A neurophysiological explanation for biases in visual localization

James C. Moreland[1] · **Geoffrey M. Boynton**[1]

Abstract Observers show small but systematic deviations from equal weighting of all elements when asked to localize the center of an array of dots. Counter-intuitively, with small numbers of dots drawn from a Gaussian distribution, this bias results in subjects overweighting the influence of outlier dots – inconsistent with traditional statistical estimators of central tendency. Here we show that this apparent statistical anomaly can be explained by the observation that outlier dots also lie in regions of lower dot density. Using a standard model of V1 processing, which includes spatial integration followed by a compressive static nonlinearity, we can successfully predict the finding that dots in less dense regions of an array have a relatively greater influence on the perceived center.

Keywords Summary statistics · Localization · Vision · Mid-level mechanisms · Robust statistics

Introduction

Much of our environment consists of "stuff" that contains complex statistical structure. The human visual system is exquisitely sensitive to the summary statistics that describe this structure across a wide variety of domains that includes orientation, size, location, speed, and facial expression (Albrecht & Scholl, 2010;

✉ James C. Moreland
   jamesm37@uw.edu

[1] Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525, USA

Alvarez & Oliva, 2009; Ariely, 2001; Chong & Treisman, 2005; Hubert-wallander & Boynton, 2015; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). What is less understood is the mechanisms by which these summary representations are formed.

Here we examine how the statistics of a sample's distribution influence the weighting of individual items in the context of one of the most commonly studied summary statistics – the ability of individuals to extract the mean or central tendency of a group of similar objects (Alvarez & Oliva, 2008; Greenwood, Bex, & Dakin, 2009; Hubert-Wallander & Boynton, 2015). Findings that representations of the group are more accurate than for any individual element has led many to conclude that the process appears to include all items in the group (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014; Ariely, 2001; Chong & Treisman, 2005; Haberman & Whitney, 2007; Oriet & Brand, 2013; Robitaille & Harris, 2011). It remains, however, an open question what contribution each element makes, and how the visual system arrives at a summary of these elements.

Some studies have shown that elements that deviate far from the group mean contribute less to the mean estimate than those that are more similar to the mean, consistent with the idea of a "robust estimator" that treats outliers as being less reliable. Such "robust weighting" has been shown both for category judgments of shape and color (de Gardelle & Summerfield, 2011), and for estimates of the centroid of a set of dots (Juni, Singh, & Maloney, 2010). However, other studies have shown the opposite finding, despite using quite similar stimuli. When observers were asked to saccade to the center of an array of dots they showed biases towards regions with *fewer* dots (McGowan, Kowler, Sharma, & Chubb, 1998).

In our study, subjects estimated the center of a two-dimensional array of dots using a procedure similar to that of McGowan et al.'s (1998) but with a mouse click response rather than a saccade. We thought it possible that explicit judgments

(Juni, Singh, & Maloney, 2010) and saccades (McGowan et al., 1998) might be mediated by different estimates of the center location and could explain the differing results. To further examine any possible dissociation between these measures we also collected free viewing eye movements alongside explicit mouse click responses. The prediction made from the previous literature would be that eye movements would be biased towards more isolated dots and click responses would show equal weighting. Our results replicated those of McGowan et al. for both clicks and fixations – dots that lay further from other dots were weighted more heavily than those in denser regions. A second experiment serves to rule out the possibility of a more object-based model where it is the number of items per-unit-space in favor of a perceptually-based filtering model.

Such results cannot easily be explained as being mediated by the process of statistical estimation per se – sensible statistical estimators of central tendency (mean, median, mode, trimmed mean, and so forth) universally weight outliers *less* heavily. We suggest here that "anti-robust" biases in statistical estimation may reflect processing within an earlier stage of processing, and show that our results can be explained using a simple model where estimates of the mean are generated through simple unbiased averaging of nonlinear V1 responses that are themselves influenced by dot density.

Experiment 1

Methods

Participants

Fifteen students from the University of Washington were recruited from the Department of Psychology. All received payment of US$20 for their participation. Data collection for each subject was completed in under 60 min, across two sessions separated by a minimum of 4 h. All participants had normal or corrected-to-normal vision. Recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

Analysis of simulated data using variance based on two similar studies (Juni et al., 2010 - see Supplementary Material; McGowan et al., 1998) suggested that ten subjects would be sufficient to detect the predicted effect and would replicate previous findings. We set out to collect 15 observers to ensure sufficient power given that we expected to exclude trials due to failures in eye-tracking.

Apparatus and materials

Observers sat in a dimly lit room, 50 cm from a CRT monitor subtending $40.4° \times 30.8°$ of visual angle. Two-dimensional movements of one eye were recorded by an ASL Eye-Trac® 6 at ~100 Hz. The subject's other eye remained uncovered. A chin and forehead rest was used to stabilize the head position and maintain the distance from the screen. All stimuli were generated by custom software written using Psychophysics Toolbox (Brainard, 1997) for MATLAB.

Procedure

A schematic of the task design is shown in Fig. 1. Participants viewed a fixation cross in the center of the screen for a period that varied randomly between 500 and 1,000 ms. Then ten dots (radius of 0.3°) appeared simultaneously. Participants were free to move their eyes as soon as the dots appeared. After 300 ms the dot array was removed. After a further 300 ms a cursor appeared in a random location on the screen and participants made their response by moving the cursor with the mouse and clicking on the location they perceived to be the mean of the dots shown.

