

Correspondence

Stable individual signatures in object localization

Anna Kosovicheva^{1,2,*}
and David Whitney^{1,3,4}

Perceptual processes in human observers vary considerably across a number of domains, producing idiosyncratic biases in the appearance of ambiguous figures [1], faces [2], and a number of visual illusions [3–6]. This work has largely emphasized object and pattern recognition, which suggests that these are more likely to produce individual differences. However, the presence of substantial variation in the anatomy and physiology of the visual system [4,7,8] suggests that individual variations may be found in even more basic visual tasks. To support this idea, we demonstrate observer-specific biases in a

fundamental visual task — object localization throughout the visual field. We show that localization judgments of briefly presented targets produce idiosyncratic signatures of perceptual distortions in each observer and suggest that even the most basic visual judgments, such as object location, can differ substantially between individuals.

To reveal this bias, observers (N = 5; 2304 trials each) reported the location of a brief (50 ms), stationary random dot noise patch shown at one of 48 random angular stimulus positions along an invisible isoecentric ring with a radius of 7 degrees of visual angle (d.v.a.; see Figure 1A and Supplemental Information). Subjects fixated the display center during the stimulus window, and then indicated perceived patch location by moving a cursor from the display center to the previously seen target location ('outward adjustment'), or by adjusting the position of a cursor constrained at an eccentricity of 7 d.v.a., starting from a random angular location ('angular adjustment').

In a separate session, to determine whether errors could be reproduced when the retinal location of the stimulus was dissociated from the retinal location of the cursor, subjects completed the outward adjustment method while moving their eyes freely during the response window. Finally, subjects completed a separate session in which they made a saccade as quickly as possible to the center of the target. For each of the four methods (see Figure 1C legend), we calculated the mean angular difference between the subject's response (or saccade landing location) on each trial and the angular location of the target center.

Figure 1B shows the errors from each observer in the outward adjustment response method at each location. Subjects' errors revealed large, idiosyncratic mislocalizations, up to 9.15° (1.11 d.v.a.) or 3.43 times the just-noticeable difference at a single location. Across all response methods, the average absolute angular deviation was 4.94° (0.60 d.v.a.). To evaluate between-observer mislocalization similarity for each

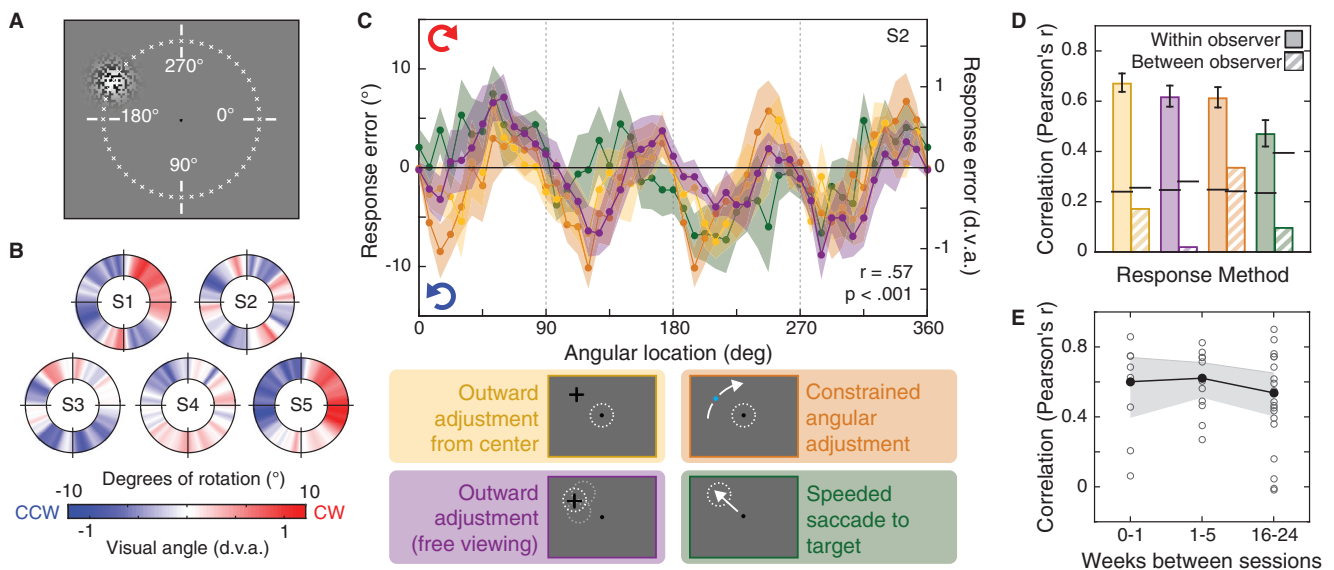


Figure 1. Stimulus arrangement and localization errors.

