Cortical idiosyncrasies predict the subjective experience of object size

Christina Moutsiana[1]*, Benjamin de Haas[1,2]*, Andriani Papageorgiou[1], Jelle A. van Dijk[1,2], Annika Balraj[1,3], John A. Greenwood[1] & D. Samuel Schwarzkopf[1,2]#

[1] Experimental Psychology, University College London, 26 Bedford Way, London, U.K.

[2] UCL Institute of Cognitive Neuroscience, 17 Queen Square, London, U.K.

[3] Columbian College of Arts and Sciences, The George Washington University, 2125 G St. NW, Washington, 20052, U.S.A.

* These authors contributed equally

To whom all correspondence should be directed

Abstract

Perception is subjective. Even basic judgments, like those of visual object size, vary substantially between observers and also across the visual field within the same observer. How the visual system determines the size of objects remains unclear. We hypothesize that size is inferred directly from stimulus representations in V1 and predict that idiosyncrasies in cortical architecture should explain individual differences in size judgments. Indeed, using novel behavioral methods we demonstrate that biases in size perception are related to both the spatial tuning of neuronal populations and local cortical magnification in V1, as measured with functional magnetic resonance imaging in healthy volunteers. Our results indicate that the visual system infers size from V1 representations and that individual perception of even simple stimuli is warped by idiosyncrasies in brain organization.

Introduction

The possibility that the apparent size of objects could be inferred from the cortical representation in area V1 (Schwarzkopf, 2015) is supported by the observation that the spatial spread of neural

activity in this area is related to apparent size under a range of contextual modulations (Murray et al., 2006; Fang et al., 2008; Sperandio et al., 2012; Pooresmaeili et al., 2013). Indeed, both the strength of contextual illusions and the objective discrimination of visual stimuli have recently been linked to the spatial extent of the cortical area in V1 that represents the central visual field (Schwarzkopf et al., 2011; Schwarzkopf and Rees, 2013; Song et al., 2013b). This hypothesis, however, makes an essential prediction that these findings fail to address: tuning properties of local neuronal populations in V1 should explain idiosyncratic biases in *basic size perception* in the absence of contextual illusions.

To test this we developed the Multiple Alternative Perceptual Search (MAPS) task. This approach combines a matching task with analyses similar to reverse correlation (Abbey and Eckstein, 2002; Li et al., 2004), thus minimizing decision confounds when measuring subjective appearance (Morgan et al., 2012, 2013; Jogan and Stocker, 2014). It estimates perceptual biases while several stimuli are presented simultaneously. We consider this a more naturalistic task, akin to our daily perceptual judgments (Figure 1A), compared with traditional psychophysical tasks involving single, isolated objects.

Results and Discussion

Idiosyncratic biases in size perception

Thirteen normal, healthy participants viewed an array of 5 circles on each trial and made a perceptual judgment (Figure 1B). The central circle was constant in size and served as the reference. Participants reported which of the four target circles appeared most similar in size to the reference. We fit a model to explain each participant's behavioral responses. For each of the four target locations we modeled the output of a detector tuned to stimulus size. In each trial the detector showing the strongest response predicted the participant's behavioral choice. This procedure thus estimated the raw perceptual bias and uncertainty at each location (Figure 1 – figure supplement 1).

Peripheral stimuli appeared smaller on average than the central reference, confirming earlier reports (Helmholtz, 1867; Bedell and Johnson, 1984; Anstis, 1998). This reduction in apparent size increased with stimulus eccentricity (Figure 1C). When instead of isolated circles we presented the target circles inside larger concentric circles, perceptual biases were predictably shifted in the other direction (the Delboeuf illusion; Delboeuf, 1892) so that targets appeared on average larger than the reference (Figure 1C). This, however, interacted with the effect of eccentricity on apparent size, with a gradual reduction in (illusory) size as the stimuli moved into the periphery. In the classical Delboeuf

illusion, perceptual biases are compared to a reference, either at the same eccentricity or even at the same *stimulus location*. In contrast, in our task the reference is at fixation. To disentangle the illusion from the effect of eccentricity, we therefore also calculated the *classical Delboeuf illusion strength*, that is, the difference in perceptual bias for isolated circles and the illusion stimuli at each location. This effect (here an *increase* in apparent size) also increased with eccentricity (Figure 1C).