The instructions to the participant were to maintain fixation on the cross until it disappeared from the screen and to click on the location they thought was the center of the dots shown. It was made clear that this was not a sample from a larger population whose mean they were estimating.

A response had to be made within 2,000 ms of the cursor appearing or the trial would be discarded and a warning displayed. In order to make it clear that the saccade was not an explicit response, both eye movement data and click responses were collected within the same trials.

For each trial, the dot locations were sampled from a bivariate Gaussian probability distribution with a standard deviation of 2.3°. The center of the bivariate Gaussian distribution was drawn randomly on each trial from a uniform rectangular region, subtending the inner 70% of the full screen ($20.2° \times 15.1°$). If a dot's sampled location was outside the borders of the screen, it was resampled until it was within the full screen region.

Participants completed practice trials before beginning the experimental trials and were given the opportunity to clarify any instructions. Each participant completed 8–10 blocks of 50 trials with eye-tracking recalibration between each block.

Eye-tracking inclusion criteria

Fixations were defined as periods of eye position velocity below .015°/s for interpolated first derivatives of the eye position.

Trials were excluded if there were disturbances in the eye-tracking data collection such as no detected fixations, eye-tracking locations beyond the screen area or periods of no eye-tracking (including blinks). Trials were also excluded if the first fixation was not within 1.5° visual angle of the
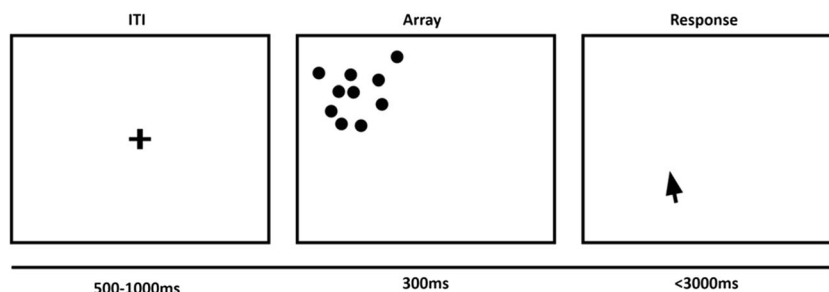
Fig. 1 Trial schematic. Participants were instructed to maintain fixation on the cross until the ten dots appeared. The dots were presented for 300 ms. Participants made their response by moving the cursor to the location they perceived to be the mean dot location. A faint grid of dots were always present to help participants localize objects on the screen and minimize effects of other non-uniform visual landmarks such as screen edges (e.g., Deubel, 2004)

fixation cross at the beginning of the trial, if the second fixation occurred before 200 ms or after 600 ms and if the second fixation was made further than 7° visual angle from the trial array center (based on the dot distribution SD = 2.3 this would include almost all possible dots).

A number of subjects struggled to consistently meet these criteria and were excluded. In total, five of 15 subjects were excluded from *all* analyses on this basis, leaving ten subjects with 300–500 trials each.

Results

Our measures of interest were the click response reporting the subjects' explicit estimate of the center of the dots with a mouse click, and the location of fixation at the end of the first saccade (note that subjects were not specifically told to fixate on the center of the dot array). Both responses consistently landed near the mean of dot arrays. This can be seen in Fig. 2 where the response location is centered on the mean of the dot array. Error in locating the mean of the array was smaller in the click responses (M = 0.64°, SD = 0.43°) than the first fixation (M = 1.22°, SD = 0.82°). There was no difference in size of error dependent on the eccentricity of the trial (Click: t(9) = 0.58, p = .58, 95% CI [−0.001, 0.002]; Fixation: t(9) = 0.80, p = .44, 95% CI [−0.003, 0.006]). All subjects showed an undershoot in the saccades with a group mean proportion of 0.91 (SD = 0.23) compared with a group click mean proportion of 1.00 (SD = 0.14). This can be seen in Fig. 2b where the distribution of fixation responses is shifted to the left of the origin. These undershoots are consistent with previous saccade to target literature (e.g., McGowan et al., 1998).

Dot weighting as a function of distance from the true mean

We used maximum likelihood estimation to estimate the weight given to dots as a function of the distance from the true mean. Dots were sorted into distance bins with edges defined using a cumulative normal distribution so that each bin contained, on average, the same number of dots.

We then calculated the average weight applied to each dot within a bin by assuming that the perceived center of the array on trial t is the average of the individual dot locations, weighted by how far each dot fell from the mean. Let $X_t = \{x_{1t}, x_{2t}, \ldots, x_{10t}\}$ and $Y_t = \{y_{1t}, y_{2t}, \ldots, y_{10t}\}$ be the X and Y positions of the 10 dots for trial t, and let the weights for the bins be $w[1]$, $w[2]$, …, $w[nbins]$. The perceived center $G(X)$, $G(Y)$ for trial t is then predicted by:

$$G(w, X_t) = \frac{\sum_{i=1}^{10} x_{it} w[\beta_{it}]}{\sum_{i=1}^{10} w[\beta_{it}]} \text{ and } G(w, Y_t)$$

$$= \frac{\sum_{i=1}^{10} y_{it} w[\beta_{it}]}{\sum_{i=1}^{10} w[\beta_{it}]}$$

where β_{it} is the index $[1, 2, \ldots, nbins]$ for the bin that dot i falls in on trial t. If the weights are all equal, then the model predicts that subjects will estimate the center of the dots to be located at the true center of gravity of the dots:

$$G(X_t) = \frac{\sum_{i=1}^{10} x_{it}}{10} \text{ and } G(Y_t) = \frac{\sum_{i=1}^{10} y_{it}}{10}$$

The best-fitting weights were found using a maximum likelihood method that assumed that the variability in the subject's responses has a bivariate normal distribution and is independent across trials (supported by observing descriptive plots of the responses, Fig. 2a and b) (Wilks, 2011).