(A) Subjects viewed a 50 ms noise patch at one of 48 locations (indicated by 'x's, not visible to subjects) and reported its center using one of the methods shown in panel C. (B) Mean response errors at each location reveal substantial between-subject variations in error (outward adjustment method; red: clockwise errors, blue: counterclockwise errors). (C) Response errors from a representative subject at all locations show high within-observer consistency across the four methods used (white dotted circles show gaze location at time of response; positive values: clockwise errors, negative values: counterclockwise errors). (D) To quantify within-observer similarity (solid bars), errors for each response method were correlated with the other three methods within the same observer (see legend for panel C). Between-observer similarity for each method (hatched bars) was calculated by averaging all pairwise between-subject comparisons (horizontal bars: upper bound of the central 95% of the permuted null). (E) To determine within-observer stability over time, correlations between pairs of sessions within a subject were sorted into three bins based on temporal separation (open circles: individual pairs of sessions, filled circles: binned averages). Mean correlations were calculated from Fisher z values and transformed to Pearson's r. Error bars represent bootstrapped 95% confidence intervals.

method, we first calculated pairwise comparisons of errors at each of the 48 locations between observers — for example, Subject 1's error at the 90° location compared to Subject 2's error at 90°, and so on for each location — and then computed the average of all pairs of subjects. This analysis produced weak inter-observer correlations across the four response methods (Figure 1D), significant only in the angular adjustment condition ($p = 0.004$, all other p -values > 0.10 , permutation test using a Bonferroni-corrected alpha, $\alpha_B = 0.006$; see Supplemental Information).

In contrast, response errors within any individual observer were highly consistent across the four response methods (see Figure 1C for errors from a single subject). To quantify this degree of similarity for each condition, we correlated errors from one response method with the other three within an observer, and then averaged the resulting values. Average within-observer correlations for each method (see Figure 1D) were significantly greater than those expected by chance (all p -values < 0.001 based on permutation tests; $\alpha_B = 0.006$, see Supplemental Information). We also assessed the stability of each observer's localization signature over time by carrying out these measurements over the course of several months. Figure 1E shows the correlations between all pairs of sessions within an observer as a function of the length of time separating them (mean: 11.0 weeks, range: 0–24 weeks), sorted into three time bins. Mean correlations indicated a high degree of stability over time, with significant correlations within each time bin (all p -values < 0.001 ; $\alpha_B = 0.017$).

The stability of subjects' errors over time and across different types of adjustments suggests that they are unlikely to be a product of motor response biases. We further excluded the possibility of response bias in a second experiment, in which subjects reported patch position relative to a stable reference dot in a two-alternative forced choice task. Subjects' responses in this task indicated that the patch appeared aligned with the reference dot only when they were physically

misaligned, in a pattern consistent with their individual errors in the main experiment (see Figure S1).

If there are systematic localization errors, why is the perceived cursor position unaffected? The presence of identical perceptual shifts in the perceived location of the noise patch and cursor should cancel out any measurable error. One possibility is that these localization errors emerge under spatial or temporal uncertainty — for instance, when the noise patch is briefly presented or spatially diffuse. We tested this by measuring subjects' errors, varying both stimulus duration and size. When either spatial or temporal noise was reduced, such that the noise patch more closely resembled the cursor, the magnitude of the errors also decreased. Variations in patch size also shifted the pattern of errors, as indicated by reduced within-observer correlations across different patch sizes (see Figure S2).

Our finding of stable, idiosyncratic localization signatures overturns long-standing assumptions about perceptual judgments of basic visual attributes — that they are homogenous across individual observers, and invariant to retinal location within an observer. While it is often assumed that different observers generally agree about the locations of objects, our results demonstrate that this judgment can result in wildly different responses across individuals. Moreover, these errors were reproduced with saccadic responses, similar to correlations between perception and action observed in some illusions [6,9]. As saccadic responses and cursor adjustment responses occur on very different timescales, the observed errors are unlikely to be memory-driven, and we observe no correlation between reaction time and the magnitude of response errors (Figure S2). The stability of the errors observed suggests that these biases may be anatomically driven [10], similar to previously reported relationships between V1 anatomy and perceived size [4]. Further work will be needed to determine the anatomical locus of these errors, and to establish any relationship between these low-level biases and more cognitive, high-level effects [1].

SUPPLEMENTAL INFORMATION

Supplemental Information includes experimental procedures and two figures, and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2017.06.001>.

AUTHOR CONTRIBUTIONS

A.K. and D.W. conceived and designed the experiments. A.K. performed the experiments and analyzed the data. A.K. and D.W. wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH EY018216 and an NSF Graduate Research Fellowship to A.K. The authors thank Benjamin Wolfe for his feedback on earlier drafts of this article.

REFERENCES

1. Wexler, M., Duyck, M., and Mamassian, P. (2015). Persistent states in vision break universality and time invariance. *Proc. Natl. Acad. Sci. USA* 112, 14990–14995.
2. Afraz, A., Pashkam, M.V., and Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Curr. Biol.* 20, 2112–2116.
3. Grzeczowski, L., Clarke, A.M., Francis, G., Mast, F.W., and Herzog, M.H. (2017). About individual differences in vision. *Vision Res.* in press.
4. Schwarzkopf, D.S., Song, C., and Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nat. Neurosci.* 14, 28–30.
5. Wade, N.J. (1994). A selective history of the study of visual motion aftereffects. *Perception* 23, 1111–1134.
6. Morgan, M., Grant, S., Melmoth, D., and Solomon, J.A. (2015). Tilted frames of reference have similar effects on the perception of gravitational vertical and the planning of vertical saccadic eye movements. *Exp. Brain Res.* 2115–2125.
7. Andrews, T.J., Halpern, S.D., and Purves, D. (1997). Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract. *J. Neurosci.* 17, 2859–2868.
8. Curcio, C.A., Sloan, K.R., Packer, O., Hendrickson, A.E., and Kalina, R.E. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236, 579–582.
9. Melmoth, D., Grant, S., Solomon, J.A., and Morgan, M.J. (2015). Rapid eye movements to a virtual target are biased by illusory context in the Poggendorff figure. *Exp. Brain Res.* 233, 1993–2000.
10. Kanai, R., and Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nat. Rev. Neurosci.* 12, 231–242.

