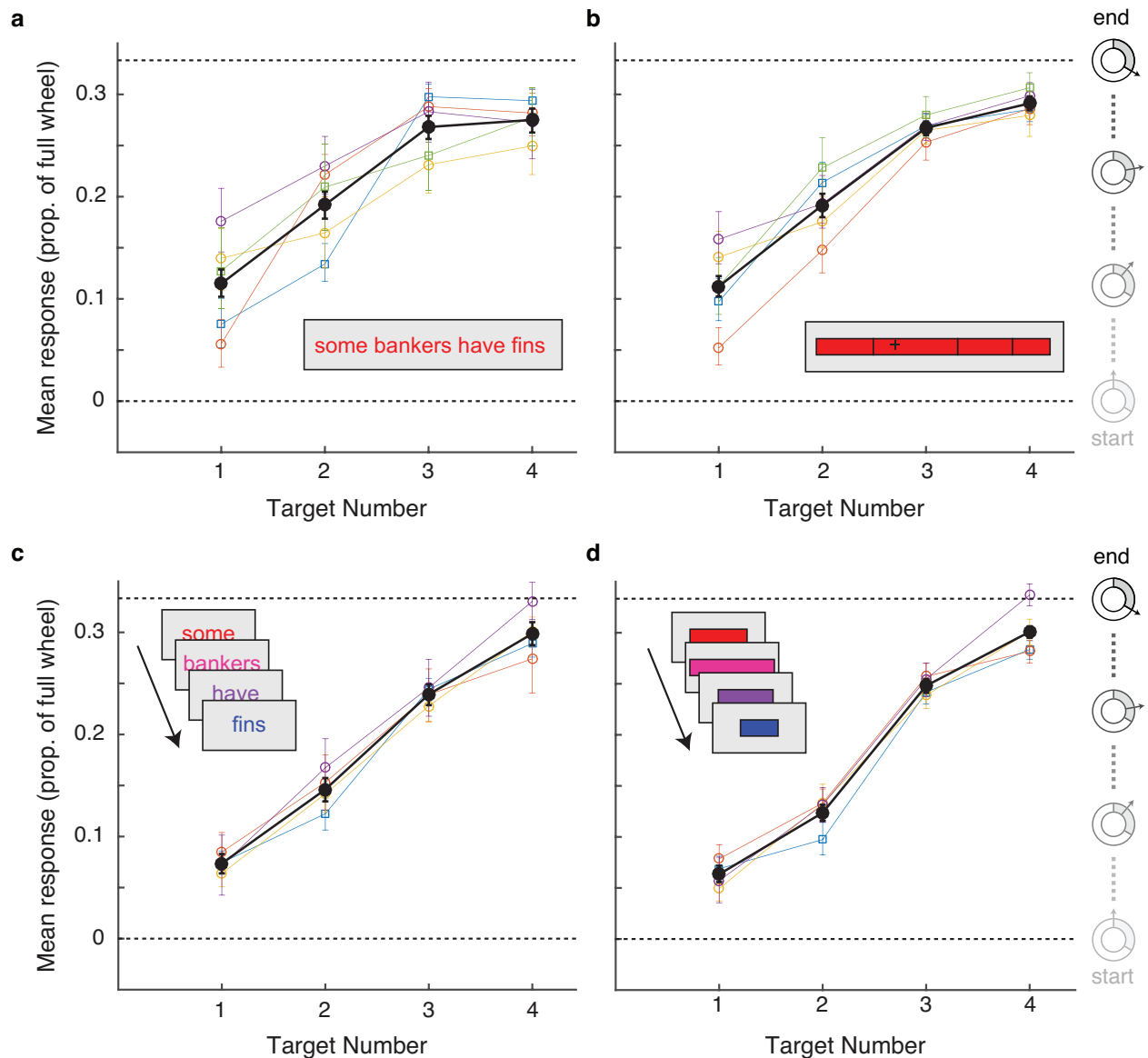


Figure S1. Mean responses relative to the beginning of each trial for each target word in Experiment 1 (a and b) and Experiment 2 (c and d), where a value of 0 represents the beginning of the trial, and a value of 1 represents one complete cycle. Observers viewed one third of a cycle on each trial. Responses are shown separately for the reading task in each experiment (a and c), and for the passive viewing task (b and d). Colored lines represent individual subjects (authors indicated by square symbols) and the black line shows the group average. Dashed lines at 0 and 1/3 mark the beginning and end of a trial, respectively. Error bars represent bootstrapped 95% confidence intervals.