While the slope of the best-fitting weights for the click responses appears qualitatively to increase as a function of distance, the deviation from equal weighting is just shy of statistical significance (Mean slope = 0.089, SD = 0.12, t(9) = 2.18, p = .057; Fig. 3b). Equal weighting as a function of distance is more apparent with the fixation response. The location of the first fixation does not deviate significantly from the flat line prediction of a weighting of 1 regardless of dot distance (Mean slope =−0.003, SD = 0.13, t(9) = −0.08, p = .94; Fig. 3c).
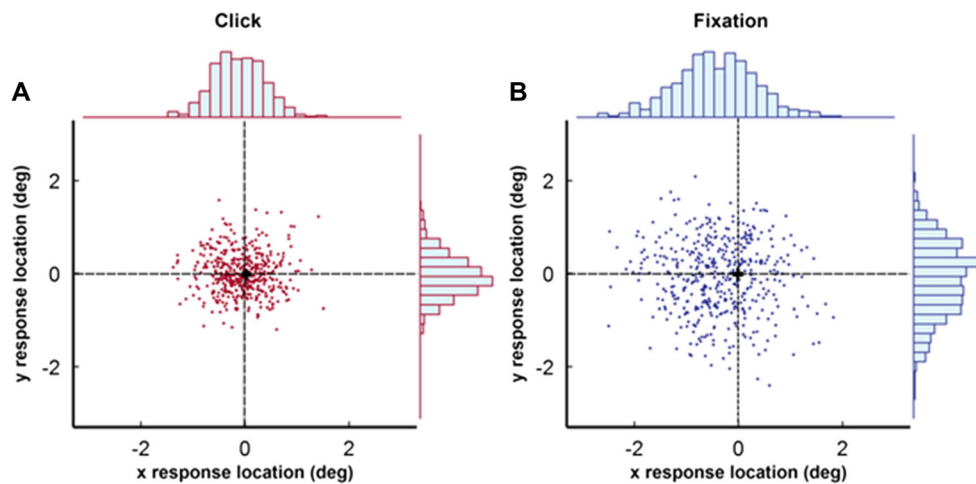
Fig. 2 Scatterplots showing responses from one representative subject for the (**a**) mouse-click and (**b**) first fixation. Responses for each trial are rotated so that the true mean is to the right of fixation and normalized so that the true mean is at the origin. The central black cross indicates (0,0)

Dot weighting as a function of dot proximity

Because our dots were drawn from a Gaussian distribution, density falls off with distance from the sample mean. Thus, an alternative explanation for our findings is that subjects put more weight on dots that are isolated from other dots. We first tested this by measuring how dot contribution is influenced by the number of other dots in near proximity.

For each trial, dot proximity was measured as the average distance to the other 9 dots. For isolated dots this results in a larger number. This is a parameter free measure of density. Weights as a function of average proximity were then estimated by fitting the model described above, but this time dots were sorted into discrete average proximity bins instead of into discrete distance to the mean bins.

Unlike the distance results we do not see equal weighting as a function of proximity for the click responses (Mean slope = 0.011, SD = .007, t(9) = 4.88, p < .001; Fig. 3e). Greater weight is associated with dots that lie further from other dots. Weights for fixation responses did not deviate from equal weighting (Mean slope = 0.03, SD = 0.14, t(9) = 0.60, p = .56; Fig. 3f).

Dot weighting as a function of density

Average proximity as implemented above is a measure of density over the whole array of ten dots and is therefore a somewhat implausible calculation for the brain. An alternative which can be measured over a more local area implements a linear filter which has been widely observed in the visual system. Our third model takes this approach.

For each trial, dot density was determined by convolving an image of the dot field with a 2-D Gaussian. The standard deviation of the Gaussian used to define density was allowed to be a free parameter for each observer which led to slightly different boundaries of the density bins. Density was then defined as the amplitude of this "density map" at each dot location. An isolated dot takes the lowest possible density of one. Weights as a function of density were then estimated by fitting the model described above, but this time dots were sorted into discrete density bins instead of discrete distance bins. The results reported here were found by repeating the bin fits using the average Gaussian width separately for click responses (M = 0.83, SD = 0.17) and first fixations (M = 0.80, SD = 0.23). We tested a range of Gaussian standard deviations between 0.4° and 2.0° and found the weights to be robust to the choice of Gaussian standard deviation.

Consistent with the proximity results, we do not see equal weighting as a function of density for the click responses (mean slope = −0.10, SD = 0.07, t(9) = −4.54, p = .001; Fig. 3h); lower weights were assigned to dots in high density regions (Fig. 3h). Free viewing data examining the location of the first fixation shows the same pattern of results (mean slope = −0.05, SD = 0.06, t(9) = −2.75, p = .022; Fig. 3i).