¹Department of Psychology, University of California, Berkeley, Berkeley, CA 94720, USA. ²Department of Psychology, Northeastern University, Boston, MA 02115, USA. ³Vision Science Program, University of California, Berkeley, Berkeley, CA 94720, USA. ⁴Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA.

*E-mail: akosov@neu.edu

Supplemental Figures

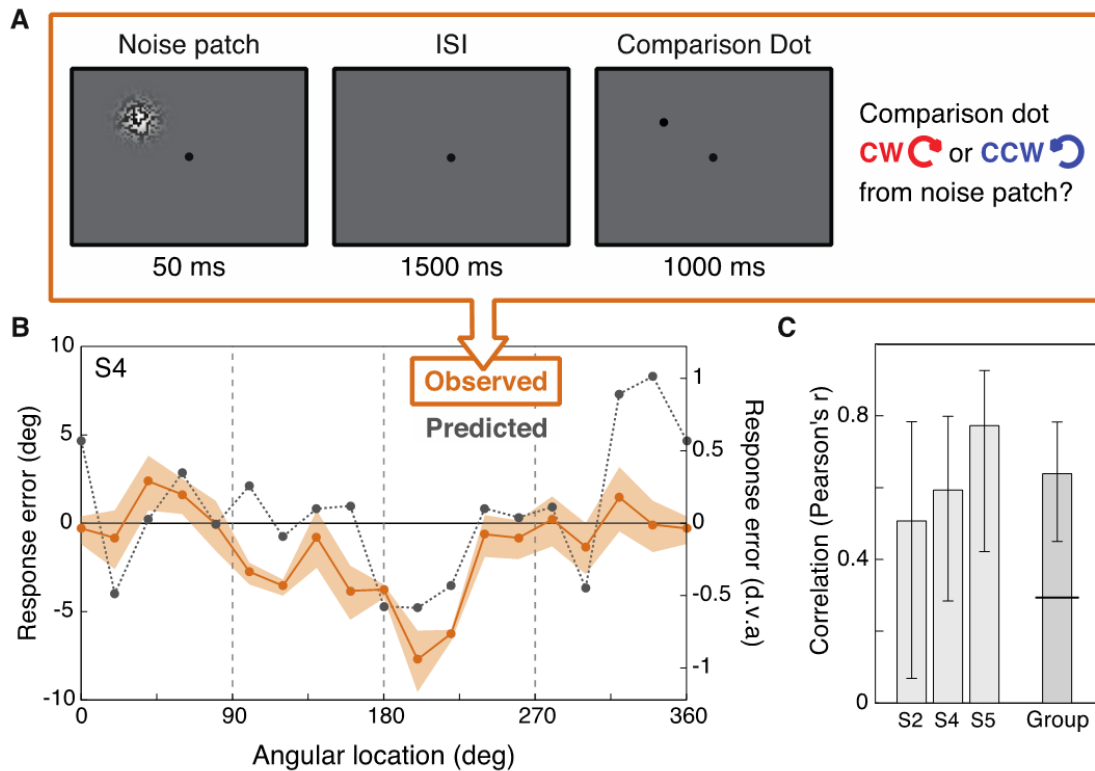


Figure S1. Stimuli and response errors in Experiment 2. (A) Subjects viewed a brief noise patch (50 ms), followed by a 1500 ms ISI, which was then followed by a 1000 ms comparison dot, the position of which was controlled by a 1-up, 1-down staircase (see Supplemental Methods). Following the offset of the comparison dot, subjects were instructed to indicate, by pressing one of two keyboard keys, whether the comparison dot was clockwise or counterclockwise relative to the center of the noise patch. Trials were separated by a 1000 ms ITI. (B) Observed localization errors from 2AFC responses (determined by the averaging dot position at the final 6 reversal points) were compared to the predicted localization errors from the same subject's method-of-adjustment responses in Experiment 1. (C) Response errors from the 2AFC task were similar to those predicted from method-of-adjustment responses, with a significant positive correlation across observers (horizontal line denotes upper bound of permuted null distribution, mean $r = 0.64$, $p < 0.001$). Subject labels correspond to subject number assignments in Experiment 1 (see Fig. 1B). Error bars indicate bootstrapped 95% confidence intervals.

