Error-driven learning in statistical summary perception

Judith E. Fan, **Nicholas B. Turk-Browne**, and **Jordan A. Taylor**

Department of Psychology, Princeton University

Abstract

We often interact with multiple objects at once, such as when balancing food and beverages on a dining tray. The success of these interactions relies upon representing not only individual objects, but also statistical summary features of the group (e.g., center-of-mass). While previous research has established that humans can readily and accurately extract such statistical summary features, how this ability is acquired and refined through experience currently remains unaddressed. Here we ask if training and task feedback can improve summary perception. During training, participants practiced estimating the centroid (i.e., average location) of an array of objects on a touchscreen display. Before and after training, they completed a transfer test requiring perceptual discrimination of the centroid. Across four experiments, we manipulated the information in task feedback and how participants interacted with the objects during training. We found that vector error feedback, which conveys error both in terms of distance and direction, was the only form of feedback that improved perceptual discrimination of the centroid on the transfer test. Moreover, this form of feedback was effective only when coupled with reaching movements towards the visual objects. Taken together, these findings suggest that sensory-prediction error — signaling the mismatch between expected and actual consequences of an action — may play a previously unrecognized role in tuning perceptual representations.

Keywords

ensemble statistics; perceptual learning; feedback; perception and action; sensory-prediction error

Introduction

We often encounter objects in groups: fallen leaves on the lawn, food and beverages on a dining tray, or a stack of papers on our desk. In such cases, we may have the goal of interacting effectively with the group of objects: minimizing time raking up the leaves, keeping the tray balanced while carrying it across the room, or moving the stack of papers without tipping it over. Achieving these goals relies upon accurately representing not only the individual objects that comprise each group, but general properties of the whole group of objects, such as the area of highest density or its center-of-mass.

Prior investigations have shown that people are able to quickly and accurately extract such "statistical summary features" from groups of objects. For example, naive participants can

Please address correspondence to: Judith Fan, Peretsman-Scully Hall, Department of Psychology, Princeton University, Princeton NJ 08540, jefan@princeton.edu, Telephone: (609) 258-8729, Fax: (609) 259-1113.

extract the mean size of an array of disks of varying sizes (Chong & Treisman, 2003, 2005), the mean orientation of peripherally-presented oriented gratings (Parkes et al. 2001), and the mean emotional expression of a set of faces (Haberman & Whitney, 2007). Good performance on these summary judgment tasks has been found even when participants are highly impaired at judging the properties of an individual object in the set (Ariely, 2001). However, although these findings show that certain summary features are readily accessible, the computations underlying summary perception are not fully understood (Bauer, 2009; Myczek & Simons, 2008; Ariely, 2008). Moreover, these demonstrations do not explain how people acquire this ability, nor explore conditions for further learning.

Examining how learning affects statistical summary perception may provide crucial insight into how these summary representations are formed and used to guide behavior. Generally speaking, such learning may proceed in two ways: (1) in an *unsupervised* manner, in which mere exposure to sensory inputs leads to better discriminability along task-relevant dimensions; or (2) in a *supervised* manner, in which external feedback about the correctness of discrimination judgments during training improves accuracy in other contexts. A major goal of the current study was to directly examine the contribution of feedback to learning in statistical summary perception (supervised), especially as compared to the effects of experience alone (unsupervised).

Perceptual learning entails the enhancement of performance on a perceptual task (e.g., orientation, size, position discrimination) due to practice. The role of external feedback in perceptual learning, however, is complex. While several studies report little effect of feedback on performance improvements (Ball & Sekuler, 1987; Karni & Sagi, 1991; Shiu & Pashler, 1992), others have found that feedback enhances and may even be necessary for perceptual learning (Herzog & Fahle, 1997; Seitz, Nañez, Holloway, Tsushima, Watanabe, 2006). One possible reason for this lack of consensus is that feedback has largely been treated in a unitary fashion in most studies of perceptual learning. Specifically, traditional experimental procedures used to examine perceptual learning have relied on binary judgments, and have thus been restricted to manipulating feedback in a purely success-based fashion (see Liu, Dosher, & Lu, 2014 for model and review). By contrast, there is a long history of examining more graded forms of feedback in motor learning (Salmoni, Schmidt, & Walter, 1984).

In everyday life, however, where people not only look at sets of objects, but also interact with them, sources of feedback other than binary task success are available. Specifically, while interacting with objects, we continuously experience sensory feedback arising from visual and haptic/tactile information. For example, upon lifting an unevenly loaded tray, we quickly notice if the tray is beginning to tilt, allowing us to compensate for unexpected forces by adjusting the position of our grip in a continuous manner. The mismatch between the expected position of the tray upon lifting it (balanced) and its actual position (tilted) constitutes a form of *sensory-prediction error*, which conveys visual and proprioceptive information about the deviation between the expected and actual consequences of an action (Synofzik, Their, and Linder 2006). Sensory-prediction errors have been shown to be crucial for basic kinds of motor learning (see Shadmehr, Smith, and Krakauer 2010 for a review). For example, when reaching towards a target location, the mismatch between the expected

hand position based on the motor command and actual hand position based on visual information is sufficient to induce adaptation of subsequent reaching movements to account for earlier errors (Tseng, Diedrichsen, Krakauer, Shadmehr, & Bastian, 2007). More generally, the amount and kind of sensory feedback available during training can directly affect the nature of learning and the extent of generalization of motor performance to new contexts (Izawa and Shadmehr 2010; Nikooyan and Ahmed 2014; Taylor, Hieber, & Ivry, 2013). Finally, an active movement appears to be necessary for sensorimotor adaptation, as passive viewing of errors does not produce sensory-prediction errors (Held and Gottlieb 1958; Held and Freeman 1963; Kording and Wolpert 2004).

In order to directly examine the contributions of these different forms of error feedback ('knowledge of results') to statistical summary perception, we developed a reaching task in which participants made pointing movements towards the centroid (i.e., average location) of an array of dots. We hypothesized that providing vector feedback (i.e., conveying both distance and direction error) after the pointing movement would generate sensory-prediction errors, and lead to improved perceptual discrimination of the centroid after training. Whether other forms of feedback conveying less spatially precise error information (i.e., scalar feedback conveying only distance to the centroid) might improve perceptual discrimination of the centroid was an open question. Before and after training, we tested changes in discrimination ability using a separate perceptual test.

Experiment 1 (Centroid Discrimination)

The goal of this experiment was to investigate the influence of error-related feedback on the fidelity of the perceptual representation of the centroid of a set of visual objects. We tested this by manipulating whether participants received vector feedback or scalar feedback on a trial-by-trial basis while practicing a task requiring pointing movements towards the centroid. To provide a baseline measurement of the effects of task practice alone, we also included a control condition in which participants did not receive any trial-by-trial feedback.

We hypothesized that vector feedback might lead to faster improvement during training than the other conditions, so we planned to examine the time course of performance on the pointing task. In order to distinguish between task-related practice effects from genuine changes in the fidelity of the underlying representation, we also measured the degree of improvement on a separate perceptual task performed before and after pointing-task training. This perceptual task was designed to match the pointing task insofar as it required computing the centroid location, although it required a different response.

Methods

Participants—Forty-five young adults (29 women, age range: 18–34y) were recruited for participation from the research subject pool at Princeton University or the Princeton, NJ community. In this and all subsequent experiments, each participant received course credit or $24 for participation and all provided informed consent to a protocol approved by the Princeton IRB. All participants were right-handed and reported normal or corrected-to-normal visual acuity.

Apparatus—The experiment was controlled by a computer running MATLAB 2013a and Psychtoolbox (http://psychtoolbox.org). Stimuli were presented on a vertically oriented 58.42-cm touchscreen display (Acer T232HL LCD) with 60 Hz refresh rate positioned 43 cm from the participant.

Stimuli and procedure—Stimuli were presented against a dark gray background on a touchscreen display (49.9° × 40.0°; Fig. 2A). In all phases, trials began with the onset of a central fixation cross, presented for 550–1000ms (randomly jittered). The stimulus array comprised 8 white dots (diameter=0.9°) presented for 200 ms. This number of items was chosen since it exceeds traditional estimates of the capacity of visual short-term memory (Luck & Vogel, 1997; Zhang & Luck, 2008), discouraging explicit encoding and averaging strategies, while still being small enough that our measures would remain sensitive to errors in perceptual averaging. The stimulus duration was chosen in order to match those used in earlier studies of statistical summary perception (e.g., Chong & Treisman, 2003) and be short enough to prevent multiple eye movements within a trial (Saslow, 1967). Following response, the display remained blank for 1500ms.

Individual dot locations were independently sampled from a bivariate Gaussian distribution (circularly symmetric; σ of marginal distributions=11.2°). Specifically, on each trial we first sampled 8 locations from this bivariate Gaussian distribution centered on the origin. Because the mean of a random sample is likely to deviate from the true mean of the underlying distribution, we next applied a rigid spatial shift to all 8 locations so that the sample centroid would exactly coincide with the origin. Finally, in order to generate the final display, we applied another rigid spatial shift to all 8 locations in order to move the sample centroid to the predetermined centroid location for that trial.

The centroid is defined as the average of the position coordinates of all dots. More precisely, the coordinates of the centroid, x_c and y_c, is computed as the sum of the product of the individual areas and coordinates of each dot, divided by the total number of all dots on the display:

$$x_c = \frac{1}{n} \sum x_i \quad y_c = \frac{1}{n} \sum y_i$$

The experimental protocol consisted of three phases over two consecutive days: a pre-test phase on Day 1, and training and post-test phases on Day 2. The post-test was administered immediately after the conclusion of the training phase.

Quadrant task: The pre-test and post-test phases consisted of a four alternative forced choice (4AFC) perceptual task (4 blocks, 50 trials/block), which entailed discriminating which of the four display quadrants (NE, NW, SE, SW) contained the centroid and tapping the display anywhere within that quadrant of the display after stimulus offset with their right finger (Fig. 2A). Response time (RT) was measured as the time elapsed between stimulus offset and response (Table 1).

Given the long history of using discrimination judgments to assay the quality of perceptual representations in visual psychophysics (e.g., Herzog & Fahle, 1997; Petrov, Dosher, & Lu, 2005), we thought such a task might be appropriate. In traditional perceptual learning tasks that entail discriminating perceptual features that differ along a single dimension (e.g., orientation), binary judgments are appropriate. Because the location of the centroid varied along two dimensions (i.e., horizontal & vertical position), we reasoned that the minimal space of response options would be four alternatives. Thus, dividing the display into four quadrants was the most natural way of achieving this.