A linear-nonlinear model for localization

We next show how our results are consistent with a linear-nonlinear, or *LN* model, in which the perceived center of mass is computed by linear spatial filtering followed by a static compressive nonlinearity. To implement the model we the generated a density map as described above by convolving an image of the dot field for each trial with a 2-D Gaussian. The density map was then passed through an exponential function, $U = V^p$. The perceived center was calculated as the two-dimensional centroid of the modified density map. If p < 1, then V^p a compressive nonlinearity that increases the relative influence of dots in regions with lower density. This model of linear spatial filtering followed by a static non-linearity is consistent with known psychophysical (Legge, 1981; Legge
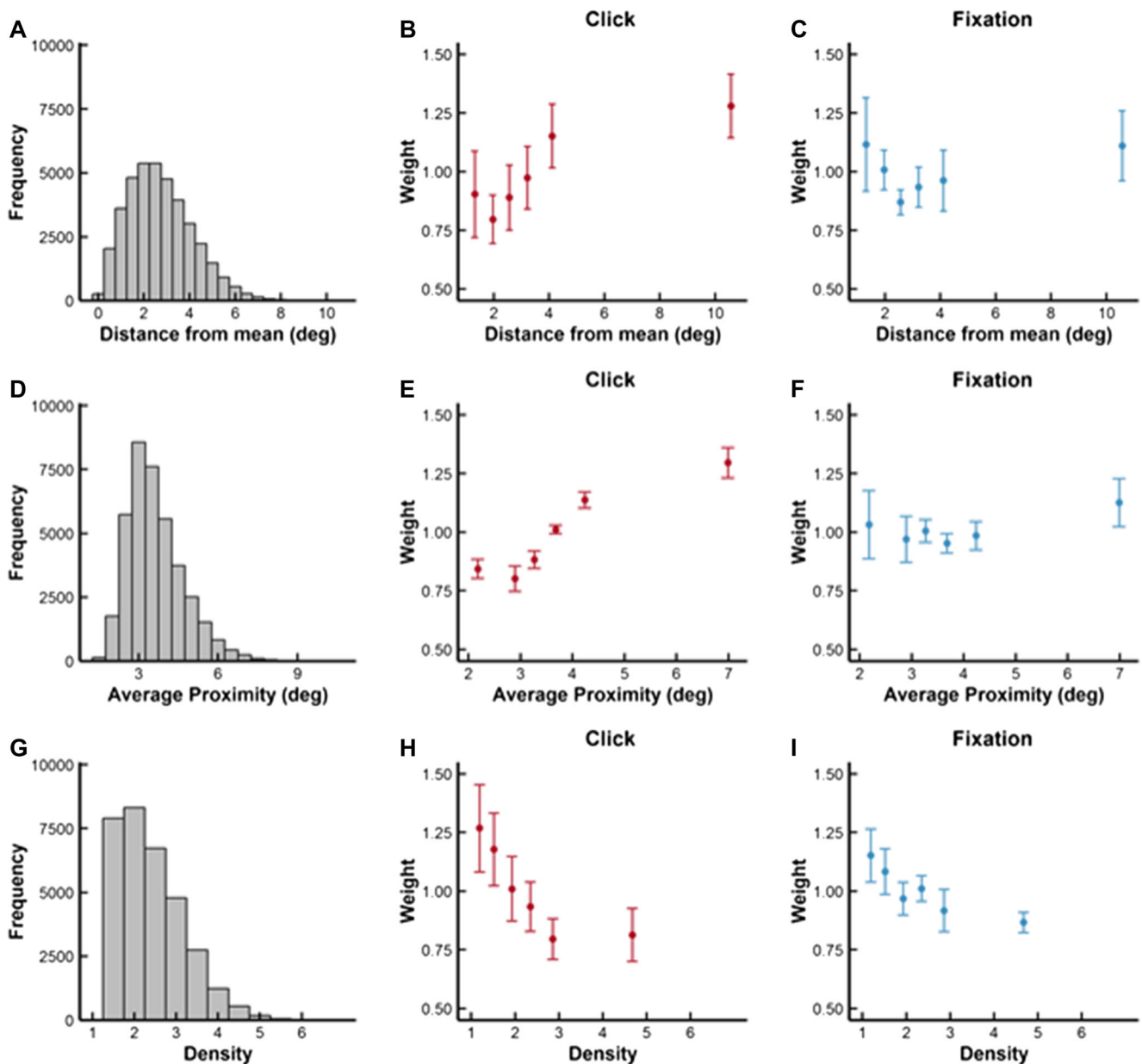
Fig. 3 Estimated weights as a function of distance from the true mean (**a-c**), average proximity (**d-f**), or density (**g-i**). All error bars are standard errors across 10 observers. **a** Histogram of observed dot distances from the true mean. **b** Weights as a function of distance for click responses. The influence of a dot on the estimated center increases as a function of how far it was from the true mean. **c** The binned weighting according to distance for the first fixation as a response. **d** Histogram of average proximity to other dots at each dot location. **e** The binned weighting according to average proximity for click responses. Dots in a trial receive weighting as a function of how far they are on average from other dots. **f** The binned weighting according to average proximity for the first fixation as a response. **g** Histogram of densities at dot locations. **h** The binned weighting according to density for click responses. Dots in a trial receive weighting as a function of how far they are from the trial mean. **i** The binned weighting according to density for the first fixation as a response

& Foley, 1980) and physiological evidence (D. G. Albrecht & Hamilton, 1982), as well as well-established "normalization" models of early visual processing (Carandini & Heeger, 2012; Heeger, 1992).