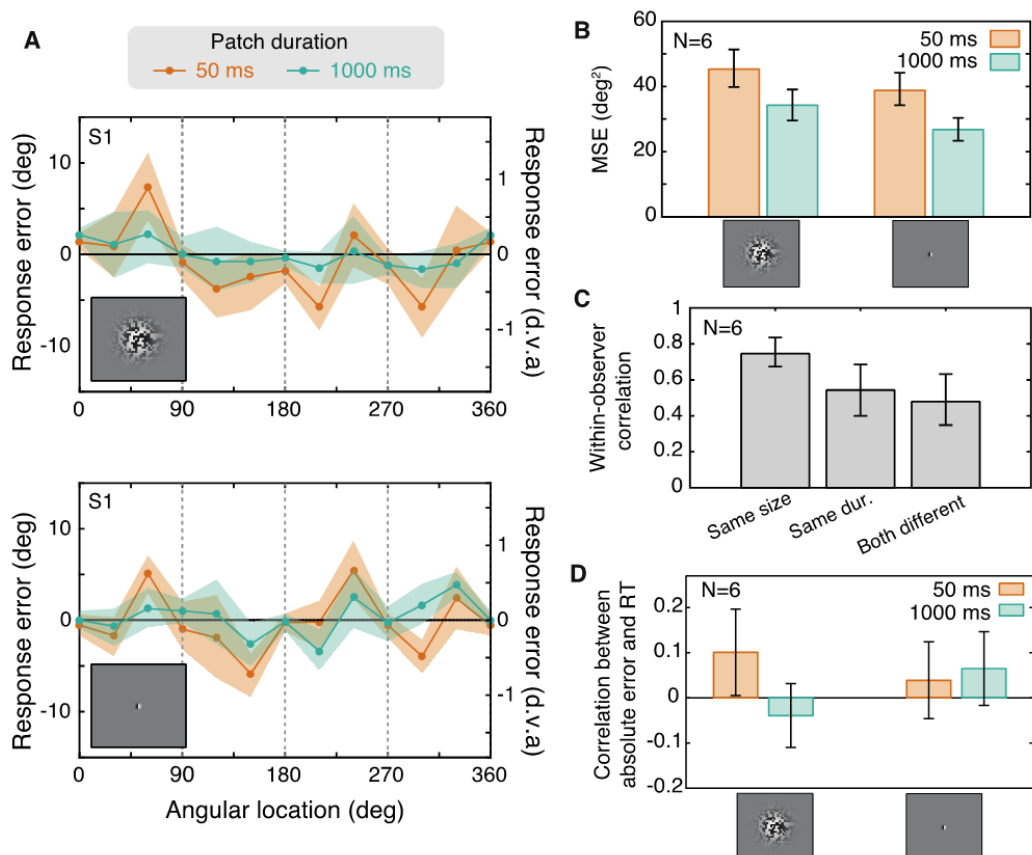


Figure S2. Response errors in Experiment 3. (A) In Experiment 3, observers viewed patches that varied in duration (50 or 1000 ms) and size (aperture SD of 1.0 or 0.1 d.v.a.; inset in upper and lower panels, respectively), interleaved across trials. Average response errors were calculated for each location for each observer within the four size and duration conditions, where positive values correspond to clockwise errors, and negative values correspond to counterclockwise errors. Upper and lower panels show example data from Subject 1. (B) To evaluate the effects of stimulus properties on the magnitude of subjects' errors, mean squared errors (MSEs) were calculated from individual trials for each subject in each condition. MSEs were larger in the 1.0 d.v.a. patch condition compared to the 0.1 d.v.a. condition ($p = 0.004$; bootstrap test, $\alpha_B = 0.0167$), and larger in the 50 ms condition compared to the 1000 ms condition ($p < 0.001$). The effect of duration was not significantly different between the two patch size conditions ($p = 0.83$). (C). To determine the effects of stimulus properties on the pattern of errors, we computed all six possible pairwise correlations between the response errors shown in (A) within each observer. These within-observer correlations were grouped into three categories, shown left to right: same-size/different-duration; same-duration/different-size; different-duration/different-size. Correlations were highest when the patch size had the same size, but varied in duration ($p = 0.002$ compared to the same-duration/different-size condition, and $p < 0.001$ compared to the different-both condition; bootstrap test; $\alpha_B = 0.0167$). Correlations between the same-duration/different size and different-both conditions were not significantly different ($p = 0.53$). (D) We tested whether response biases increased with increasing time from stimulus offset, as observed in memory-based errors [S7] or time-order errors (TOEs) [S8,S9]. Correlations between response time (RT) and the absolute magnitude of the response error were not significantly different from zero at any combination of size or duration (bootstrap test; $\alpha_B = 0.0125$; left to right: $p = 0.04$ with 1.0 d.v.a./50 ms; $p = 0.25$ with 1.0 d.v.a./1000 ms; $p = 0.38$ with 0.1 d.v.a./50 ms; $p = 0.13$ with 0.1 d.v.a./1000 ms). Error bars correspond to bootstrapped 95% confidence intervals.

Supplemental Experimental Procedures

General Method (Experiment 1)

Participants

Five observers (three female; mean age: 25.6, range: 21-28), including one author (AK), participated in the main experiment. Subjects were experienced psychophysical observers, and all except the author were naïve to the purpose of the experiment. All subjects reported normal or corrected-to-normal vision and gave informed consent prior to participating. Procedures were approved by the Institutional Review Board at UC Berkeley and conducted in accordance with the Declaration of Helsinki.