In order to make the task more consistently difficult, we constrained the location of dots to appear close to the quadrant boundaries (Fig. 1). Specifically, centroid locations in the quadrant task were distributed following a hyperbolic function: $r = c / \sqrt{cos(\theta) sin(\theta)}$, where r is the radial distance from the center of the display, θ represents randomly sampled angles, and c is a scaling factor that controls the overall distance to the nearest quadrant boundaries. Smaller values of c produce displays with centroids falling closer to boundaries, thus making discrimination more difficult. The rationale for using the hyperbolic function derives from the spatial layout of the four quadrants on the display. Specifically, we assume that the likelihood that the participant will select the correct quadrant varies according to the angular location of the centroid. At angular locations near the main diagonals (y = x and y = −x), the distance between the centroid and each quadrant boundary is more closely matched. However, as the centroid approaches angular locations near the vertical (y-axis) or horizontal (x-axis) meridians, its relative distance to the boundary dividing two adjacent quadrants is much smaller than its relative distance to the other two quadrants, effectively biasing discrimination towards a binary choice between the correct and nearest quadrant, rather than a 4-way choice among all quadrants. A hyperbolic function was chosen in order to better match discrimination difficulty at all angular locations (at least relative to a circular function, where radial distance is constant at all angles). Nevertheless, because centroid locations were sampled from the same distribution in all conditions, the specific choice of distribution does not bear on the most critical comparisons in the current study, which are between conditions.

In order to ensure that the difficulty of trials in the pre-test and post-test phases was equated, we used a separate thresholding procedure on Day 1 to estimate the value of c that would result in baseline accuracy of 62.5% on the 4AFC task (QUEST; Pelli & Watson. 1983). No feedback was given during the thresholding, pre-test and post-test phases. We restricted values of c to fall in a broad range: [0.344 cm, 3.44 cm]. In the rare case that estimated values of c fell out of this range, we extended the thresholding phase until c converged on an acceptable value. Nevertheless, estimation of c only served to calibrate the difficulty of the quadrant task such that participants achieved approximately 62.5% accuracy on the task (the midpoint between 25% and 100%). Because the value of c was fixed within-participant, changes in perceptual discrimination were measured by computing the percent change in accuracy between the pre-test and post-test phases for each participant.

Pointing task: The training phase consisted of a centroid estimation task (12 blocks × 50 trials/block = 600 total trials), which entailed tapping the display at the estimated location of

the centroid with the right index finger (Fig. 2A). Response time was measured the same way as in the quadrant task (Table 1).

Centroid locations in this pointing task were sampled from a uniform distribution over a rectangular region spanning the central 25% of the display. Note that the distribution of centroid locations in the pointing task differs from that used in the quadrant task in this and all experiments. The type of feedback provided to participants was manipulated across separate groups during the training phase. Participants in the Vector condition (N = 15) received trial-by-trial error information about the distance and direction of the actual centroid from their response. This vector error was provided by marking the participant's response with a blue crosshair and the actual centroid with a green crosshair. Participants in the Scalar condition (N = 15) received trial-by-trial error information proportional to the distance of their response from the centroid but no direction-error information. Scalar error was returned as a numerical point score (range=0–10 points), where the number of points earned followed a Gaussian reward function of deviation from the centroid (with $\sigma = 11.2° / \sqrt{8}$), rounded to the nearest integer value. Participants in the Control condition (N = 15) did not receive error feedback on a trial-by-trial basis. Regardless of condition, all participants received task performance feedback at the end of each 50-trial block in the form of a running cumulative point total and participants were encouraged to try to earn as many points as possible. To familiarize participants with the pointing task, participants completed 8 practice trials without feedback.

Data Analysis: On each trial, error was computed as the distance between the actual centroid and participant's response. Outlier responses with errors exceeding 3× the standard deviation of each participant's errors were removed prior to computing aggregate statistics of performance. Overall performance for each participant was quantified as root-mean-squared error (RMSE). RMSE is a summary statistic that captures overall deviation between the centroid and participants' responses, and reflects both the accuracy (bias) and precision (variability) of responses. We also report the mean radial error (average distance between the centroid and response) and the standard deviation of errors, computed within-participant, to assess potential dissociations between these two error components.

While the primary goal of this study was to examine the consequences of training with feedback on subsequent perceptual discrimination of the centroid, we also examined changes in performance over the course of the training phase, reasoning that such changes might help explain any training effects observed. The time course of errors was fit with an exponential function ($error_{abs} = be^{ax}$) for each participant, where the exponential parameter a captures the rate of change in performance over the course of training (where positive values indicate increasing error and negative values indicate decreasing error). The same exponential function was also fit to response times during the training phase.

We also examined the presence of any directional biases (e.g., consistently pointing above and to the right of the centroid) by measuring the angular component of the error vector. These angles were tested for uniformity using the Hodges-Ajne test, a non-parametric test of deviations from uniformity in circular data (Ajne, 1968; Zar, 1999). In the absence of a directional bias, the distribution of these angles should be circularly uniform around the

centroid on average (in the range [−180°,180°]), where θ=0 was defined as pure rightward bias, and counterclockwise rotation is positive).

In order to derive an empirical estimate of chance performance on this task, we conducted simulations informed by the actual spatial distribution of responses participants made in this task (as opposed to assuming uniform probability of responding at all locations in the display). Specifically, for each trial in the pointing task, we simulated random responding by sampling from the set of responses a given participant had made, and computed error on the basis of this random response. This procedure was repeated 1000 times for each participant in order to derive robust estimates of RMSE under completely random responding. Alternatively, participants, although not guessing completely at random, may have been able to achieve above chance-level performance by subsampling the display (e.g., pointing to the average location of only 3 or 4 dots, rather than the full set of 8). This possibility was examined by estimating the goodness-of-fit (RMSE) of each of the intermediate cases between random guessing (i.e., 0 of 8 dots) and full averaging of all dots presented (i.e., 8 of 8 dots) to each participant's actual response distribution. Specifically, this was estimated by sampling a random subset of dots from each trial of a fixed size (e.g., 3 of 8 dots), recording the distance between the centroid of this random subset and the participant's response, and then computing RMSE. This procedure was performed 1000 times for each subset size, for each participant, in order to derive robust estimates of average deviation between participant responses and the centroids of these random subsets.

Although dot locations were sampled from the same underlying distribution, individual trials varied in how dispersed the dots were on the display. Trials with high dispersion might be more difficult (i.e., result in less precise centroid representation) than trials with low dispersion. To quantify how dot dispersion affected centroid estimation, we used multiple regression to model the contribution of horizontal location variability (s_x) and vertical location variability (s_y) to error magnitude across trials in the pointing task for each participant ($error = s_{xi}^T \beta_{sx} + s_{yi}^T \beta_{sy} + \varepsilon_i$).

Results

Pre-test—Pre-test performance on the quadrant task in all groups was significantly better than chance (59.0% correct vs. 25%; $t(44)=25.6$ $p \ll 0.001$, $d=3.81$), and did not differ among groups ($F(2,42)=1.76$, $p=0.184$, $\eta^2=0.078$). Estimates of c, which controlled quadrant task difficulty (see Methods), also did not differ between groups at baseline (Vector: 0.958 cm [s.e.m. = *0.123 cm*], Scalar: 1.01 cm [s.e.m. = 0.097 cm], Control: 1.21 cm [s.e.m. = 0.194 cm]; $F(2,42)=0.814$, $p=0.450$, $\eta^2=0.0373$), nor did median response time ($F(2,42)=1.46$, $p=0.244$, $\eta^2=0.065$).

Training

Overall: Participants received different kinds of information following a response in the pointing task during the training phase (Fig. 2B). Performance in all groups highly exceeded that expected by chance ($p\ll0.001$), verifying that participants were engaged with the task (Fig. 3A). Overall performance (RMSE) during the training phase did not reliably differ among groups ($F(2,42)=1.76$, $p=0.184$, $\eta^2=0.078$; Vector RMSE: 1.35cm; Scalar RMSE:

1.41 cm; Control RMSE: 1.54 cm). Radial error (i.e., reflecting the accuracy component of performance) also did not differ among conditions (Vector: 1.19 cm, Scalar: 1.24 cm, Control: 1.35 cm; $F(2,42)=1.61$, $p=0.212$, $\eta^2=0.071$; Fig. 3B); nor did error variability (i.e., reflecting the precision component of performance; Vector: 0.636 cm, Scalar: 0.678 cm, Control: 0.748 cm; $F(2,42)=2.17$, $p=0.127$, $\eta^2=0.094$).

Time course: We thought it reasonable to expect participants to show steady improvement over the course of training, where the rate of improvement would be highest for the groups receiving either Vector or Scalar feedback. Surprisingly, these time series were generally flat (i.e., non-significant exponential learning rate) over the training period tested in all three groups (Vector: $t(14)= 0.996$, $p= 0.336$; $d= 0.257$; Scalar: $t(14)= 0.905$, $p= 0.381$, $d= 0.233$; Control: $t(14)= 0.676$, $p= 0.510$, $d=0.175$), suggesting that participants may have been performing close to ceiling. Although response accuracy did not improve over time, response times did significantly decrease over the course of training in all three conditions (Vector: $t(14)= 2.27$, $p=0.040$; $d= 0.585$; Scalar: $t(14)= 4.52$, $p<0.001$; $d=1.17$; Control: $t(14)= 4.92$, $p<0.001$; $d=1.27$). Neither the rate of decrease in RTs ($F(2,42)=1.12$, $p=0.335$, $\eta^2=0.051$) nor overall RTs ($F(2,42)=0.488$, $p=0.617$, $\eta^2=0.023$) differed between conditions, suggesting that this temporal trend in RT more likely reflects generic task practice effects, rather than the consequences of differences between feedback conditions.

Directional bias: Despite the lack of overall RMSE effects during training, there may have been differences in the degree to which participants were biased to respond between conditions. Using the Hodges-Ajne test for non-uniformity in the angular component of errors, we found significant evidence for directional bias Control participants ($p=0.012$; 95% CI: $[–23.9°, –90.0°]$): a bias below and to the right of the centroid. There was a trend toward deviation from uniformity in the Scalar group ($p=0.071$; 95% CI: $[5.1°, –77.6°]$), and no reliable non-uniformity in Vector participants ($p=0.250$; 95% CI: $[–22.9°, 109.6°]$). Nevertheless, 95% confidence intervals for the direction of bias overlapped heavily in the quadrant lying southeast of the centroid for all groups. These results are consistent with the possibility that there may be a southeast bias in the pointing task, which was reduced for participants in the Vector condition who received spatially precise directional error information on every trial.