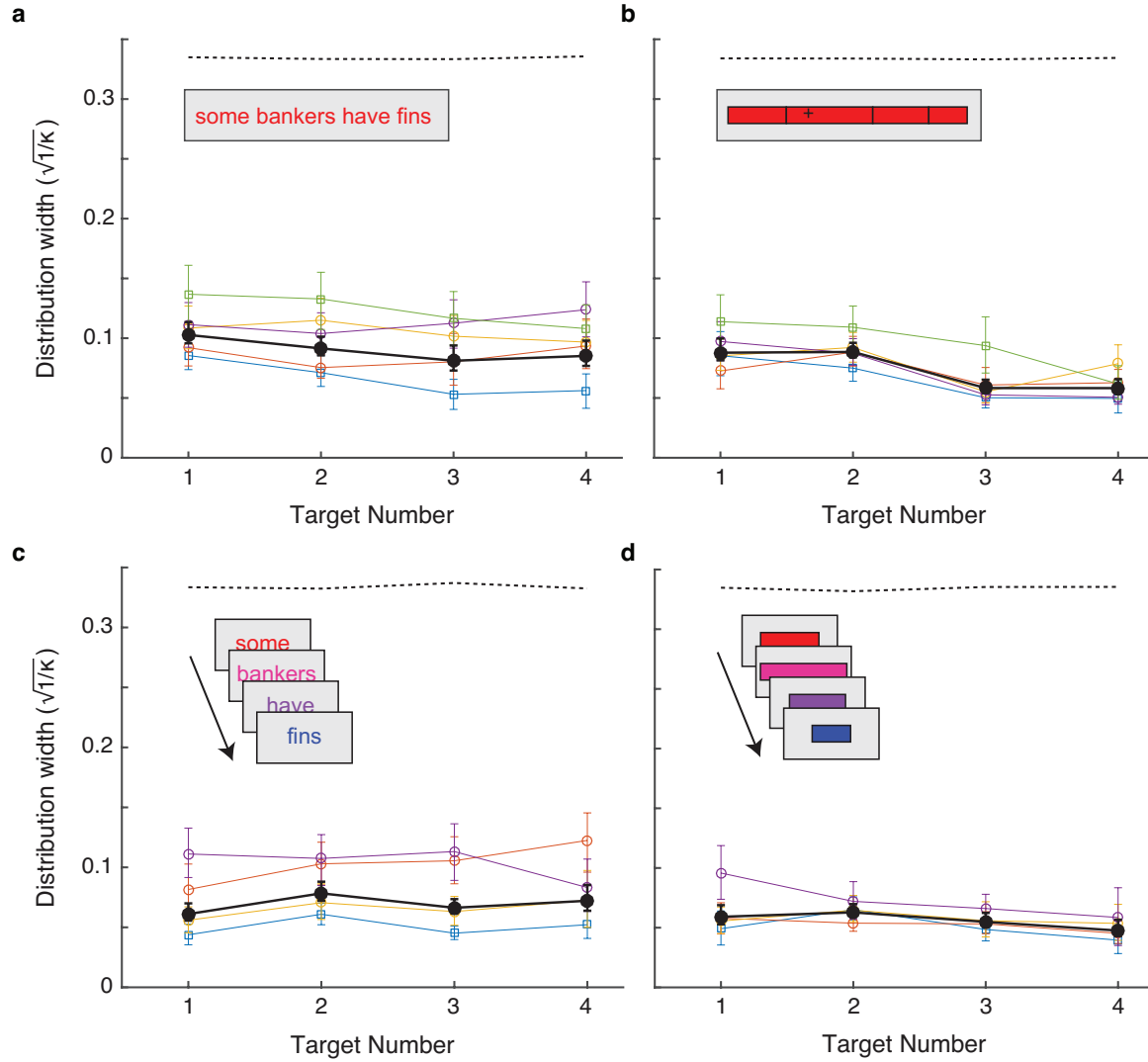


Figure S2. The widths of the distributions of subjects' responses relative to trial onset (see S1) for Experiments 1 and 2, in both the reading task (a and c) and the passive tracking task (b and d). Dashed lines show the expected distribution width based on chance performance (i.e., random guessing), and correspond to the mean of the permuted null distribution for each condition (see Experiment 1 Results). Distribution width is reported as  $\sqrt{\frac{1}{\kappa}}$ , which corresponds approximately to the standard deviation of the fitted von Mises distribution. Colored lines represent individual subjects and the black line shows the group average. Error bars represent bootstrapped 95% confidence intervals.