Stimuli and Procedure

Stimuli were presented on a gamma-corrected 19" Samsung Syncmaster 997DF CRT monitor and run on an iMac computer (Apple, Inc., Cupertino, CA). The experiment was programmed in Matlab (The MathWorks, Inc., Natick, MA) using the Psychophysics Toolbox [S1,S2]. Display resolution was set to 1024 × 768 and the refresh rate to 100 Hz. Subjects viewed the display binocularly at a distance of 57 cm from the monitor and a chinrest was used to minimize head motion. Where reported in the methods and results, "d.v.a." refers to degrees of visual angle, and "degrees" (°) refers to degrees of rotation (i.e., from 0° to 360°).

Stimuli were presented on a gray background (41.6 cd/m²). At the beginning of each trial, subjects were instructed to fixate a black (0.75 cd/m²) 0.31 d.v.a. diameter dot presented at the center of the screen. After 1000 ms, a noise patch appeared for 50 ms at one of 48 angular positions along a 7 d.v.a. invisible isoeccentric ring. Noise patches consisted of a grid of squares, each measuring 0.21 × 0.21 d.v.a., inside a two dimensional Gaussian contrast aperture with a standard deviation of 1 d.v.a. Each square was either black or white with equal probability, and the peak contrast of the aperture was 100%. The set of possible angular positions were at evenly spaced 7.5° angular intervals and included the 4 cardinal locations—directly to the right, below, to the left, and above fixation, corresponding to 0, 90, 180, and 270°, respectively. Figure 1A shows the set of 48 possible angular locations, superimposed on a display with a single noise patch.

Following the offset of the noise patch, the fixation dot changed to dark gray (14.5 cd/m²), and after 500 ms, subjects were shown a response screen, in which they were instructed to report the location of the noise patch using one of the methods described below. Each response method was completed in a separate experimental session lasting approximately 45 minutes. Subjects completed the sessions in the same order (speeded saccades, angular adjustment, outward adjustment, outward adjustment with free-viewing; see descriptions below). In each session, observers completed 12 trials at each of the possible 48 patch locations (576 trials in total). The order of trials within each session was randomly drawn without replacement.

Response Methods

Outward adjustment. On the response screen, subjects were shown a 0.45 × 0.45 d.v.a. black crosshair surrounded by a white outline, directly on top of the fixation dot at the center of the display. Observers were able to move the crosshair horizontally and vertically using the mouse to any location on the display, and were instructed to match the position of the center of the noise patch, while maintaining fixation on the dot in the center of the screen. After making their response, observers proceeded to the next trial by clicking the mouse.

Outward adjustment with free-viewing. The procedure was identical to the one described above, except observers were instructed to move their eyes to track the crosshair, as they normally would when using a computer mouse. Once they made their response, subjects were instructed to saccade back to the fixation dot at the center of the display in preparation for the next trial and maintain fixation while the noise patch was presented.

Constrained angular adjustment. Subjects were shown a 0.31 d.v.a. diameter blue (8.9 cd/m²) adjustment dot at a random angular location at an eccentricity of 7 d.v.a. While maintaining fixation at the center of the display, subjects moved a mouse to adjust the angular position of the blue dot to match the location of the noise patch. Unlike the two response methods described above, the eccentricity of the adjustment dot was always constrained to 7 d.v.a. In other words, subjects were only able to move the dot clockwise or counterclockwise along an invisible ring to match the perceived location of the patch. The blue adjustment dot was assigned a new, randomly selected angular location at the beginning of the response screen on every trial.

Speeded saccades. The experiment was run on an identical testing setup, with the exception of the computer (Mac Mini; Apple, Inc., Cupertino, CA). Gaze position was recorded with a desktop eye tracker (see Eye Tracking). The presentation of the noise patch was preceded by a 500 – 1000 ms fixation interval (selected randomly on each trial). During this interval, subjects were required to maintain fixation continuously before the patch was

presented. If the eye deviated more than 1 d.v.a. horizontally or vertically from the center of the fixation dot, the counter would restart. Subjects were instructed to saccade as quickly as possible to the noise patch as soon as it appeared. If eye position was still within 1 d.v.a. from the center of the fixation dot 500 ms after the onset of the noise patch, subjects were given feedback in the form of a red fixation dot (26.1 cd/m²), and a brief (98 ms; 651.9 Hz) tone. Otherwise, the fixation dot changed to dark gray (25.8 cd/m²) until the subject re-fixated the dot in preparation for the next trial. The next trial began following a 1000 ms intertrial interval (ITI). As with the other response methods, subjects completed 12 trials for each of the 48 locations. Sessions were divided into 2 blocks of 288 trials each, and eye position was re-calibrated after each block.

Eye Tracking

Eye movements were recorded with an EyeLink 1000 desktop mounted infrared eye tracker (SR Research Ltd., Mississauga, Ontario, Canada), used together with the EyeLink Toolbox for Matlab [S3]. Gaze position was recorded from the right eye of each subject at a sampling rate of 1000 Hz. Prior to each block of trials, subjects completed a 13-point calibration procedure (mean error on validation: 0.28 d.v.a.).