Relationship to stimulus variability: While the locations of dots were drawn from the same distribution, their exact locations varied from trial-to-trial. We reasoned that trials with higher variability in dot location would be more difficult, and thus be associated with larger errors on average. To test this possibility, we used multiple regression to model the contribution of horizontal location variability and vertical location variability to error magnitude across trials for each participant. Across all conditions, we found that both horizontal location variability and vertical location variability were both highly predictive of error magnitude (Vector: $\beta_{sx}=0.056$: $t(14)=8.48$, $p<0.001$; $\beta_{sy}=0.028$: $t(14)=3.19$, $p=0.007$; Scalar: $\beta_{sx}=0.083$: $t(14)=6.57$, $p<001$; $\beta_{sy}=0.068$: $t(14)=6.66$, $p<0.001$; Control: $\beta_{sx}=0.084$: $t(14)=7.27$, $p<0.001$; $\beta_{sy}=0.076$: $t(14)=7.75$, $p<0.001$). To examine potential differences between conditions due to feedback, we submitted these linear coefficients to a 2 axis (horizontal, vertical) $\times$ 3-condition (Vector, Scalar, Control) ANOVA. We found that the

degree of dependence between variability and error was lower for Vector participants than for Scalar or Control participants (main effect of condition: $F(2,84)=8.45$, $p<0.001$, $\eta^2=0.167$), suggesting that vector error feedback may serve to reduce noise in centroid estimation for more difficult trials.

Post-test—Despite showing similar performance during training, post-test phase performance in the quadrant task revealed the consequences of training: perceptual discrimination was reliably enhanced in the Vector group (% change: 12.6; $t(14)=3.05$, $p=0.0087$, $d=0.787$; Fig. 3C), but not in either of the Scalar (% change: 3.17; $t(14)=0.058$, $p=0.954$, $d=0.015$), or Control groups (% change: −0.002; $t(14)=0.058$, $p=0.954$, $d=0.015$). The degree of enhancement was greater for Vector participants than for Control participants ($t(28)=0.027$, $p=0.027$, $d=0.853$; uncorrected), and Scalar and Control did not differ from one another ($t(28)=0.639$, $p=0.528$, $d=0.233$). The difference in improvement did not reach significance for Vector vs. Scalar groups, however ($t(28)=1.66$, $p=0.108$, $d=0.607$). Taken together, these findings suggest that vector error feedback tuned centroid representations in a more general manner, promoting transfer to a separate perceptual task.

For completeness, we additionally performed a 2 phase (pre-test, post-test) × 3 condition (Vector, Scalar, Control) ANOVA on RTs, which yielded no significant interaction between phase and condition ($F(2,82)=0.556$, $p=0.576$, $\eta^2=0.013$), although there was a main effect of condition ($F(2,82)=6.14$, $p=0.003$, $\eta^2=0.130$) and marginally significant effect of phase ($F(1,82)=3.84$, $p=0.053$, $\eta^2=0.045$). Follow-up analyses of post-test RTs showed that median response times were lower in the post-test than pre-test for all conditions ($ps<0.039$) and critically, in the Vector condition ($t14)=3.06$, $p=0.008$, $d=0.791$), arguing against the possibility that a speed-accuracy tradeoff explained the improvement in accuracy for Vector participants. The rate of decrease in RTs (exponential term in model fit) during the training phase did not correlate with the degree of accuracy improvement for Vector participants ($r=−0.119$, $p=0.673$).

Experiment 2 (Center-of-Mass Discrimination)

In Experiment 1, we found that participants who received Vector feedback during training showed better perceptual discrimination of the centroid in the post-test, although this was not accompanied by significant performance changes during training. On the other hand, we observed no learning in either of the Scalar and Control conditions. The current experiment sought to expand these findings in two additional ways:

First, to explore the possibility that the lack of improvement during training might reflect ceiling performance in this task, we sought to make the task more difficult. Having found that greater dot location variability was associated with larger errors in Experiment 1, we reasoned that increasing the amount of overall stimulus variability in the display might make the task more difficult, providing a broader dynamic range in which to observe performance changes during the pointing task. We accomplished this by varying the size as well as location of dots, and eliciting estimates of the center-of-mass of the array, where the area of each dot was taken into account.

Second, we sought to test the importance of error feedback, *per se*, to learning. To accomplish this, we included a Guided condition in which participants were passively guided to the correct location of the center-of-mass immediately after array offset. Thus, this condition was identical to the Vector condition except that pointing response error was virtually eliminated by the onset of the response cue.

Methods

Participants—Sixty young adults (40 women, age range: 18–34y) participated.

Stimuli and procedure—On each trial, participants were presented with an array of 8 dots that varied in both size[1] and location (Fig. 2C). The diameter of individual dots was determined by sampling independently from a Gaussian distribution centered on the base diameter used in Experiment 1 (0.9°) with $\sigma = 0.3°$ (only dots with positively valued diameter were used). Participants were instructed to aim for the center-of-mass of the array, taking into account the area of the dots, and treating the dots as flat (as opposed to having volume). The center-of-mass is defined as the average of the position coordinates of all dots, weighted by their surface area. The coordinates of the center-of-mass, x_c and y_c, may be computed as the sum of the product of the individual areas and coordinates of each dot, divided by the total area of all dots on the display:

$$x_c = \frac{\sum m_i x_i}{\sum m_i} \quad y_c = \frac{\sum m_i y_i}{\sum m_i}$$

This experiment included Vector and Scalar conditions (N = 20 each) that were identical to those in Experiment 1. Because the Scalar condition provides a tighter control for the Vector condition than the Control condition (which did not provide trial-by-trial feedback at all), and the two conditions did not differ in Experiment 1, we did not recruit a separate Control group.

Another goal of this study was to isolate the contribution of error information in feedback, *per se*. To accomplish this, we included a Guided condition (N = 20), identical to the Vector condition except that the actual center-of-mass location was revealed 100ms after stimulus offset but *prior* to response. Guided participants were still instructed to point to the location of the center-of-mass on each trial, but now it was marked with a green crosshair until they made their response. If pairing each dot array with its center-of-mass is sufficient to result in enhanced perceptual discrimination, then Guided participants should show positive transfer to the quadrant task. On the other hand, if response error is necessary to produce transfer, then Guided participants should not show positive transfer to the quadrant task.

Data Analysis: Response data were submitted to the same set of analyses as in Experiment 1. To quantify how much stimulus variability affected center-of-mass estimation in the

---

[1]In pioneering studies of statistical summary representation by Chong & Treisman (2003), size was operationalized as the diameter (of circular disks). They found that psychological estimates of the mean size of two disks systematically fall between the arithmetic mean of the diameter and the arithmetic mean of the surface area of the two disks. Prior work has also demonstrated that observers are sensitive to variability in stimulus size (e.g., Solomon, Morgan, & Chubb, 2011).

pointing task during the training phase, here we used multiple regression to model the contribution of horizontal location variability (s_x), vertical location variability (s_y), and variability in dot size (i.e., area, s_{area}) to error magnitude across trials for each participant ($error = s_{xi}^T \beta_{sx} + s_{yi}^T \beta_{sy} + s_{area\_i}^T \beta_{s_{area}} + \varepsilon_i$).

Results

Pre-test—Pre-test performance on the quadrant task in all groups was significantly better than chance (60.2% correct vs. 25%; $t(59)$=32.3, $p << 0.001$, d=4.17), and did not differ among groups ($F(2,57)$=0.998, p=0.375, η^2=0.034). Estimates of c also did not differ between groups (Vector: 1.10 cm [s.e.m. = 0.070 cm], Scalar: 1.11 cm [s.e.m. = 0.041 cm]; Guided: 1.20 cm [s.e.m. = 0.094 cm]; $F(2,57)$=0.441, p=0.645, η^2=0.015), nor did median response time ($F(2,57)$=2.14, p=0.127, η^2=0.070).

Training

Overall: During training, participants received either Vector or Scalar error feedback following their response on each trial, or were shown the location of the center-of-mass prior to making their response (Guided). As Guided participants had direct access to the location of the center-of-mass while pointing to the display, their responses provide an approximation of noise in the motor response when perceptual uncertainty about the centroid location is minimized. Overall, Guided participants made substantially smaller pointing errors than Vector and Scalar participants (ts>15.4, p<<0.001, d>4.88; Vector RMSE: 1.94 cm; Scalar RMSE: 1.91 cm; Guided RMSE: 0.358 cm). Consistent with Experiment 1, neither overall training phase performance (RMSE; $t(38)$=0.285, p=0.777, d=0.090; Fig. 4A), radial error (Vector: 1.70 cm; Scalar: 1.67 cm; Guided: 0.304 cm; ts<1) nor error variability (Vector: 0.935 cm; Scalar: 0.928 cm; Guided: 0.187 cm; ts<1; Fig. 4B) differed between Vector and Scalar conditions.

Time course: Our overall manipulation of task difficulty was successful (Experiment 2: Vector/Scalar pooled RMSE=1.93 cm vs. Experiment 1: Vector/Scalar/Control pooled RMSE = 1.43 cm; $t(83)$ = 7.23, p<0.001, d=1.57). However, there were no significant performance changes (i.e., nonzero exponential learning rate) over the course of the training phase in either the Scalar ($t(19)$= 1.32, p= 0.203, d=0.295) or Guided conditions ($t(19)$= 0.770, p= 0.451, d=0.172); performance actually tended to get worse in the Vector condition ($t(19)$= 2.40, p= 0.027 d=0.538), perhaps due to fatigue. This shows that increasing stimulus variability was not sufficient to bring about performance changes over the course of the pointing task.

Response times significantly decreased over the course of training in the Vector ($t(19)$= 3.77, p=0.001; d=0.843) and Scalar conditions ($t(19)$= 4.95, p<0.001; d=1.11), though there was no significant trend in the Guided condition ($t(19)$= 0.785, p=0.442; d=0.175). This suggests that some component of response planning may generally speed up with additional task practice, but participants likely reached the minimum response time in the Guided condition.