Figure 4 shows an example of this model for an example set of dots. The leftward panel shows the center-of-mass after linear spatial filtering. The center-of-mass remains identical to the Euclidian mean of the unfiltered dots. The rightward panel shows the center-of-mass after the output of the linear spatial filters is passed through a saturating nonlinearity ($p = 0.5$). The effect is a small shift of the center-of-mass of the image *away* from high density regions.

For each subject we estimated the value of p that minimized the difference between predicted and obtained center
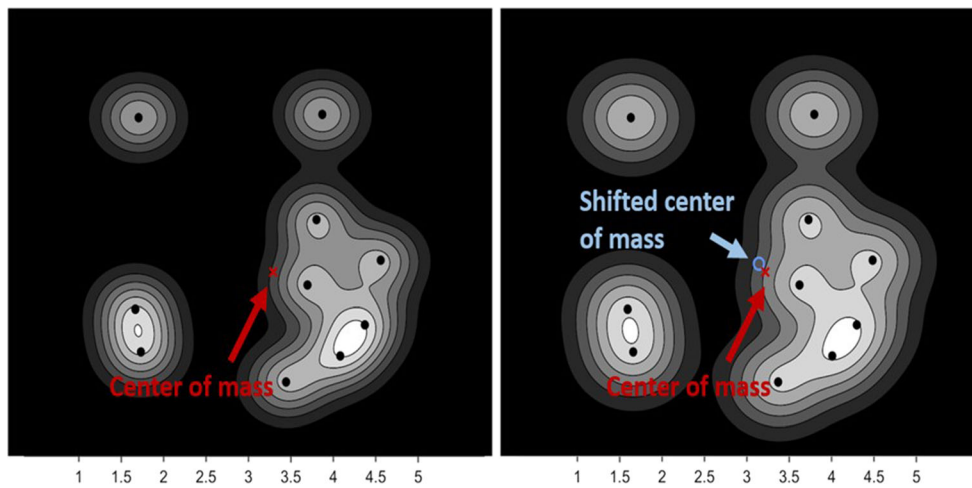
Fig. 4 Model of local detectors as mechanism of localization. **Left**: The Euclidian center of mass calculated after convolving a dot field with a Gaussian. The center-of-mass remains identical to the Euclidean mean.

Right: Passing the output of the linear spatial filters through a saturating nonlinearity results in the center-of-mass being shifted towards regions of the image with lower dot density

of mass estimates. The best fitting value of p ranged between 0.31 and 0.84 for our ten subjects. The mean value of p across subjects was significantly lower than 1 (M = 0.63, SD = 0.16), t(9) = −7.28, p < .001, 95% CI [0.52, 0.75]. Thus, adding a compressive nonlinearity significantly improves our ability to predict subjects' performance.

We next demonstrate the relationship between the *LN* model and the binned distance and density models by simulating responses generated by the *LN* model and fitting the simulated data with the weighted models. If the *LN* model is valid, we expect to find the similar weights as for the observed data.

Responses were generated, for the same dot stimuli that were presented to the observers, using a power function, $U = V^p$, and the values of p found for each observer. Noise was added to the simulated responses using the observers own click response standard deviation for draws from a zero mean 2-D Gaussian.

The weights for the density and distance binning models are shown in Fig. 5. The weights found for the predicted responses closely follow those observed in the behavioral data.

Interim discussion

This first experiment investigated whether the visual system equally weights all elements when localizing a group of dots. We found that observers were not equally weighting all dots. Instead there is support for an overweighting of dots in low density regions as defined by both a parameter free proximity model and a linear filter model. We were able to account for these findings by using a simple *LN* model of early visual processing where a linear spatial filter is followed by a non-linear compression leading to an emphasis on lower densities relative to higher densities.

Experiment 2

Experiment 1 showed that observers were not equally weighting all of the dots when estimating the sample center. These results are consistent with two main classes of models. For the "position-based" model, elements are assigned
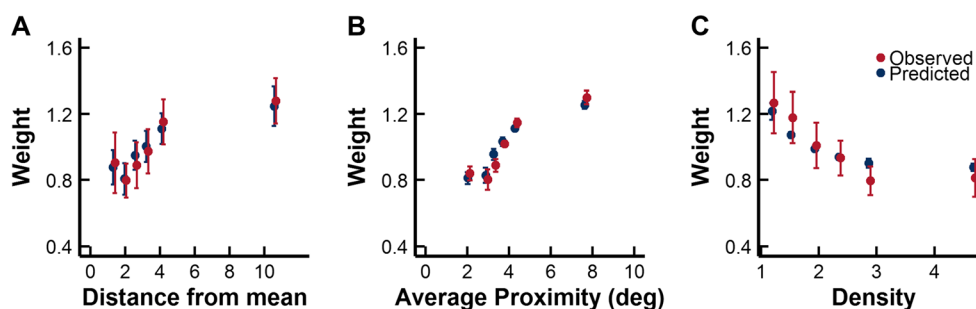
Fig. 5 Results of the *LN* model fit to click responses. Using the same set of stimuli presented to observers, responses were generated using a linear filter then a compressive non-linearity power function using the power values for each observer. Bins are jittered in the x-axis for clarity. Predicted responses are shown in black and observed in red. **a** Predicted response weights for the binned distance model compared to observed response data model. **b** Predicted response weights for average proximity model compared to observed response data model. **c** Predicted response weights for density model compared to observed response data model

weights based entirely on their position relative to other elements, so that outlying and/or dots in regions of low density receive relatively higher weights. For the "perceptually-based" model, the center of mass is computed on the representation of the stimulus after an early stage of perceptual processing. A compressive nonlinearity in the early filtering process leads to an effective overweighing of dots in low-density regions.