To reduce noise artifacts, a heuristic filtering algorithm was applied to the raw gaze position samples (see [S4] for details). Gaze information was then parsed into saccades, fixations, and eye blinks. The first time point at which the velocity exceeded 30 d.v.a./s and the acceleration exceeded 8000 d.v.a./s² was used to delineate the beginning of a saccade. In addition, the onset of a saccade required a minimum eye motion of 0.15 d.v.a. Time points at which the velocity and acceleration fell below their respective thresholds marked the end of each saccade. Finally, to compensate for slow drifts in x- and y- gaze position across trials, the x- and y- gaze coordinates during fixation (i.e., coordinates immediately preceding the saccade) were first fitted to a third-degree polynomial across the set of trials within a session. Then, we estimated the drift by calculating the difference between the average x- and y- fixation coordinates for the first 20 trials and the fitted value for each trial. The resulting drift values for each trial were subtracted from the observed x- and y- gaze coordinates.

Data Analysis

For each of the cursor response methods, we calculated the mean angular difference between the subject's response on each trial and the angular location of the center of the noise patch (see Figures 1B and 1C for examples). Positive values corresponded to a clockwise angular error, and negative values corresponded to a counterclockwise angular error. To exclude lapses, trials in which the angular deviation was more than four standard deviations from the center of the noise patch were removed from the analysis (0.15% of all trials). The just-noticeable-difference (JND) was estimated by calculating half the interquartile range of subjects' response errors [S5]. First, we calculated the overall standard deviation by averaging the response variance at individual stimulus locations, and then multiplied the SD by 0.6745 (the x-value at which the cumulative normal distribution is equal to 75%). The overall just-noticeable difference ranged from 2.11° to 3.38° (0.26 to 0.41 d.v.a) across observers.

For the eye tracking data, the horizontal and vertical position of the endpoint of the first large saccade (minimum amplitude 1.7 d.v.a.) on each trial were used to determine the saccade landing location. Similar to the cursor responses, we calculated error as the angular difference between the saccade landing location and the noise patch center. We excluded trials in which the saccade latency was greater than 500 ms or in which the angular error of the saccade landing location was more than four standard deviations from the center of the noise patch. This resulted in the exclusion of 1.8% of trials from the analysis.

Mean angular error was computed at each of the 48 possible locations. Unless otherwise noted, confidence intervals were estimated using a nonparametric bootstrap procedure [S6]. For each observer and for each of the 48 patch locations, individual trial errors were resampled with replacement and then averaged. This procedure was repeated 1000 times to estimate a 95% confidence interval. Where reported, average correlations were calculated by transforming individual Pearson's r values to Fisher z values, computing the average, and then back-transforming the values to Pearson's r.

Permutation Test Procedure

To determine the similarity in response error between subjects, we computed the correlations between pairs of observers, based on the errors at each of the 48 locations (e.g., Subject 1's error at 90° was compared to Subject 2's error at 90°, etc). We computed all possible pairwise correlations between the five observers and then calculated the mean of these pairwise correlations. The resulting value (e.g., 0.17 for the outward adjustment method) was compared to a permuted null distribution that would be expected if observers' errors were uncorrelated with each other. To generate this distribution, on every iteration, each observer's set of errors across all angular locations (see Figure 1C) was assigned one of 48 random phases from 0 to 360°, randomizing the position of the curve while

leaving adjacent errors intact. This procedure randomized observers' responses relative to each other while simultaneously preserving (within each subject), the relationship between the response errors at adjacent locations. Then, we recomputed all possible pairwise correlations between observers, and calculated the average of the correlations. For each response method, the permuted null distribution consisted of the set of correlation values from 10,000 iterations.

To estimate similarity in response error within observers, we correlated the errors from one response method with those from the remaining three response methods for each observer (e.g., method 1 vs. method 2, 1 vs. 3, and 1 vs. 4). The values for each response method (shown in Figure 1D) were compared to a permuted null distribution generated with a similar procedure to the between-subjects analysis. On each iteration, we assigned a random phase to each observer's pattern of response errors (independently selected for each of the four response methods), and then calculated all pairwise correlations between the response method and the remaining three methods. We then averaged the values within an observer, and then across observers. This procedure was repeated for 10,000 iterations to produce a null distribution of within-subject correlation values for every response method. The resulting p-values were compared to a Bonferroni-adjusted alpha for eight comparisons (four within-observer, and four between-observer correlations, $\alpha_B = 0.006$).

Analysis by Time

Within each observer, we calculated the length of time separating all possible pairs of sessions. On average, pairs of runs were separated by 10.4 weeks, ranging from 0 to 21.6 weeks. Two observers also completed an additional session of the constrained angular adjustment method, increasing the mean to 11.0 weeks, and the range to 0-24 weeks. Based on the temporal separations between pairs of runs, data were sorted into one of three bins: less than 1 week, 1-5 weeks, and 16-24 weeks. Figure 1E shows the Pearson's r values for every pair of sessions within an observer, sorted into the three time bins. The averaged values within each bin were compared to a permuted null distribution generated with the procedure described above—a null distribution was generated individually for each pair, and then averaged within each time bin. Permutation tests were performed using a Bonferroni-adjusted alpha for three comparisons ($\alpha_B = 0.017$).