Relationship to stimulus variability: To examine the relationship between stimulus variability and pointing task performance, we again used multiple linear regression to model the contributions of variability in horizontal location, vertical location, and dot area to errors over time. We found that all three of these factors reliably predicted error magnitude for Vector (β_{sx}=0.118: $t(19)$=11.0, p<0.001; β_{sy}=0.150: $t(19)$=16.2, p<0.001 β_{ssize}=0.754; $t(19)$=3.96, p=0.001) and Scalar (β_{sx}=0.112: $t(19)$=14.3, p<0.001; β_{sy}=0.145: $t(19)$=16.1, p<0.001 β_{ssize}=0.719; $t(19)$=3.44, p=0.003) participants, but not Guided (ts<1) participants. To compare the degree to which these factors predicted error magnitude between conditions, we performed a 2 condition (Vector, Scalar) $\times$ 3-predictor (horizontal location, vertical location, size) ANOVA, which did not reveal a main effect of condition ($F(1,114)$=0.028, p=0.869, η^2=0.0002). This suggests that returning vector error feedback does not necessarily reduce the influence of stimulus variability on error.

Directional bias: Using the Hodges-Ajne test of uniformity in the angular component of errors, we found that neither Vector (p=0.11) nor Scalar groups (p=0.59) showed systematic evidence of directional bias.

Post-test—Post-test phase performance was again reliably enhanced in the Vector group (% change: 0.071; $t(19)$=2.63, p=0.017, d=0.588; Fig. 4C), but not in the Scalar group (% change: −0.007; $t(19)$=0.257, p=0.800, d=0.058) nor the Guided group (% change: −0.008; $t(14)$=0.413, p=0.685, d=0.092). The degree of enhancement for Vector participants was marginally greater than for Scalar participants ($t(38)$=2.09, p=0.044, d=0.660) or Guided participants ($t(38)$=2.38, p=0.023, d=0.752), and Scalar and Guided performance did not differ from one another ($t(38)$=0.045, p=0.965, d=0.014). Median response time did not differ between groups ($F(2,57)$=0.873, p=0.423, η^2=0.030). These results provide additional evidence that displaying the location of the center-of-mass as vector error feedback during training — beyond merely pairing this location with the dot array or returning scalar error feedback — is necessary to observe positive transfer to another task.

We additionally performed a 2 phase (pre-test, post-test) v 3 condition (Vector, Scalar, Guided) ANOVA on RTs, which yielded no significant interaction between phase and condition ($F(2,112)$=0.333, p=0.717, η^2=0.006), and no main effects of either condition ($F(2,112)$=1.96, p=0.145, η^2=0034) nor phase ($F(1,112)$=0.416, p=0.520, η^2=0.004). Median response times were numerically (though not significantly) lower in the post-test than pre-test in the Vector condition ($t(19)$=1.19, p=0.249, d=0.266), inconsistent with a speed-accuracy tradeoff. The rate of decrease in RTs (exponential term in model fit) during the training phase did not significantly correlate with the degree of accuracy improvement for Vector participants (r=−0.305, p=0.191).

Experiment 3 (Role of Pointing Movements in Learning)

Experiments 1 and 2 manipulated access to vector error information by returning spatially precise information about the location of the centroid with respect to the location of the participant's response, registered at the point of contact between the participant's fingertip and the display. Although prior studies of statistical summary perception have not involved action toward the summary features, prior work in the field of motor learning suggests that

such actions are required to generate sensory-prediction error (for a review, see Shadmehr, Smith, & Krakauer, 2010). Experiment 3 sought to test whether sensory (i.e., haptic and visual) feedback of this kind may have contributed to the learning observed in the first two experiments. Towards this end, we designed a task that elicited perceptual estimates of the centroid during training without a pointing movement towards the display. Instead, participants typed the spatial coordinates of the centroid according to its position in a grid overlaying the display (grid task).

Participants

Twenty young adults (11 women, age range: 18–34y) participated in Experiment 3.

Stimuli and procedure

The presentation of the dot array was identical to that used in Experiment 1. During the training phase, the entire display was overlaid by a rectangular grid whose cells measured $1.2° \times 1.2°$, about the size of a fingertip, in order to approximate the precision of responses afforded by the pointing task. Cells lying within the central 25×25-cell region ($30.2° \times 30.2°$) were labeled with an arbitrary 2-character alphanumeric code that varied from trial-to-trial (Fig. 2D). Dot stimuli were of equal size and sampled in an identical manner to that used in Experiment 1. Participants identified the grid cell containing the centroid and typed in the corresponding code. Response entry caused the selected cell to become highlighted with a white border. Error feedback was returned by marking the true location of the centroid with a green crosshair. The grid was visible throughout the entire trial, whereas the 2-character codes were only visible during response.

A continuous measure of error was derived by computing the distance between the center of the selected cell and the true centroid on each trial, allowing us to submit these data to the same set of analyses used in Experiments 1 and 2. This metric provides an unbiased measure of response accuracy because the centroid was distributed uniformly across trials, but may result in slight underestimation of response precision owing to additional measurement error. Performance greatly exceeded that expected by chance ($p<<0.001$) or by only averaging a subsample of items on the display ($p<<0.001$), verifying that participants could perform the task.

Prior to beginning the training phase, participants completed 8 practice grid-task trials where only one target dot appeared on each trial and the goal was to select the cell corresponding to this dot's location. The pre-test and post-test phase structure was identical to that used in the first two experiments.

Results

Pre-test—The pre-test and post-test phases were identically structured to those in the first two experiments. Pre-test performance on the quadrant task was significantly better than chance (60.9% correct vs. 25%; $t(19)=18.0$, $p << 0.001$, $d=4.02$), and the value of c (mean = 1.04 cm, s.e.m. = 0.079 cm) did not differ from that of participants in Experiment 1 ($t(63)=0.113$; $p=0.911$; $d=0.03$).

Training—Overall performance (RMSE=1.71 cm; mean radial error=1.47cm; error s.d.=0.862cm; Fig. 5A) greatly exceeded chance (p<<0.001). Moreover, although the grid task differed from the pointing task both in the presence of gridlines on the display and the type of response, performance was comparable to that for the pointing task in Experiment 1 (Vector RMSE=1.35cm) & Experiment 2 (Vector RMSE=1.94cm). This shows that these changes to the display and response modality did not greatly impact participants' ability to estimate the centroid location. Using the same Hodges-Ajne test of angular uniformity as in the previous experiments, we found no evidence of directional bias (p=0.591). As in the previous experiments, there was also no evidence of improvement over the course of training (t(19)=0.275, p=0.787, d=0.061).

One potential concern is that if the grid cells were too large, participants might not have treated trials on which they selected the correct grid cell as errors, even though their estimate of the centroid and the true centroid differed. However, the proportion of trials on which the correct grid cell was selected was only 11.1% (s.e.m.=0.63%), showing that the grid cells were small enough for participants to make errors on most trials.

Post-test—If visual comparison (i.e., absent proprioceptive feedback) between the location of the selected grid cell and the precise centroid location provides a sufficiently strong error signal, then training on the grid task should also result in positive transfer to the quadrant task. However, post-test performance did not reliably differ from pre-test performance for participants who trained on the grid task (% change: 2.69; t(19)=1.24, p=0.231, d=0.277; Fig. 5C). These results are consistent with the possibility that sensory-prediction error — conveying information about the mismatch between the expected centroid location (where response movement was aimed) and the actual centroid location (green crosshair) — may play a crucial role in tuning the perceptual representation of the centroid.

Experiment 4 (Change in the Fidelity of Representation)

In what ways can training alter centroid representations? Improvements on the quadrant task in the Vector group (Experiments 1 and 2) showed that some change in the perceptual representation occurred, but this was based on a relatively coarse measure — proportion correct in a categorical judgment task. Experiment 4 sought to better quantify changes in the accuracy and precision of the perceptual representation of the centroid. To accomplish this, we used the grid task (from Experiment 3) to collect continuous reports of the centroid location in the pre-test and post-test, allowing us to directly measure the quality of the perceptual representation of the centroid before and after training on the pointing task.

Methods

Participants—Twenty young adults (14 women, age range: 18–34y) participated in Experiment 4.

Stimuli and procedure—Dot stimuli were identical to those used in Experiments 1 and 3. Participants performed the pointing task with vector error feedback during training, and the grid task without feedback during the pre-test and post-test phases (Fig. 2E).

Results

Pre-test—Pre-test performance on the grid task (RMSE=1.92 cm) greatly exceeded chance ($p \ll 0.001$; Fig. 6A) and did not differ from that of participants during the training phase (RMSE=1.71cm) in Experiment 3 ($t(38)=1.38$; $p=0.176$; $d=0.436$).

Training—Likewise, overall pointing task performance (RMSE=1.37cm) during the training phase did not differ from that of the Vector group in Experiment 1 (RMSE=1.43cm; $t(33)=0.156$; $p=0.877$; $d=0.053$). Moreover, radial error (1.20 cm) and error variability (0.659 cm) also did not differ (ts<1). However, participant's pointing errors significantly deviated from angular uniformity in($p=0.03$; 95% CI: [−56.6°, 15.5°]), consistent with an overall bias below and to the right of the centroid. There was no evidence of performance changes (i.e., non-zero learning rate) over the course of the training phase ($t(19)= 1.52$, $p= 0.146$, $d=0.339$).

Post-test—Consistent with the findings in Experiments 1 and 2, training on the pointing task with vector error feedback resulted in generalization to a different task requiring fine perceptual discrimination of the centroid, as quantified by the reduction in RMSE on the grid task in the post-test ($t(19)=2.93$; $p=0.009$; $d=0.654$; Fig. 6C and 5D). This change in overall grid task performance reflected both decreases in average radial error ($t(19)=2.37$; $p=0.029$; $d=0.530$), as well as in the variability of errors ($t(19)=2.80$; $p=0.011$; $d=0.626$).