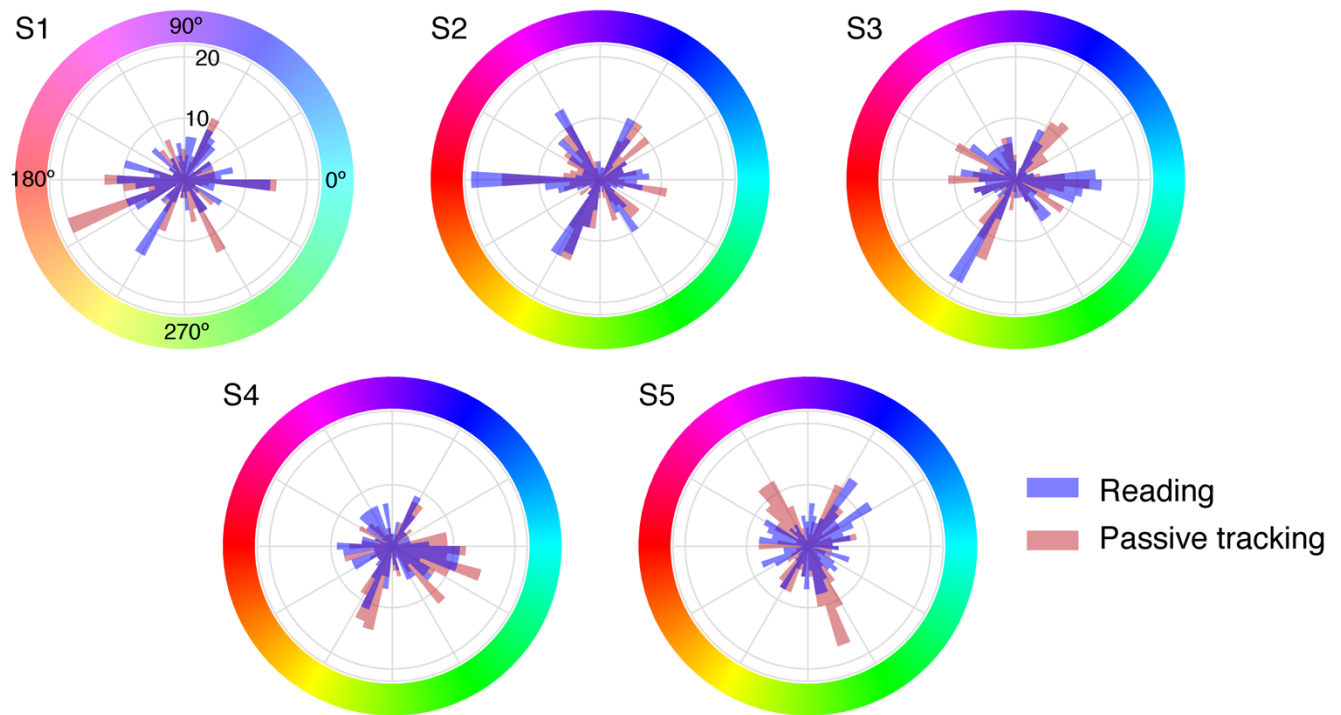


Figure S3. Distributions of reported colors for each observer (S1 through S5) in the reading and passive tracking conditions in Experiment 1 (blue and red, respectively; see observer S1 for axis tick labels). Responses generally reflect biases toward category centers, with some variability between individual observers. Since the starting point colors were randomly selected on each trial, we would expect these response distributions to introduce additional variance, but not systematic directional biases, when averaged over trials (see main text).