Two-alternative Forced Choice (Experiment 2)

Participants

Three observers who completed Experiment 1 also participated in Experiment 2 (mean age: 26.3, range: 25-27).

Stimuli and Procedure

Stimuli and procedure (see Figure S1), were identical to Experiment 1 with the following exceptions. On each trial, subjects were shown a noise patch for 50 ms (see Figure S1A), followed by a 1500 ms ISI, which was followed by a black comparison dot (0.31 d.v.a. diameter) shown for 1000 ms. Following the offset of the comparison dot, subjects were instructed to press one of two keyboard keys to indicate whether the comparison dot was clockwise or counterclockwise relative to the center of the previously presented noise patch.

The use of a small, stable reference dot as our comparison stimulus reduces position uncertainty, and allows us to differentiate between two explanations for errors observed in Experiment 1. If a brief, spatially distributed stimulus is necessary to produce the errors observed in Experiment 1, then the comparison dot will appear aligned with the center of the noise patch when the two are physically offset, consistent with the errors reported in Experiment 1. In contrast, if the errors observed in Experiment 1 are simply due to a bias in subjects' adjustments, the comparison dot will appear aligned with the noise patch only when they are physically in the same location.

Due to the greater number of trials required to produce an estimate of perceived location with 2AFC responses, noise patches were shown at one of 18 (rather than 48) angular locations at an eccentricity of 7 d.v.a. Noise patch locations ranged from 0° to 360° in steps of 20°, including locations immediately to the right and left of fixation (0 and 180°, respectively). The position of the comparison dot on each trial was controlled by one of 36 interleaved staircases (two staircases per noise patch location). To determine the position at which the comparison dot appeared to match the center of the previously shown noise patch, each staircase was controlled by a 1-up, 1-down rule, with a fixed step size of 0.5° (0.06 d.v.a.). One of the two staircases at each location started 3° (0.37° d.v.a.) clockwise relative to the center of the noise patch, and the other started 3° counterclockwise relative to the center.

The experiment terminated once each of the 36 staircases had completed 8 reversals. To ensure that the staircases would terminate at approximately the same time, the staircase (and therefore the patch location) on each trial was selected pseudorandomly, weighted by the number of reversals remaining. In other words, on a given trial, a staircase with 7 reversals remaining had a greater probability of being selected than a staircase with 3 reversals remaining to completion, and completed staircases were assigned a weight of 0 (i.e., they were no longer selected). On average, subjects completed the experiment after 680 trials (range: 639 – 730 trials).

Data Analysis

Position estimates for each of the 36 staircases were calculated by averaging the dot location (relative to the center of the noise patch) at the last 6 reversal points. The resulting estimates across pairs of staircases at each patch location showed a high degree of internal consistency (mean $r = 0.87$, range across subjects: 0.85-0.89), and were therefore averaged to obtain 18 estimates of perceived location, one for each noise patch location (see Figure S1B for an example subject).

Estimates of perceived location using 2AFC responses in Experiment 2 were compared to the perceived locations for each subject in Experiment 1 using the constrained angular adjustment method (the most similar method to the 2AFC task, as it requires a clockwise vs. counterclockwise judgment of the position of a peripheral dot relative to the noise patch). However, as the 18 patch locations tested in Experiment 2 did not perfectly correspond to the 48 locations in Experiment 1, we calculated predictions of localization error at the set of 18 locations from each subject's response errors in Experiment 1 through linear interpolation of neighboring points (dotted line in Figure 3B).

For each observer, we calculated the correlation between the perceived position estimates based on subject's 2AFC responses and the predicted localization error from Experiment 1, and calculated the average across observers (Figure S1C). As in Experiment 1, the observed mean was compared to a null distribution generated by randomizing the phases of subjects' response errors and recalculating the correlation.

Effects of Stimulus Properties on Localization Errors (Experiment 3)

Participants

Six observers participated in Experiment 3 (4 female, mean age: 22.2, range: 18-30). Two subjects were experienced psychophysical observers (including one author) who had previously completed Experiment 1, and the remainder were undergraduate students receiving course credit for their participation. All subjects reported normal or corrected-to-normal vision and gave informed consent prior to participating.

Stimuli and Procedure

Stimuli and procedure were identical to those used in the constrained angular adjustment procedure in Experiment 1, with the following exceptions. Stimuli were presented at a viewing distance of 48 cm on a gamma-corrected 27" BenQ XL2720Z LED monitor controlled by a Dell Optiplex 9020 desktop computer with a Quadro K420 graphics card. The display resolution was set to 1920 × 1080 and the refresh rate to 120 Hz.

On each trial, subjects were shown a noise patch with a pseudorandomly selected duration (50 ms or 1000 ms) and size (Gaussian contrast envelope SD of 1.0 or 0.1 d.v.a.; see insets in Fig S2A). The combination of the short stimulus duration (50 ms) and large patch size (SD of 1.0 d.v.a.) reproduced the stimulus parameters used in Experiments 1 and 2. The smaller patch size was comparable to the size of the cursor dot in Experiment 1 (full envelope width of approximately 0.4 d.v.a.). The noise patch on each trial was centered on one of 12 possible angular positions (evenly spaced in 30° intervals, including the 4 cardinal locations). As before, following a 500 ms inter-stimulus interval (ISI), subjects adjusted a blue response dot (constrained at 7 d.v.a.) to match the perceived location of the patch while maintaining fixation at the center of the display.