Importantly, we further analyzed the *idiosyncratic patterns* of perceptual biases within each participant, both for isolated circles and for Delboeuf stimuli. Biases were very strongly correlated for both conditions and across the three eccentricities tested (Figure 1D and Figure 1 – figure supplements 2-3). That is, if observers perceived a large reduction in the apparent size of an isolated stimulus at a given location, they tended to show large reductions both for isolated circle and Delboeuf stimuli within the same visual field quadrant, regardless of eccentricity. We further confirmed that bias estimates were highly reliable by comparing bias estimates from two sessions conducted on different days (Figure 1 – figure supplement 4).

Long-term reliability of bias estimates

Nine participants who had taken part in the size eccentricity bias experiment also participated in the artificial size bias experiment approximately one year later. Despite the long time between these experiments and the fact that stimulus values were sampled differently (see Materials and Methods) the estimates of perceptual biases at target eccentricity 3.92° (which was common to both experiments) were well correlated between them ($r=0.37$, $p=0.028$). This correlation was largely driven by the within-subject variance. It was considerably greater after subtracting the mean across the four targets for every condition ($r=0.58$, $p=0.0002$). In contrast, removing the within-subject variance by averaging bias estimates across the four targets reduced the correlation substantially ($r=0.22$, $p=0.5759$).

Perceptual biases correlate with spatial tuning in V1

Next we employed functional magnetic resonance imaging (fMRI) with population receptive field (pRF) mapping to estimate the tuning of V1 voxels to spatial position (Figure 2A-B; Dumoulin and Wandell, 2008; Schwarzkopf et al., 2014). Interestingly, this revealed a systematic relationship between pRF size and perceptual biases: both isolated circles (pooled across eccentricity: $r=0.41$, $p<0.0001$; Figure 2C-E) and Delboeuf stimuli ($r=0.34$, $p=0.0001$; Figure 2F-I) were perceived as smaller when they were presented at visual field locations covered by voxels with larger pRFs. This pattern was also present across the visual field when data were averaged across participants: increasing eccentricity gives both an increase in pRF size (Dumoulin and Wandell, 2008; Schwarzkopf

et al., 2014) and a decrease in apparent size. Our data now demonstrate that this is also true for *idiosyncratic variations* in both pRF and apparent size at a *fixed* eccentricity.

The fact that apparent size of our circle stimuli is smaller when pRFs are larger could directly follow from the hypothesis that object size is inferred from the spatial position of neuronal activity inside the retinotopic map. Larger pRFs give a coarser retinotopic representation of the stimulus, with a greater spread of activation in the representation of the object's edges. It is possible that a decoding mechanism (e.g. based in higher brain regions involved in perceptual decisions) is required to 'read out' the position of the circle edges based on this activation (Pouget et al., 2000). Provided that the stimulus is so small that cortical receptive fields undersample it (as should be the case in our experiments), as pRF size increases the activity from the opposing edges of the circle would increasingly overlap thus shifting the resulting position estimates closer together. This in turn results in a reduction in apparent size. Alternatively, the less precise stimulus representation with larger pRFs may simply interact with the concordant reduction in local cortical magnification. Higher-level decoding mechanisms would thus consistently underestimate the size of the stimulus because they are inadequately calibrated to the idiosyncratic differences in cortical magnification (Anstis, 1998).

Dissociation between basic perceptual bias and contextual illusions

In an indirect replication of our earlier findings (Schwarzkopf et al., 2011; Schwarzkopf and Rees, 2013; Song et al., 2013b), we also observed an inverse relationship between classical Delboeuf illusion strength and V1 area ($r=-0.31$, $p=0.0006$; Figure 3A-C). Raw perceptual biases for isolated circles also correlated with V1 area ($r=0.27$, $p=0.0032$; Figure 3D-F). In contrast, raw biases for Delboeuf stimuli (relative to the central reference) were not correlated with V1 area ($r=-0.09$, $p=0.3063$; Figure 3G-I). Because the classical Delboeuf illusion is the *difference* in apparent size between these stimuli at the *same location*, it instead reflects the contextual interaction between target and surround.