Discussion

In the present study, we sought to understand the role of learning in the formation of statistical summary representations. Specifically, we examined how error-related feedback altered the fidelity of the perceptual representation of the centroid of a set of visual objects. We found converging evidence across four experiments that providing vector error feedback while participants practiced making pointing movements toward the centroid improved the fidelity of their perceptual representation of the centroid, as measured by a separate perceptual test conducted before and after training (quadrant task accuracy in Experiments 1 and 2; grid task RMSE in Experiment 4). By contrast, participants receiving no trial-by-trial feedback (Control condition in Experiment 1) or only scalar feedback (Scalar condition in Experiments 1 and 2) did not show transfer to the separate perceptual test, suggesting that spatially precise error feedback was necessary to improve the fidelity of centroid representations. Critically, feedback had to be in the form of a vector error; participants who were passively guided toward the correct centroid location did not show transfer despite having precise centroid information (Guided condition in Experiment 2). Moreover, this form of feedback was effective only when coupled with reaching movements towards the visual objects; training on a centroid estimation task that did not involve a pointing movement did not result in learning (Experiment 3). To quantify changes in the perceptual representation, we measured the quality of the perceptual representation before and after training using a continuous measure (Experiment 4) and found that training with vector error feedback enhanced both the accuracy (i.e. reduced bias) and precision (i.e., reduced variability) of centroid estimates. As a whole, these findings suggest that sensory-prediction

error — signaling the mismatch between expected and actual consequences of an action — may play a previously unrecognized role in tuning perceptual representations.

Role of feedback in learning

What might underlie the specific advantage for training with vector error feedback? In the Vector condition, the true spatial location of the centroid was precisely marked immediately following a pointing response to the estimated centroid location. When these two points do not coincide, the result is a sensory-prediction error defined as the difference between the pointing movement and the spatial position of the true centroid.

On the other hand, in the Scalar condition, numerical feedback was presented that reflected the accuracy of the estimate, but the direction of displacement between the participant's estimate and the true centroid location was not provided. When the expected point score and the actual score earned do not coincide, this type of feedback produces reward-prediction error, reflecting a quantitative difference not tied to a particular sensory modality. Although this type of reward-based feedback has been shown to exert an important influence on learning and decision-making (Rescorla & Wagner, 1972; Schultz, Dayan, & Montague, 1997; O'Doherty, 2004), the role of reward-prediction error in sensorimotor adaptation is less certain (Izawa & Shadmehr, 2011; Nikooyan and Ahmed 2014).

In the Guided condition, participants were not given the opportunity to actively predict the location of centroid and thus did not experience sensory-prediction error during training. Consistent with the notion that sensory-prediction error may be necessary to induce perceptual learning, Guided participants, despite being provided with exactly the same visual information as Vector participants, did not improve on the perceptual task.

An alternative explanation is that Guided participants may have ignored the array, because the task did not demand them to compute its centroid. However, previous investigations of statistical summary perception have shown that observers are able to extract summary features even at short presentation durations (∼50ms, Chong & Treisman, 2003), leading to the general belief that such features are automatically computed in a pre-attentive manner. If true, participants would not be able to simply disregard this statistical information upon being shown the array (without exerting cognitive control to actively ignore the instructions). Future experiments could directly examine this possibility by including catch trials designed to test whether participants were ignoring the array.

Taken together, our findings suggest that the pointing response itself may have been critical for learning, as such visually guided movements provide rich and spatially precise sensory feedback about the position of the fingertip with respect to the centroid target. In the Grid training condition, both haptic feedback about finger position, as well as visual feedback about the location of the response, was less precise than in the Vector condition. Owing to this lack of precise sensory feedback during response, these participants only had access to a crude visual error signal, even though the true location of the centroid was marked on every trial. These findings are consistent with prior studies of motor adaptation that show that sensory feedback is weighted more heavily when this information is precise (Kording & Wolpert, 2004; Izawa and Shadmehr 2011).

One intriguing and counterintuitive aspect of these results is that perceptual performance in the quadrant and grid tasks was enhanced in the Vector condition even though pointing task performance did not significantly improve over the course of the training phase. There are multiple possible reasons for this, though at present we can only offer speculative accounts that could be directly tested in future studies. One straightforward possibility is that a single training phase was insufficient to see significant improvements in response accuracy because such improvement depends on a slow form of visuomotor learning that may be stable within-session but improves over multiple sessions (Karni & Sagi, 1993).

A second possibility is that additional engagement of the motor system in the pointing task vs. the grid task may have been required for generalization. Although we have emphasized the informational content carried in error feedback, classic neuropsychological findings (Mishkin, Ungerleider, & Macko, 1983; Goodale & Milner, 1992) support a distinction between an action-oriented representation formed along the dorsal visual stream and a more object-based representation along the ventral visual stream. Consistent with this account, it is possible that the 'dorsal representation' supplemented the 'ventral representation' of the centroid during visually-guided reaching to the centroid in the pointing task, yielding high pointing accuracy. Moreover, in the Vector condition, pointing movements in conjunction with precise spatial feedback about the centroid location generated sensory-prediction errors, which may have served to further tune the ventral representation, the benefits of which were only revealed in the perceptual test after training.

We also note that while the lack of online improvement during training is puzzling, it was also observed in a prior study (Bauer, 2009) that examined the time course of performance in a task requiring estimation of a statistical summary feature (i.e., mean length of a set of lines) when participants received different kinds of feedback. In that study, some participants were told their response was correct when it was close to the arithmetic mean, while others received feedback based on the quadratic or harmonic means, other mathematically valid ways to compute an average. Participants in the harmonic mean and quadratic mean conditions initially made large errors, which decreased over time. By contrast, participants in the arithmetic mean condition made smaller errors throughout, and did not significantly improve with additional practice. These results suggest that participants default to the arithmetic mean, although binary feedback can nudge them to bias their responses towards a different mean.

We use this prior work to support the idea that learning in the Vector condition more likely reflects fine tuning of the centroid representation, rather than coarse calibration to task demands. In the latter case, reduced error would not reflect enhancement in the underlying perceptual representation, but rather participants' figuring out which feature dimensions mattered for task performance. In our study, calibration to the centroid estimation task may have even been accomplished in the first few practice trials prior to the start of training. Nevertheless, vector error feedback may be required for this training to further tune the underlying representation.

Given that participants did not improve over the course of 600 trials of training, an open question concerns the amount of training required to give rise to the kind of learning we

observed. For example, it is possible that fewer trials may have been sufficient to induce some amount of generalization. On the other hand, if learning driven by sensory-prediction errors proceeds in a more incremental fashion, then a relatively long training session (as long or longer than that used in this study) may be necessary. Future investigations that characterize the 'dose-response' curve relating the amount of training and degree of learning would help to further illuminate the underlying components of error-driven learning in statistical summary perception.

Learning mechanisms

What learning mechanisms might explain how sensory-prediction errors tune the perceptual representation of the centroid? One possibility is that the representation of the individual items in the array is enhanced by directly altering the tuning properties of units residing in early visual areas (Fahle, Edelman, & Poggio, 1994; Gilbert, Sigman, & Crist, 2001; Karni & Sagi, 1991). Changes in the representation at that level would be predicted to result in highly specific perceptual expertise at trained locations, but be unlikely to result in generalization to novel locations (e.g., items in the left hemifield, if trained on the right). However, because such early changes alter the fidelity of the inputs to the rest of the system, such training might be predicted to transfer to other tasks involving precise representations of spatial position at these locations (e.g., local change detection). In the present study, we are unable to directly test this possibility, as the locations of individual items varied widely from trial to trial. Nevertheless, testing the spatial specificity of learning in statistical summary perception merits closer examination in future research.

Another possibility is that sensory-prediction errors lead to adjustment of the pattern of weights applied to the output from units in early visual areas in accordance with the demands of the task (Dosher & Lu, 2009; Petrov, Dosher, & Lu, 2005). On this account, errors in the centroid estimation task arise from improper weighting of units coding for specific locations. Most models of learning proceed by incremental updating of the weights associated with active units that contributed to the error (i.e., coding for position vectors that are correlated with the error vector), in proportion to error magnitude (i.e., gradient descent). Over several such trials of a centroid estimation task, such error-based updating would result in convergence upon uniform weights, eliminating systematic bias due to improper integration of individual item location information. This kind of learning mechanism would be predicted to facilitate generalization to novel displays with random item locations, but not necessarily to new tasks requiring a different set of weights. Although the centroid may be an already well-learned statistical summary feature for an adult observer, such a learning mechanism may help to explain how novel (Bauer, 2009) or high-level summary features (e.g., mean facial expression, Haberman & Whitney, 2007) are learned.

A third possibility is that participants learn statistical properties of the distribution that was most likely to have generated the visual arrays. This form of statistical learning would allow the participant to generate perceptual expectations about the likely location and dispersion of items on each trial, resulting in a more efficient representation of arrays drawn from this distribution (Simoncelli & Olhausen, 2001; Chalk, Seitz, & Series, 2010; Michael & de Gardelle, 2014). Such an internal model of stimulus statistics does not prescribe how

information about individual item locations should be combined in order to make a response on any given trial. However, it may nevertheless exert an influence on how apparent outliers are weighted in the centroid computation, that is, by discounting these items that appear unlikely to have arisen from the same distribution as the others (Haberman & Whitney, 2010). In this scenario, vector error feedback may serve to correct for such a tendency to discount outliers by increasing the internal estimate of the variability in item locations. This account has intuitive appeal, as it posits that learning occurs by tuning a higher-order statistical representation that mirrors the demands of the task (i.e., estimating summary statistics from randomly sampled items). However, as an account of error-driven learning, it is complicated by the fact that the correspondence between trial-by-trial vector error feedback and updating of these model parameter estimates is underdetermined. Future studies employing more complex stimulus distributions may elucidate how learning such statistics interacts with feedback.

Although we interpret our findings as supporting the notion that spatially precise sensory-error feedback enhanced the underlying centroid representation, these results are also consistent with the possibility that such transfer may only occur in the case of perceptual tasks requiring a discrete response, as in the quadrant task (Experiments 1 and 2) and the grid task (Experiment 4). Indeed, although participants do select a particular grid cell when making their response, their responses carried spatial precision that was on the order of what they were able to select with their fingertip in the pointing task. Thus, the grid task provided quasi-continuous response data that were well suited for measuring continuous changes to the fidelity of the centroid representation, irrespective of the discrete nature of grid cell selection.

What kinds of statistical summary representations are most responsive to training? In this study, we focused on how feedback and task practice influenced averaging over a simple visual feature, namely, location. However, it has been previously shown that people are also able to extract averages for complex visual features (Haberman & Whitney, 2007). There is recent evidence that multiple distinct mechanisms might underlie the ability to extract low-level and high-level summary features (Haberman, Brady, & Alvarez, in press). The relative plasticity of different statistical summary representations may be constrained by the type of mechanism that is responsible for performing these summary operations.