The addition of a long stimulus duration (1000 ms) introduced the possibility of subjects fixating the noise patch while it was visible on the screen. As demonstrated in Experiment 1, subjects' mislocalizations are robust to eye movements during the response phase; however, variations in the retinal location of the stimulus could influence the pattern of localization errors. Therefore, gaze-contingent feedback was introduced to monitor fixation compliance during the experiment. On each trial, if the observer's gaze position moved outside an invisible 3 × 3 d.v.a. screen-centered box, or any eye blinks occurred at the time the noise patch was onscreen, it was immediately removed. On these trials, the entire screen was filled with a red box (69.8 d.v.a. width × 39.6 d.v.a. height) for the remainder of the patch duration, and subsequent adjustment responses were removed from the analysis (5.6% of all trials). In addition, subjects were instructed to fixate the center of the display while adjusting the dot position, and were prevented from making any adjustments while fixating outside the same screen-centered box. Whenever gaze

position deviated from this region, the adjustment dot remained in place regardless of any mouse movements, and was changed to from blue to red (to indicate that it was “locked”). As before, observers proceeded to the next trial by clicking the mouse. Following a 500 ms delay, the presentation of the next noise patch was withheld until the subject had continuously maintained gaze position in the same central region for 500 ms.

Observers completed a total of 480 trials, consisting of 10 trials for every unique combination of stimulus location (12 angular positions), patch duration (50 or 1000 ms), and envelope size (SD of 1.0 or 0.1 d.v.a.), presented in a random order. Trials were divided into five blocks of 96 trials each, and subjects were recalibrated at the beginning of each block.

Eye Tracking

Eye tracking procedures were the same as those described in Exp. 1, with the exception that gaze position was initially calibrated binocularly, and the eye with the lowest average error on validation was used for gaze-contingent control of the display.

Data Analysis

As before, trials in which the angular response error was greater than four standard deviations from the center of the noise patch were removed from the analysis. These accounted for 0.66% of all trials, resulting in the exclusion of a total of 6.25% trials (including trials in which the subject’s point of gaze deviated from fixation).

We investigated the effects of stimulus properties (patch duration and size) on both the magnitudes and the patterns of errors produced by individual subjects. Figure S2A shows the mean response errors for one subject across the set of 12 patch locations for each size and duration condition. To compare the magnitude of localization errors between the four combinations of stimulus size and duration, we first calculated the mean squared error (MSE) of individual trials for each subject in each condition. The MSE serves as an estimate of goodness-of-fit to a perfect accuracy model, where the signed error is 0° across all locations. Figure S2B shows the mean MSEs across the six subjects for the two size and duration conditions. Effects of size and duration were separately evaluated using a Bonferroni-adjusted alpha for three comparisons ($\alpha_B = 0.017$).

To determine the effects of stimulus properties on the pattern of signed (i.e., clockwise and counterclockwise) errors, we calculated the correlations between mean response errors across the set of 12 locations between stimulus conditions within each observer. Correlations were calculated for all six pairwise combinations of the four unique stimulus conditions (e.g., 1.0 d.v.a./50 ms vs. 1.0 d.v.a./1000 ms; 1.0 d.v.a./50 ms vs. 0.1 d.v.a./50 ms, and so on). The resulting six correlation values were grouped into one of three categories by averaging two values per category (shown in Figure S2C): (1) same size and different duration, (2) same duration and different size, and (3) different size and different duration. Pairwise comparisons between the three correlations were performed using bootstrap tests with a Bonferroni-adjusted alpha for three comparisons ($\alpha_B = 0.017$).

Finally, to determine whether responses were influenced by memory [S7] or by time-order errors (TOEs) [S8,S9], correlations were calculated between reaction time and absolute response error on individual trials, within each size and duration condition (Figure S2D). Across subjects, the central 95% of reaction times fell between 725 and 2976 ms. Correlations in each size and duration condition were separately evaluated using bootstrap tests (1,000 iterations) with a Bonferroni-adjusted alpha for four comparisons ($\alpha_B = 0.0125$).

Supplemental References

- S1. Brainard, D.H. (1997). The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436.
- S2. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- S3. Cornelissen, F.W., Peters, E.M., and Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods, Instruments, Comput.* 34, 613–617.
- S4. Stampe, D.M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behav. Res. Methods, Instruments, Comput.* 25, 137–142.
- S5. Woodworth, R.S., and Schlosberg, H. (1954). *Experimental psychology* (Oxford and IBH Publishing).
- S6. Efron, B., and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap* (London: Chapman & Hall).
- S7. Sheth, B.R., and Shimojo, S. (2001). Compression of space in visual memory. *Vision Res.* 41, 329–341.
- S8. Hellström, Å. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychol. Bull.* 97, 35–61.
- S9. Allan, L.G. (1977). The time-order error in judgments of duration. *Can. J. Psychol.* 31, 24–31.