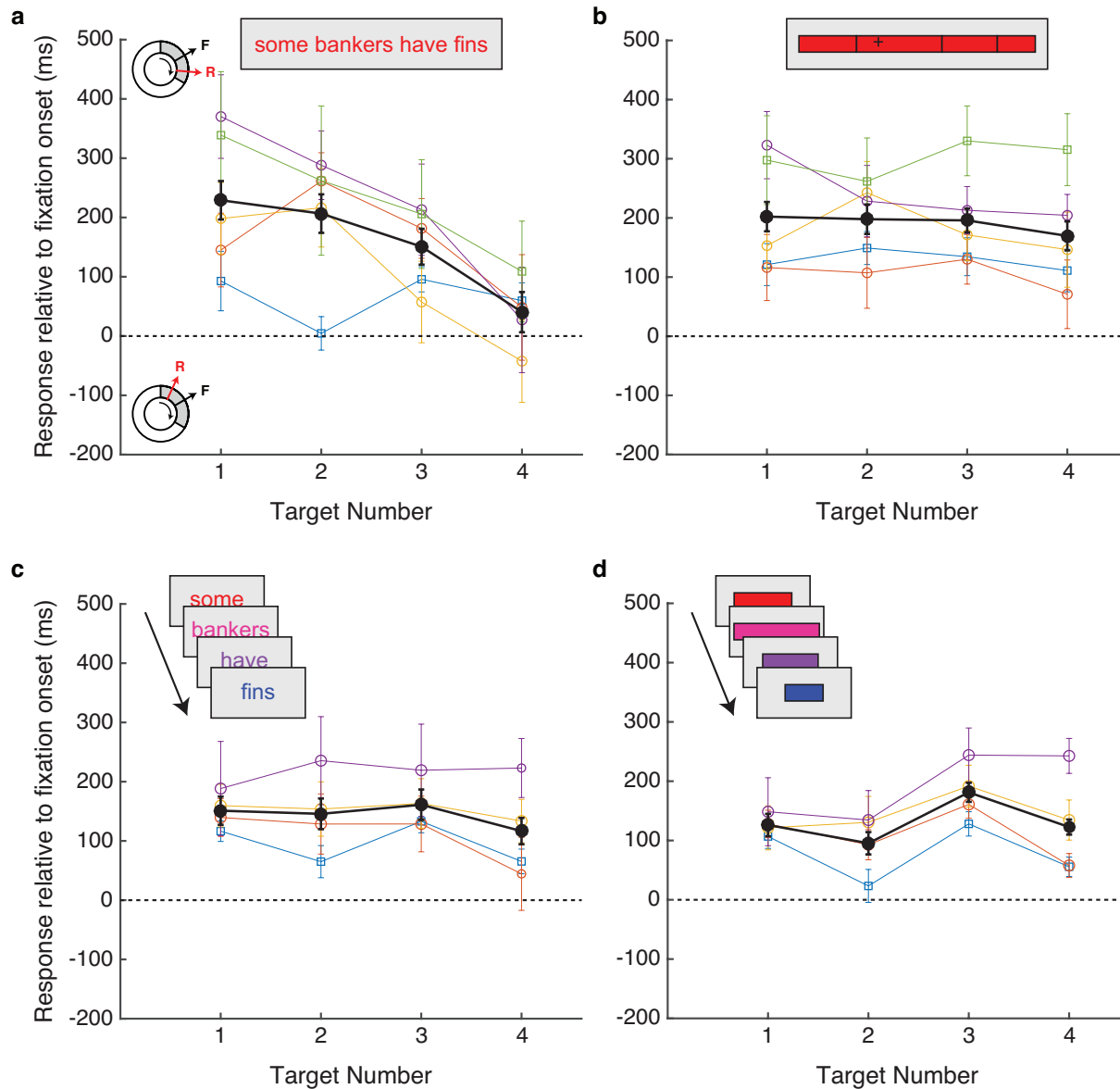


Figure S4. Mean responses relative to onset of the first fixation on the target word for each subject and each condition in Experiment 1 (panels a and b) and Experiment 2 (c and d). Positive values represent responses (R) that correspond to time points following fixation onset (F), and negative values represent responses prior to fixation onset. Colored lines represent individual subjects and the black line shows the group average. Error bars represent bootstrapped 95% confidence intervals.