In the case of low-level summary perception, these features may be readily and automatically computed (e.g., mean orientation) at a coarse level of resolution, but acuity may be enhanced based on the actual statistics and range of feature values actually experienced. However, in the case of high-level summary perception, experience and feedback may be especially critical for shaping these summary representations over time. Characterizing the type of feedback and training required to induce improvement in the fidelity of statistical summary representations of abstract, complex visual features would be a fruitful avenue for future research.

Conclusions

Overall, there are two main implications of our findings. First, they suggest that statistical summary representations are not merely an automatic consequence of visual experience, but

can be tuned based on external feedback to further optimize performance given specific perceptual goals. Moreover, they show that changes in the fidelity of these perceptual representations may be dependent on the type of sensory error information available during learning, suggesting constraints on the mechanisms by which such tuning may occur.

Acknowledgements

References

Ajne B. A simple test for uniformity of a circular distribution. Biometrika. 1968; 55:343–354.

Alvarez GA. Representing multiple objects as an ensemble enhances visual cognition. Trends in Cognitive Sciences. 2011; 15(3):122–131. [PubMed: 21292539]

Alvarez GA, Oliva A. The representation of simple ensemble visual features outside the focus of attention. Psychological Science. 2008; 19(4):392–398. [PubMed: 18399893]

Ariely D. Better than average? When can we say that subsampling of items is better than statistical summary representations? Perception & Psychophysics. 2008; 70(7):1325–1326. [PubMed: 18927014]

Ball K, Sekuler R. Direction-specific improvement in motion discrimination. Vision Research. 1987; 27(6):953–965. [PubMed: 3660656]

Bauer B. The danger of trial-by-trial knowledge of results in perceptual averaging studies. Attention, Perception, & Psychophysics. 2009; 71(3):655–665.

Chalk M, Seitz AR, Series P. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision. 2010; 10(8):1–18.

Chong SC, Treisman A. Representation of statistical properties. Vision Research. 2003; 43(4):393–404. [PubMed: 12535996]

Dosher BA, Lu Z-L. Hebbian Reweighting on Stable Representations in Perceptual Learning. Learning & Perception. 2009; 1(1):37–58. [PubMed: 20305755]

Fisher, NI. Statistical Analysis of Circular Data. Cambridge University Press; 1995. Revised edition

Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. Neuron. 2001; 31(5):681–697. [PubMed: 11567610]

Goodale MA, Milner AD. Separate visual pathways for perception and action. Trends in Neurosciences. 1992; 15(1):20–25. [PubMed: 1374953]

Haberman J, Brady TF, Alvarez GA. Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. Journal of Experimental Psychology: General. 2015; 144(2):432–446. [PubMed: 25844624]

Haberman J, Whitney D. Rapid extraction of mean emotion and gender from sets of faces. Current Biology. 2007; 17(17):R751–R753. [PubMed: 17803921]

Haberman J, Whitney D. The visual system discounts emotional deviants when extracting average expression. Attention, Perception, &Psychophysics. 2010; 72:1825–1838.

Held R, Freedman SJ. Plasticity in human sensorimotor control. Science. 1963; 142(3591):455–462. [PubMed: 14064442]

Held R, Gottlieb N. Technique for studying adaptation to disarranged hand-eye coordination. Perceptual and Motor Skills. 1958; 8(3):83–86.

Herzog MH, Fahle M. The role of feedback in learning a vernier discrimination task. Vision Research. 1997; 37(15):2133–2141. [PubMed: 9327060]

Izawa J, Shadmehr R. Learning from Sensory and Reward Prediction Errors during Motor Adaptation. PLoS Computational Biology. 2011; 7(3):e1002012. [PubMed: 21423711]

Karni A, Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proceedings of the National Academy of Sciences of the United States of America. 1991; 88(11):4966–4970. [PubMed: 2052578]

Karni A, Sagi D. The time course of learning a visual skill. Nature. 1993; 365(6443):250–252. [PubMed: 8371779]

Kording KP, Wolpert DM. Bayesian integration in sensorimotor learning. Nature. 2004; 427(6971): 244–247. [PubMed: 14724638]

Kording KP, Wolpert DM. The loss function of sensorimotor learning. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101(26):9839–9842. [PubMed: 15210973]

Liu J, Dosher B, Lu Z-L. Modeling trial by trial and block feedback in perceptual learning. Vision Research. 2014; 99((C)):46–56. [PubMed: 24423783]

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997; 390(6657):279–280. [PubMed: 9384378]

MATLAB and Statistics Toolbox Release. The MathWorks, Inc. United States: Natick, Massachusetts; 2013a.

Michael E, de Gardelle V. Priming by the variability of visual information. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(21):7873–7878. [PubMed: 24821803]

Mishkin M, Ungerleider L, Macko K. Object vision and spatial vision: Two cortical pathways. Trends in Neurosciences. 1983:414–417.

Myczek K, Simons DJ. Better than average: Alternatives to statistical summary representations for rapid judgments of average size. Perception & Psychophysics. 2008; 70(5):772–788. [PubMed: 18613626]

Nikooyan AA, Ahmed AA. Reward feedback accelerates motor learning. Journal of Neurophysiology. 2015; 113:633–646. [PubMed: 25355957]

O'Doherty JP. Reward representations and reward-related learning in the human brain: insights from neuroimaging. Current Opinion in Neurobiology. 2004; 14(6):769–776. [PubMed: 15582382]

Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M. Compulsory averaging of crowded orientation signals in human vision. Nature Neuroscience. 2001; 4(7):739–744. [PubMed: 11426231]

Pearce JM, Hall G. A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review. 1980; 87(6):532–552. [PubMed: 7443916]

Petrov AA, Dosher BA, Lu Z-L. The dynamics of perceptual learning: an incremental reweighting model. Psychological Review. 2005; 112(4):715–743. [PubMed: 16262466]

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0, URL http://www.R-project.org/

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II: Current Research and Theory. 1972; 2:64–99.

Salmoni AW, Schmidt RA, Walter CB. Knowledge of results and motor learning: a review and critical reappraisal. Psychological Bulletin. 1984; 95(3):355. [PubMed: 6399752]

Saslow MG. Latency for saccadic eye movement. Journal of the Optical Society of America. 1967; 57(8):1030–1033. [PubMed: 6035297]

Solomon J, Morgan M, Chubb C. Efficiencies for the statistics of size. Journal of Vision. 2011; 11(11): 1041–1041.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275(5306):1593–1599. [PubMed: 9054347]

Seitz AR, Nanez JE, SrHolloway S, Tsushima Y, Watanabe T. Two cases requiring external reinforcement in perceptual learning. Journal of Vision. 2006; 6(9):966–973. [PubMed: 17083288]

Shadmehr R, Smith MA, Krakauer JW. Error correction, sensory-prediction, and adaptation in motor control. Annual Review of Neuroscience. 2010; 33:89–108.

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. Annual Review of Neuroscience. 2001; 24(1):1193–1216.

Synofzik M, Thier P, Lindner A. Internalizing agency of self-action: perception of one's own hand movements depends on an adaptable prediction about the sensory action outcome. Journal of Neurophysiology. 2006; 96(3):1592–1601. [PubMed: 16738220]

Taylor JA, Hieber LL, Ivry RB. Feedback-dependent generalization. Journal of Neurophysiology. 2012; 109(1):202–215. [PubMed: 23054603]

Tseng YW, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory-prediction Errors Drive Cerebellum-Dependent Adaptation of Reaching. Journal of Neurophysiology. 2007; 98(1):54–62. [PubMed: 17507504]

Watson AB, Pelli DG. QUEST: A Bayesian adaptive psychometric method. Perception & Psychophysics. 1983; 33(2):113–120. [PubMed: 6844102]

Zar, JH. Biostatistical Analysis. 4th edition. Prentice Hill; 1999.