Experiment 2 was designed to distinguish between these two classes of models by varying the contrasts of the dots within each array. A spatially-based model only considers dot location and should therefore not be affected by dot contrast. However, the filtering process in the *LN* model should reduce the influence of low-contrast dots on the perceived center.

Methods

Participants

Eight subjects from the University of Washington participated in the experiment. All received payment of US$20 for their participation. Data collection was completed in under 60 min. All participants had normal or corrected-to-normal vision.

Apparatus and materials

The apparatus was the same as for Experiment 1.

Procedure

The procedure was as for Experiment 1, but the dot stimuli were not of uniform contrast. Instead five of the ten dots were presented at 75% (high) contrast, and the other five at 18.75% (low) contrast (See Fig. 6a for an example stimulus). Participants were asked to report the mean location of the dots weighting the presence of a dot equally without regard to the contrast. We did not collect eye movement data. Each of the eight participants completed 12 blocks of 50 trials.

Results and discussion

The average error in locating the trial center was comparable to Experiment 1 (M = 0.70°, SD = 0.41°). A multivariate regression using high and low contrast dots as independent variables found that observers weigh higher contrast dots more than those of lower contrast (mean difference = 0.007, SD = 0.009, t(7) = 2.37, p = .050, 95% CI [0.00, 0.15]; Fig. 6b).

These data support a perceptually-based model that depends at least somewhat on contrast. We tested this hypothesis by using the *LN* model, with the same subject parameters as fit in Experiment 1, to predict the center response and comparing

the weights estimated for high and low contrast dots. Indeed, the results show that high contrast dots are contributing more to the center estimation (Fig. 6b). The effect size is larger in the predicted responses compared with the experimental data. This could be explained by some equal weighting of all the dots with a bias based on contrast. The equal weighting would reduce the effect size of contrast. This stage could operate on the output of the visual systems linear-nonlinear cascade where observers actively attempt to account for all the dots in their estimate.

While contrast may not be the sole characteristic of dot localization, Experiment 2 provides evidence that contrast is an important contributor to the decision and can be accounted for in the *LN* model.

Experiment 3

Experiments 1 and 2 provided evidence that observers are basing their center judgements on the basis of density as defined by contrast rather than proximity. In both experiments, dot positions were drawn from a 2-D Gaussian distribution for which dots in low density regions also tended to fall far away from the geometric mean.

For Experiment 3, we used a "doughnut-shaped" non-Gaussian distribution for which, on average, dot density was maximal at an intermediate distance from the geometric mean. This allowed us to compare weights for dots in low density regions either near or far away from the geometric mean.

Methods

Participants

Fourteen subjects from the University of Washington Psychology Department Subject Pool participated in the experiment. All received course credit for their participation. Data collection was completed in under 60 min. All participants had normal or corrected-to-normal vision.

Apparatus and materials

The apparatus was the same as for Experiments 1 and 2.

Procedure

The procedure was as for Experiment 1, but the dot stimuli were drawn from an alternative distribution. Dots were located by drawing first from a one-dimensional distribution consisting of two overlapping Gaussians with means of ±2° and SDs of .5°. Dots drawn from this distribution were then rotated by a random angle around the origin, and the whole field was then shifted to a new position on the screen based on

A

B
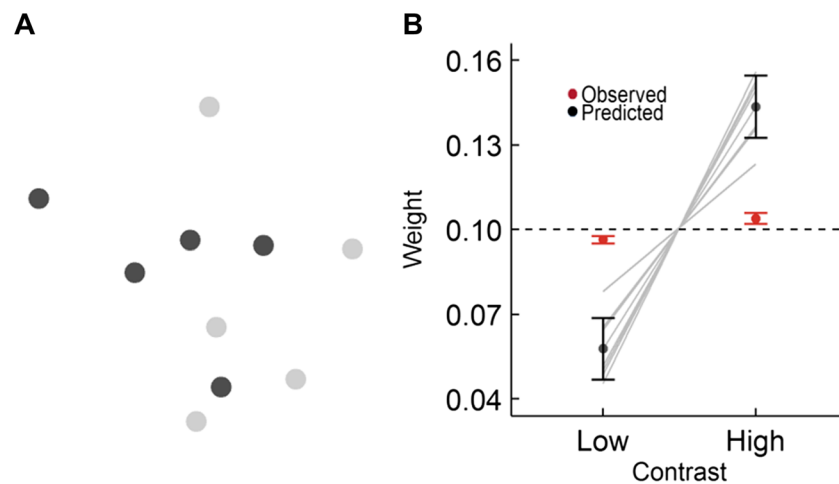
Fig. 6 **a** Schematic representation of the stimuli for Experiment 2. Half of the dots on each trial were of high and low contrast. **b** Weight given to dots in center estimation for low and high contrast dots. In red are the weights from the observed data of Experiment 3. In grey and black are the predicted weights from the *LN* model using the parameters (compressive power and standard deviation of click responses) from Experiment 1

a uniform distribution. Participants were asked to report the mean location of the dots weighting the presence of a dot equally. We did not collect eye movement data. Each of the 14 participants completed 12 blocks of 50 trials.