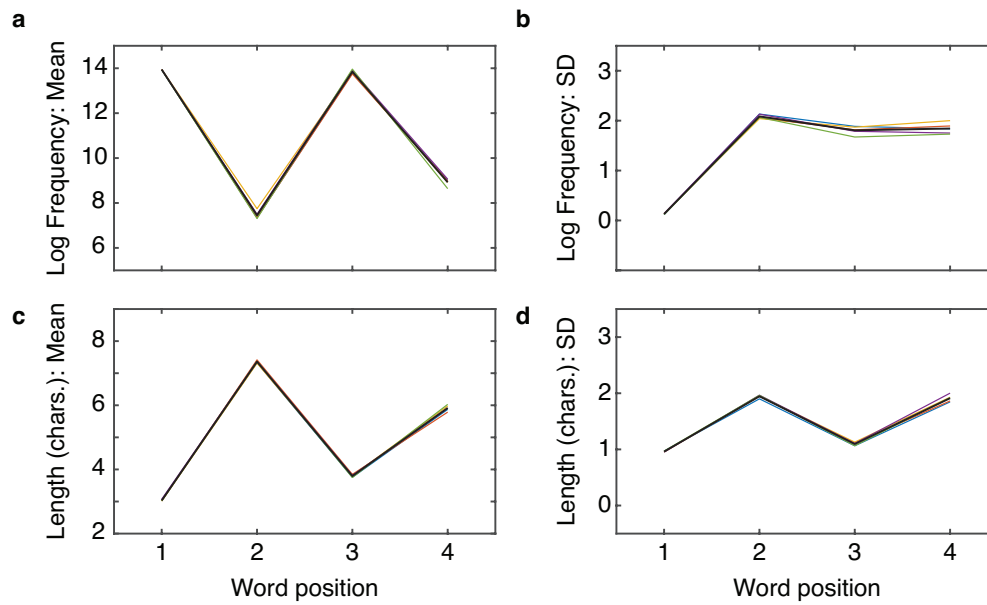


Figure S5. Word statistics as a function of word position in the sentence (1 through 4) in Experiment 1. (a) Mean log frequency across all trials, with colored lines representing individual observers, and black lines showing the group mean (note that many of the lines are overlapping). Word frequencies were extracted from the Hyperspace Analogue to Language (HAL) corpus (Lund & Burgess, 1996) using the English Lexicon Project (ELP) search tool online (Balota et al., 2007). (b) Within-subject standard deviation of log frequency across trials. (c) Mean word length (in characters) as a function of position in the sentence. (d) Within-subject standard deviation of word length across trials.

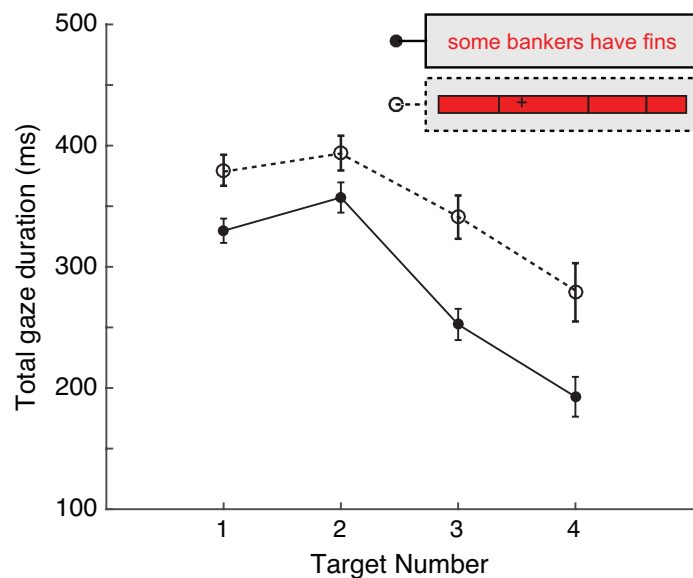


Figure S6. Average time spent fixating each target word in the reading (solid line) and passive tracking task (dashed line) in Experiment 1. For each trial, durations were summed across all fixations on the target word. Note that the gaze durations for each target number are calculated across different trials (therefore, the sum of gaze durations across the four targets is not equal to the total trial duration). Error bars represent bootstrapped 95% confidence intervals.