Zhang W, Luck SJ. Discrete fixed-resolution representations in visual working memory. Nature. 2008; 453(7192):233–235. [PubMed: 18385672]
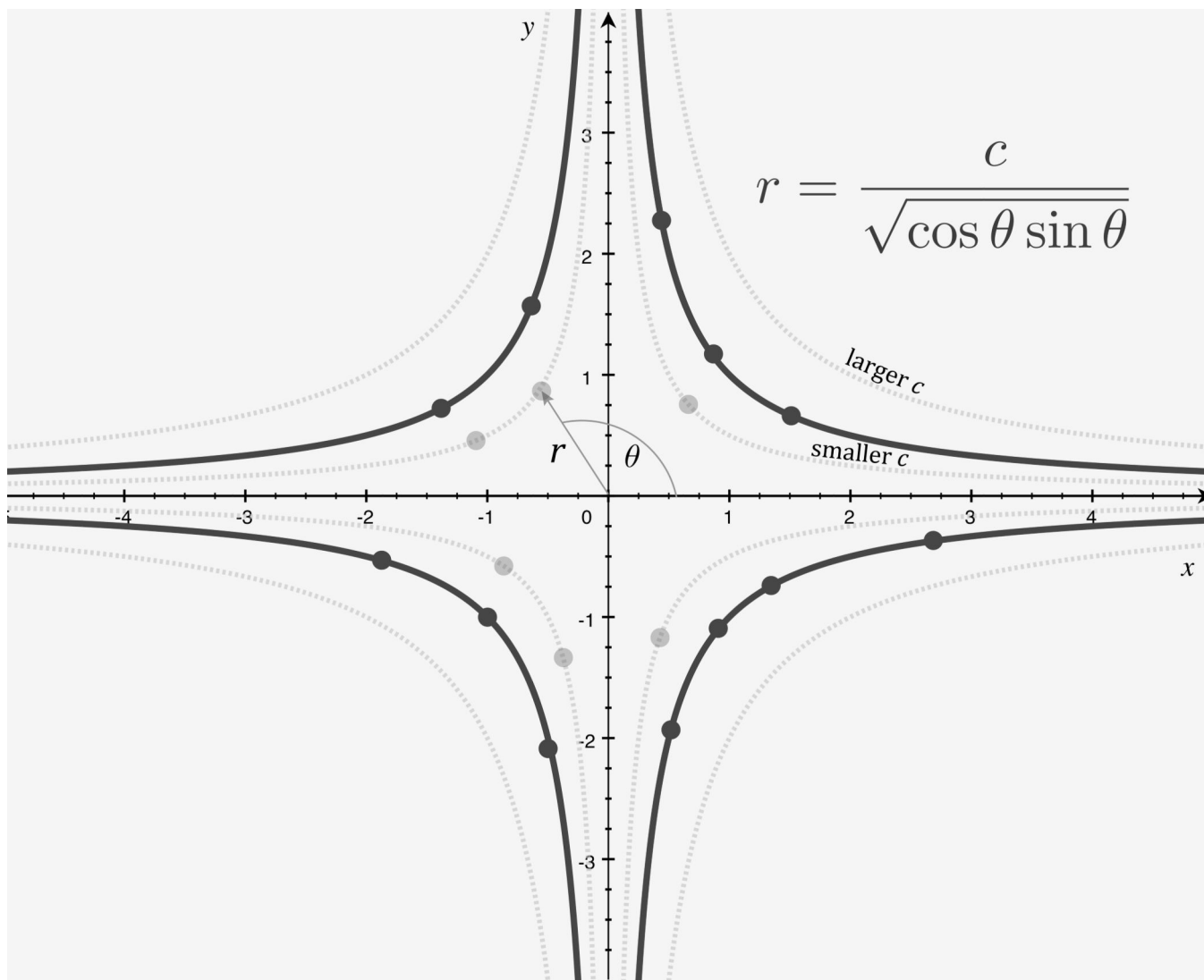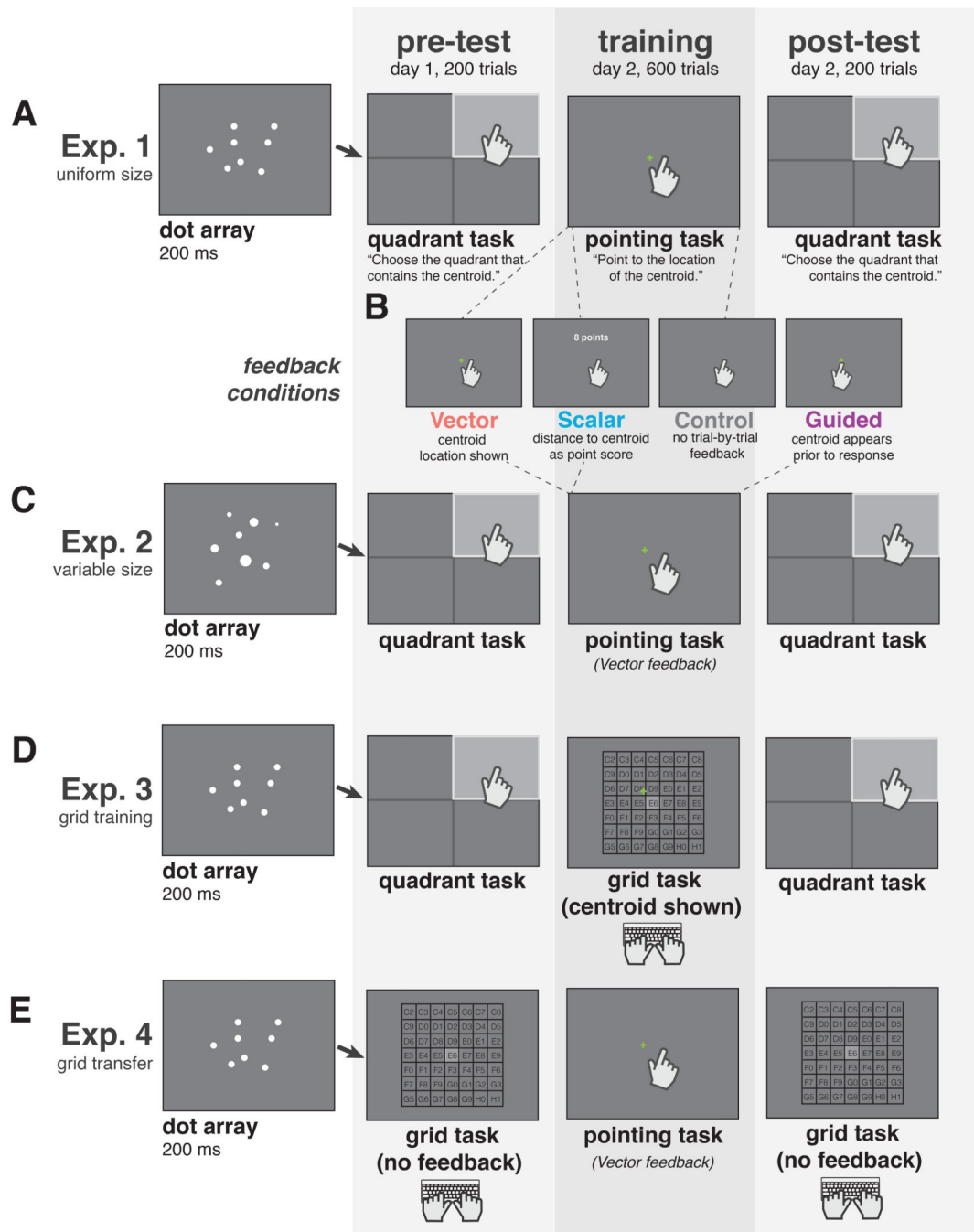
Figure 1.
Distribution of centroid locations in the quadrant task followed a hyperbolic function:
$r = c/\sqrt{cos(\theta)sin(\theta)}$, where r is the radial distance from the center of the display, θ represents randomly sampled angles, and c is a scaling factor that controls the overall distance to the nearest quadrant boundaries. Smaller values of c produce displays with centroids falling closer to boundaries, thus making discrimination more difficult.

Figure 2.

Task and procedure. (A) Eight dots of uniform size presented for 200ms in random locations on each trial. During training, participants pointed to the centroid on touchscreen. During the pre-test and post-test, participants discriminated which quadrant contained the centroid. (B) Groups received different trial-by-trial feedback during training: true centroid location shown after response (Vector), point score reflecting distance to centroid (Scalar), no feedback (Control), or centroid location shown prior to responding (Guided). (C) In Experiment 2, dots appeared in random sizes and locations. (D) During training in