Results and discussion

The average error in locating the trial center was comparable to Experiments 1 and 2 (M = 0.90°, SD = 0.62°). Using the same maximum likelihood estimation procedure as Experiment 1 we find that each of the models showed a similar qualitative pattern of weighting to that of the Gaussian distribution, although weighting by distance was now also significantly different from zero (Distance: t(13) = 3.30, p = 0.006; Average proximity: t(13) = 6.70, p < .001; Density: t(13) = −7.36, p < .001) (Fig. 7).

The "doughnut" distribution has, on average, a maximal density at an intermediate distance from the geometric mean. This distribution has regions of low density both near and away from the mean and therefore allows for the dissociation between density and distance by examining weights for regions of low density both near and far away from the geometric mean. We binned dots into equal frequency bins according to both distance and density, resulting in a three-by-three matrix crossing low, medium and high density with near, medium and far distances from the mean. A two-way ANOVA found a main effect of distance F(2,117) = 23.45, p < .001, density F(2,117) = 4.10, p = .019, but no interaction of distance and density F(4,117) = 0.169, p = .954. The main effects were expected from the slopes found above and the lack of interaction indicates that low density regions have a greater influence than high density regions not solely because they fall far away from the geometric mean, demonstrating that density has a strong influence on the perceived center of mass.

Discussion

When asked to determine the center of a group of dots, observers show biases that deviate from equal weighting of each element. Dots which lie further from other dots are given more weight. This result is highly counter intuitive because it cannot easily be explained as being mediated by the process of statistical estimation per se. Contrast appears to matter in this density weighting as low contrast dots received less weight than high contrast dots. Finally, we showed that this was not limited to samples drawn from Gaussian distributions by replicating these findings in a "doughnut-shaped" distribution which allowed us to disassociate the contributions of distance and density to weights. Our results show that weights for dots at different distances from the mean are not weighted differently when they fall in regions of similar density.

We therefore conclude that local density is a major factor in influencing a dot's contribution to the perceived center of an array. We suggest that this may be the result of processing within earlier visual areas: our results can be explained using a simple *LN* model consisting of a linear filter followed by a static compressive nonlinearity. This model is consistent with, for example, known psychophysical (Legge, 1981; Legge & Foley, 1980), physiological (Albrecht & Hamilton, 1982) and imaging data (Kay, Winawer, Mezer, & Wandell, 2013) describing V1 processing. It could also be attributed to other earlier areas (e.g., Freeman et al., 2015) but the size of the receptive fields required for the linear stage suggests later areas (Kay et al., 2013; Smith, Singh, Williams, & Greenlee, 2001). Our results do not require that the summary representation itself must be generated at this early sensory stage, rather, the output of this early stage is the available representation to be summarized.
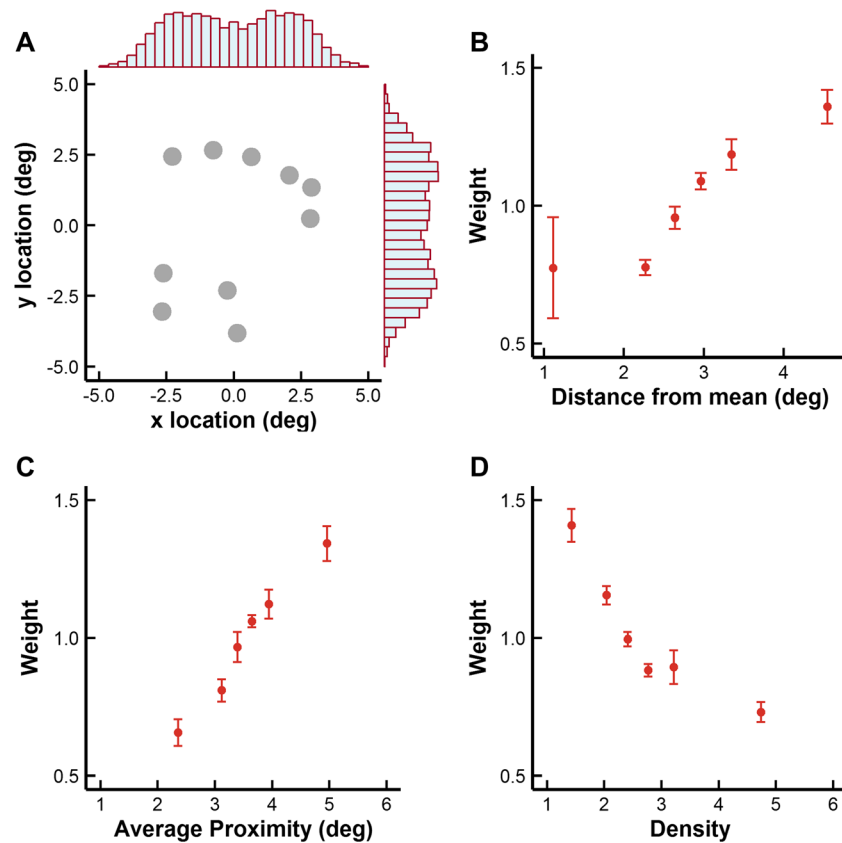
Fig. 7 **a** Example trial from the alternative "doughnut" distribution. Marginal histograms show the dot locations for all the trials for one observer. They follow the overlapping Gaussian distribution they were drawn from. **b** Weights as a function of distance from the trial mean. **c** Weights as a function of average distance to all other dots of the trial. **d** Weights as a function of density