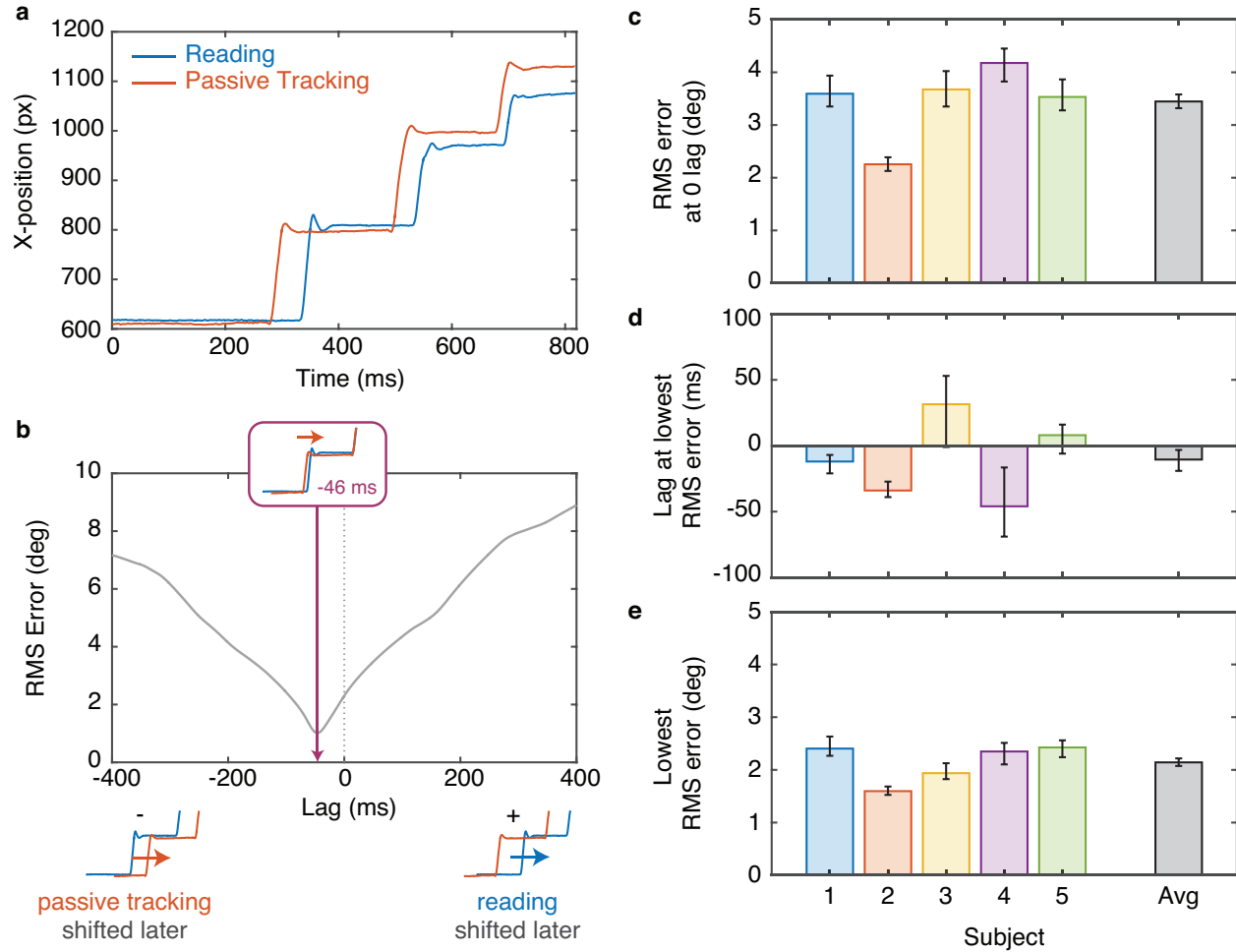


Figure S7. Tracking accuracy in passive tracking task. (a) Horizontal gaze position as a function of time for a sample trial in the reading task and its matched trial in the passive tracking task (blue and red, respectively). (b) Root-mean-square (RMS) error (in degrees) of the horizontal position traces for the trial shown in (a) as a function of temporal lag between the reading and passive tracking conditions. Positive lag values indicate that the reading trace is shifted later relative to the passive tracking condition. Negative values indicate that the passive tracking condition is shifted later. Therefore, a minimum at a negative lag value (in the example, -46 ms; purple arrow) indicates that the two traces are most similar when the passive tracking condition is delayed slightly relative to the reading condition. (c) Median RMS error between reading and passive tracing conditions across all trials, at a 0 ms temporal lag for each subject (subject colors matched to Figures S1-S4; group mean in gray). (d) Temporal lag corresponding to the lowest RMS error (e.g., purple arrow in (b)), calculated separately for each trial, and averaged across trials. Across subjects, the lowest RMS error corresponds to a lag of -10.5 ms (95% CI [-19.1, -3.2]). (e) Median RMS error at the minimum value, across all temporal lags, calculated individually for each trial. Error bars represent bootstrapped 95% confidence intervals