Experiment 3, participants typed in the code of the grid cell containing the centroid, and then were presented with the location of the centroid. (E) Participants in Experiment 4 performed the pointing task during training while receiving vector error feedback; the grid task was used to obtain a more precise measure of changes in perceptual discrimination.
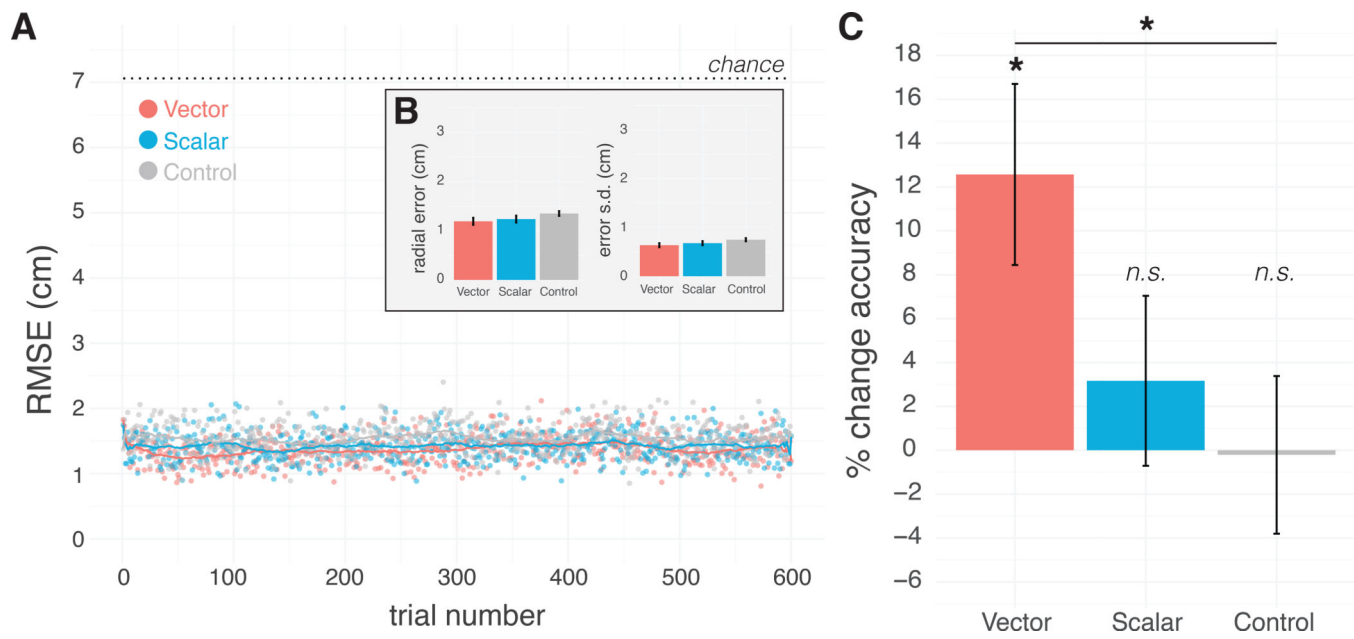
Figure 3.
Experiment 1 Results. A) Overall performance (RMSE) on the pointing task during the training phase as a function of trial number for the Vector (Salmon), Scalar (Blue), and Control (Gray) groups. All groups performed better than chance (dashed line). B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, ~<0.1.
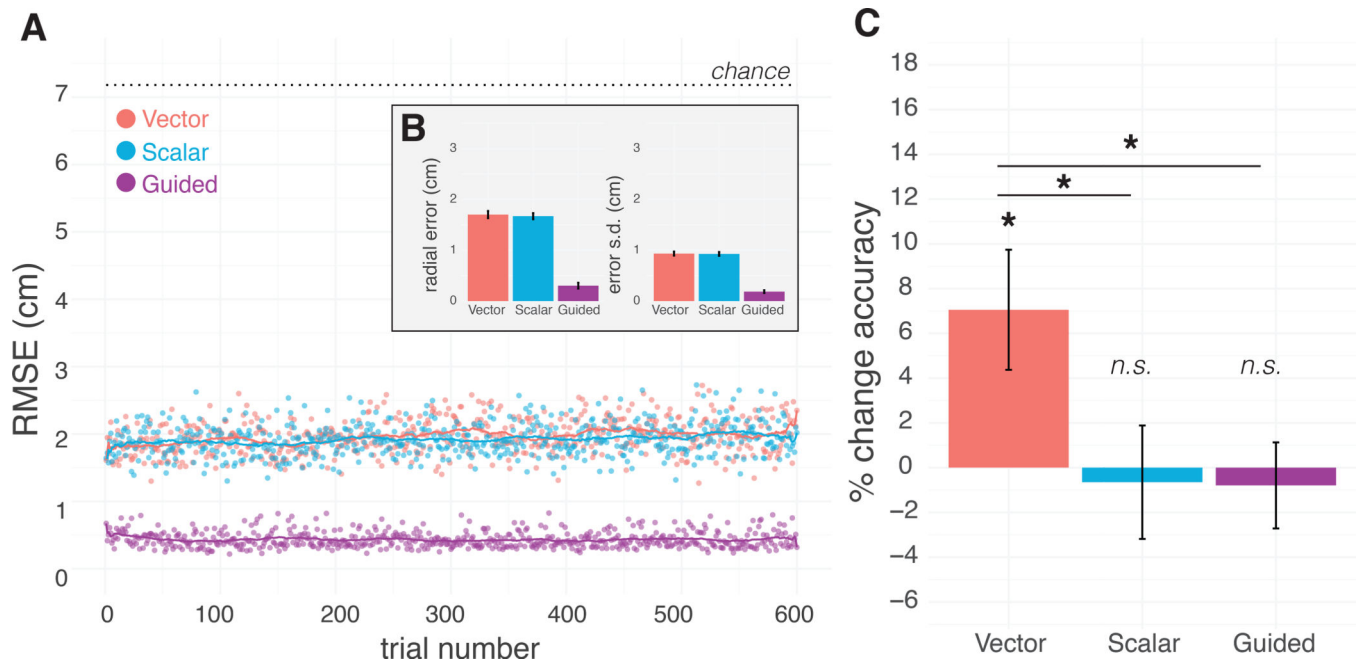
Figure 4.
Experiment 2 Results. A) Overall performance (RMSE) on the pointing task as a function of trial number for the Vector (Salmon), Scalar (Blue), and Guided (Purple) groups. All groups performed better than chance (dashed line), with Guided participants making substantially smaller errors. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, ~<0.1.
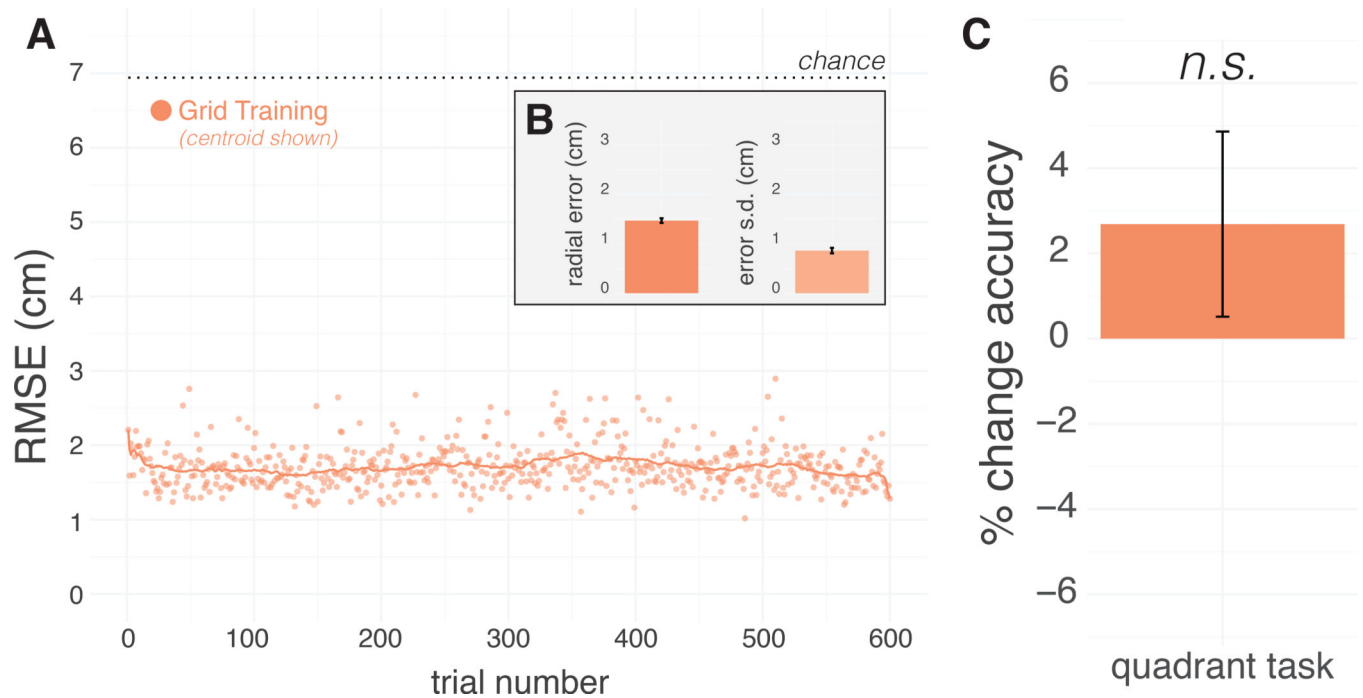
Figure 5.
Experiment 3 Results. A) Overall performance (RMSE) greatly exceeded chance, where continuous error computed as distance between center of selected grid cell and the centroid. B) Performance in the grid task broken down into accuracy (radial error) and precision (standard deviation). C) Percent change in accuracy between post-test and pre-test in the quadrant task. Error bars represent 1 SEM *<0.05, ~<0.1.
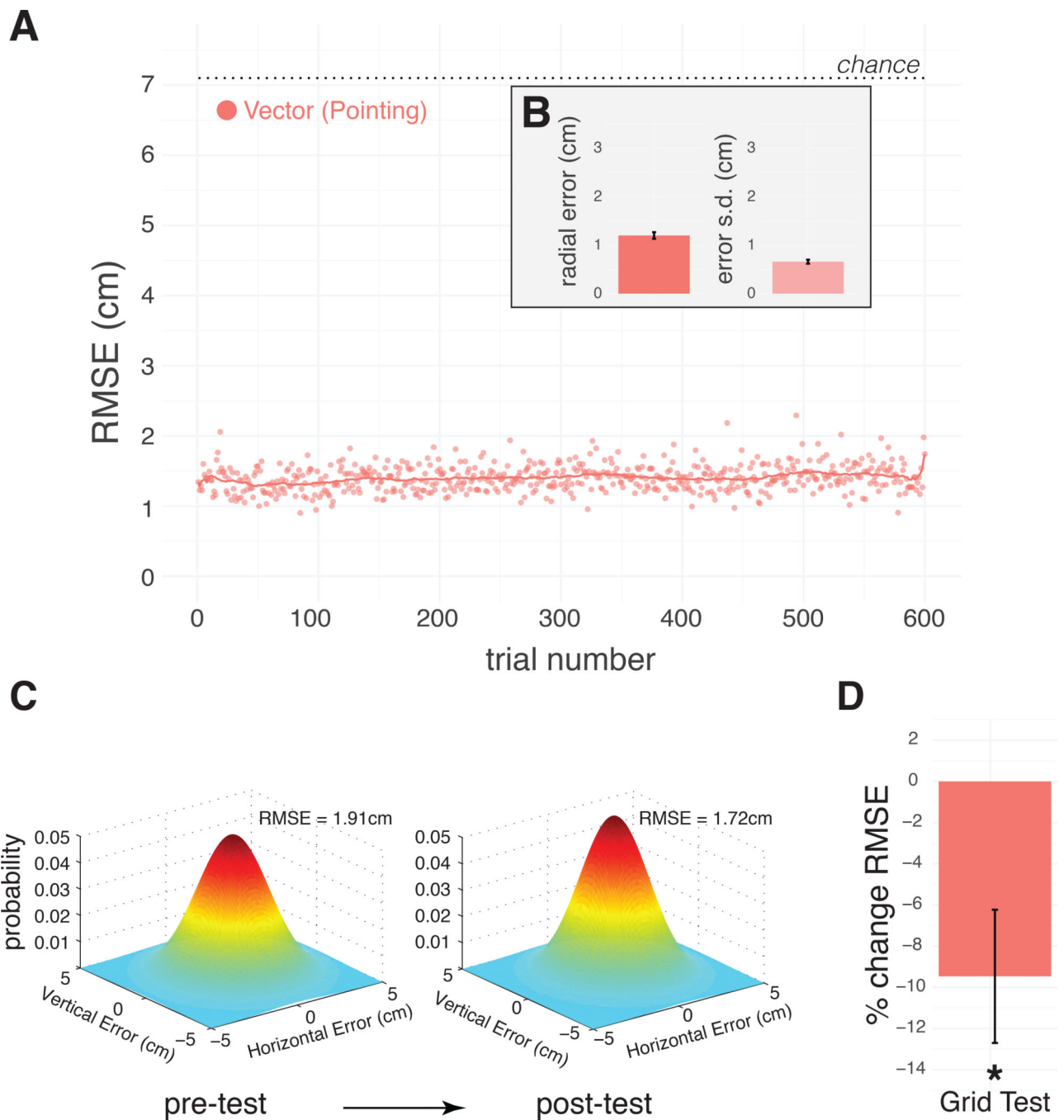
Figure 6.
Experiment 4 Results. (A) Overall performance (RMSE) greatly exceeded chance on the pointing task with Vector feedback. B) Performance in the pointing task broken down into accuracy (radial error) and precision (standard deviation). (C) Error distribution in the pre-test and post-test modeled as bivariate Gaussian distribution fit to participant errors. (D) Percent change in grid task performance (RMSE) between post-test and pre-test; negative value reflects enhanced fidelity. Error bars represent 1 SEM *<0.05, ∼<0.1.

Table 1

Median response time (SEM) by condition and phase, in milliseconds.

		Pre-test	Training	Post-test
E1	**Vector**	1217 (*158*)	761 (*109*)	812 (*86*)
	Scalar	1031 (*105*)	771 (*54*)	843 (*99*)
	Control	884 (*83*)	673 (*56*)	737 (*58*)
E2	**Vector**	923 (*113*)	755 (*71*)	837 (*87*)
	Scalar	887 (*73*)	913 (*90*)	829 (*54*)
	Guided	698 (*50*)	741 (*56*)	709 (*85*)
E3	**Grid Training**	880 (*84*)	1280 (*69*)[*]	796 (*67*)
E4	**Vector**	1727 (*129*)[*]	852 (*70*)	1465 (*102*)[*]

[*] Latency to first keystroke in grid task.