These results are similar to a previous study that had subjects saccade to the estimated center of a field of uniformly distributed dots (McGowan et al., 1998). Our replication of their findings demonstrates that their results were not specific to using a saccade as an explicit response to report the center of mass. Like us, McGowan et al. found that localization showed a bias toward regions of lower dot density. In their Discussion, McGowan et al. (1998) suggested that their results could potentially be accounted for by local detectors covering the array region have a response that depends on the number and intensity of dots in its "receptive field." To explain the overweighting of dots in regions with few other dots the response of a local detector must increase at a successively slower rate (saturate) as dots are added to its "receptive field." This description is very similar to our *LN* model. Furthermore, we have shown that that the density of dots not only depends on an all-or-nothing presence of dots but also on its contrast which is indicative of an early sensory process and rules out later proximity mechanisms.

One previous study (Juni et al., 2010) using samples drawn from a Gaussian distribution found that subjects' performance was consistent with equal weighting of all dots. Simulations of Juni et al.'s analysis suggest that they may not have been able to detect an effect of distance the size of that reported in

Experiment 1 (Supplementary Material). Our model also suggests that with the larger number of dots in the sample (100) they would also not expect to see a strong effect of density as the variation in dot density falls off relatively uniformly as a function of distance in all directions. As a result, the effect of variations in dot density on center-of-mass estimates is negligible with larger numbers of dots. In the Juni et al. study, when mixtures of distributions were used, and subjects were told to ignore noise dots when making their estimates of mean location, subjects behaved robustly – *down weighting* the influence of outlier dots in regions of low density. We suggest that this represents a second stage of robust integration that follows an earlier sensory stage that, under some circumstances, produces anti-robust biases in estimates of central tendency.

The effects of earlier stages of processing may explain some of the apparent differences in how summary statistics are computed across different domains. For example, we have previously shown striking differences in how summary statistics are computed when objects are presented serially over time. Summaries of mean object location were influenced very strongly by early items. Summaries of mean object size, facial expression and motion direction did not show this primacy effect, and were more influenced by later items (Hubert-Wallander & Boynton, 2015). One possibility is that these

results do not reflect differences in how summary statistics are calculated, but rather reflect differential processing at an earlier stage. For example, if spatial attention is rapidly deployed to the location of early-presented dots, then this may enhance the early sensory representations of those (and later) dots that fall near that location, producing a strong primacy effect, without overweighting of early items existing explicitly within the summary statistic calculation.

In conclusion, here we suggest that some summary statistics are *not* computed over items in the world, but over sensory representations of those stimuli. Anything that alters that sensory representation is likely in turn to influence the resulting sensory statistic.

References

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48*(1), 217–237.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*(4), 560–567. doi:10.1177/0956797610363543

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research, 101,* 34–40. doi:10.1016/j.visres.2014.05.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*(4), 392–398. doi:10.1111/j.1467-9280.2008.02098.x.The

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences of the United States of America.*. doi:10.1073/pnas.0808981106

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science: A Journal of the American Psychological Society / APS, 12*(2), 157–162. doi:10.1111/1467-9280.00327

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. doi:10.1163/156856897X00357

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51–62. doi:10.1038/nrn3136

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research, 45*(7), 891–900. doi:10.1016/j.visres.2004.10.004

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13341–13346. doi:10.1073/pnas.1104517108

Deubel, H. (2004). Localization of targets across saccades: Role of landmark objects. *Visual Cognition, 11*(2–3), 173–202. doi:10.1080/13506280344000284

Freeman, J., Field, G. D., Li, P. H., Greschner, M., Gunning, D. E., Mathieson, K., … Chichilnisky, E. J. (2015). Mapping nonlinear receptive field structure in primate retina at single cone resolution. *Elife, 4,* e05241. doi:10.7554/eLife.05241

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 13130–13135. doi:10.1073/pnas.0901352106

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), 751–753.

Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197. Retrieved from http://journals.cambridge.org/abstract_S0952523800009640

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision, 15,* 1–12. doi:10.1167/15.4.5

Juni, M. Z., Singh, M., & Maloney, L. T. (2010). Robust visual estimation as source separation. *Journal of Vision, 10*(14), 1–20. doi:10.1167/10.14.2.Introduction

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex Compressive spatial summation in human visual cortex. *Journal of Neurophysiology, 110,* 481–494. doi:10.1152/jn.00105.2013

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research, 21*(1978), 457–467.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*(12), 1458–1471. doi:10.1364/JOSA.70.001458

McGowan, J. W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Research, 38*(6), 895–909. doi:10.1016/S0042-6989(97)00232-0

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research, 79*(1), 8–16. doi:10.1016/j.visres.2012.12.004

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744. doi:10.1038/89532

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision, 11,* 1–8. doi:10.1167/11.12.18.Introduction

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex (New York, N.Y. : 1991), 11*(12), 1182–1190. doi:10.1093/cercor/11.12.1182

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences.* Academic Press.