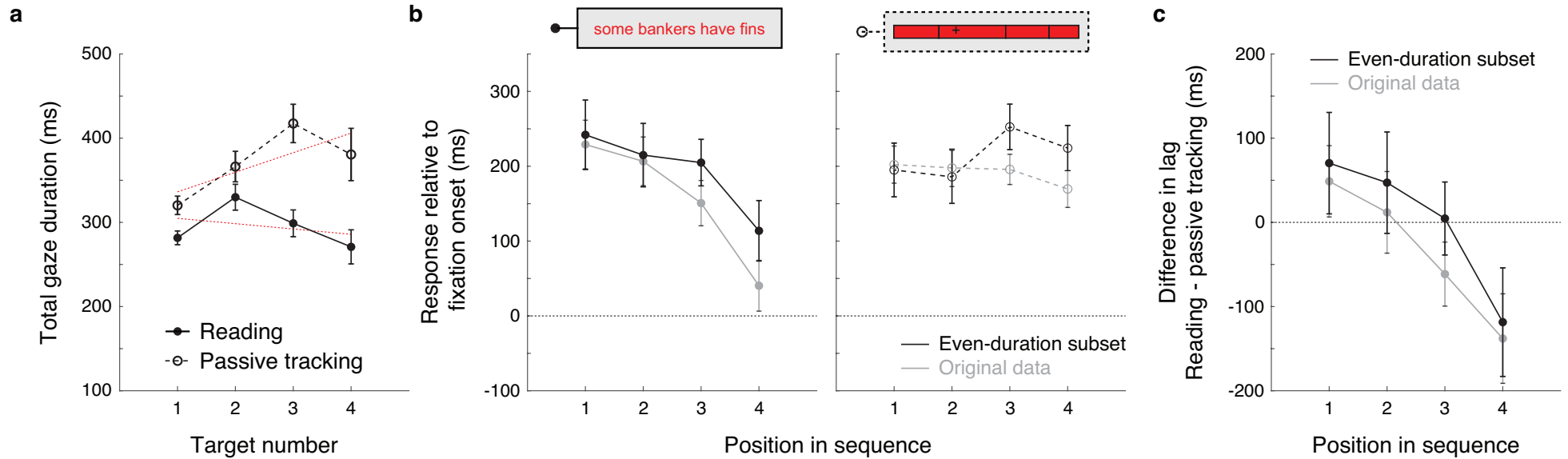


Figure S8. Re-analysis of results from Experiment 1. To examine whether the pattern of ‘flattening’ of response lags in Experiment 2 could be an artifact of having the full sentence duration divided equally among the four words, we analyzed a subset of trials from Experiment 1 that had an approximately even distribution of fixation durations among the four words. For each trial, we fitted gaze duration as a function of word position to a linear function, and analyzed only the trials with the upper 50% of slopes within each condition. (a) The average time spent fixating each target word for this selected subset of trials in Experiment 1. Solid and dashed lines show the reading and passive tracking tasks, respectively (red dotted line shows the linear fit for each condition). Compared to the main analysis in Experiment 1, gaze durations are distributed more evenly across the four words (compare to Figure S6). (b) Re-analysis of color reports relative to fixation onset for the subset of trials shown in (a). Left and right panels show the reading and passive tracking conditions, respectively. Gray lines show the full analysis (replotted from Figure 2 in the main text), and black lines show the subset of trials shown in (a). The difference in lag between the reading and passive tracking conditions for the two analyses is shown in (c). This subset of results from Experiment 1 still shows a decrease in lag for later words in the sentence, and a difference between the reading and passive tracking conditions (similar to the main analyses in Experiment 1). Therefore, it is unlikely that the even distribution of word durations in Experiment 2 is solely responsible for this pattern of flattening.

## Supplementary References

- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445–459.  
<http://doi.org/10.3758/BF03193014>
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2), 203–208.  
<http://doi.org/10.3758/BF03204766>