The bias induced by the illusion thus likely differs mechanistically from basic perceptual biases. As our main findings show, both isolated circles (Figure 2C-E) and Delboeuf stimuli (Figure 2F-I) were perceived as smaller when pRFs were large – however, *at the same location* Delboeuf stimuli were nonetheless seen as larger than isolated circles. However, even though the apparent size of both isolated circles and Delboeuf stimuli was linked to pRF size, the *difference* between them was correlated with V1 surface area only.

The illusion effect may be modulated by cortical distance, possibly via lateral intra-cortical connections (Song et al., 2013a; Schwarzkopf, 2015), rather than pRF size. We conjectured

previously that the illusion could arise due to local connectivity between V1 neurons encoding the target circle and the surrounding context. Thus the illusion may be weaker when V1 surface area (and thus cortical distance) is larger (Schwarzkopf et al., 2011; Schwarzkopf and Rees, 2013; Song et al., 2013a, 2013b). In contrast, basic perceptual biases for any stimulus seem to be linked to the coarseness (pRF size) of the retinotopic stimulus representation itself.

Taken together, these results indicate that different mechanisms influence apparent size: both isolated circles and Delboeuf stimuli generally appear smaller (relative to the central reference) when pRFs are large, but an increase in cortical surface area (and thus the scale required of intra-cortical connections) is associated with a decrease in the illusory modulation of perceived size. Because our task estimates perceptual biases under either condition relative to a *constant* reference, it was uniquely suited to reveal dissociations between these effects. A more traditional task in which reference stimuli are presented at matched locations/eccentricities would be insensitive to this difference.

MAPS reliably detects perceptual biases

Our results thus demonstrate a neural correlate of apparent size even for simple, non-contextual stimuli. However, because it is difficult to disentangle *perceptual* biases from *response* or *decision* biases (Morgan et al., 2012), we next ran a series of control experiments to validate the perceptual bias estimates from MAPS. Participants performed four runs of the experiments with circles presented at the intermediate eccentricity of 3.92°. In two of these runs we introduced artificial biases to test how well MAPS performed when the ground truth was known. In these runs, two of the four target stimuli were altered subtly without the participants' knowledge. In one experiment one target was physically larger while the other was smaller. In another experiment participants instead judged the orientation of small grating patches. Here one of the gratings was tilted 5° clockwise while another was tilted 5° counterclockwise relative to the reference. In both experiments MAPS reliably identified the locations with artificial biases although the magnitude of the estimated bias was smaller than the physically introduced bias (Figure 4).

Results from MAPS are consistent with traditional psychophysics

Next we compared perceptual biases estimated with MAPS with those measured with the traditional method of constant stimuli (MCS). Participants completed two runs of MAPS followed by an experiment with a MCS design: in different blocks of the experiment participants were instructed to attend to only one of the four target circles. They then judged whether this target was larger or smaller than the central reference circle (Figure 5A). We estimated the perceptual bias at each

location by fitting a cumulative Gaussian curve to the behavioral responses. The bias estimates from this experiment correlated strongly with those from MAPS (Figure 5B). However, under MAPS the magnitude of biases was reduced significantly ($t(7)=7.5$, $p<0.0002$), consistent with conservative biases observed in the artificial bias experiments (Figure 4).

Spatial attention modulates perceptual biases

The more conservative biases observed with MAPS may be related to prior observations that judgements of multiple simultaneous items shifts biases towards the set average (Ariely, 2001; Parkes et al., 2001; Chong and Treisman, 2003; Walker and Vul, 2014): MAPS requires participants to attend to four locations simultaneously; in the MCS attention is always focused on single targets. This could affect perceptual biases. Results from another control experiment in which we presented an attentional cue at a given location briefly before stimulus onset are consistent with this interpretation: perceptual biases were enhanced subtly for the cued location (Figure 6A) while the frequency with which participants chose the cued location was unaltered (Figure 6B).

Biases measured with MAPS are less conspicuous

To further investigate this possibility, we conducted a two-alternative forced choice (2AFC) experiment to test whether biases estimated by MAPS or MCS were a closer match to the participants' subjective experience. In each trial participants were presented sequentially with two stimulus arrays separated by a blank interval (Figure 7A). One stimulus contained five equally sized circles (the unbiased stimulus). The other contained circles whose sizes had been altered according to the perceptual biases measured either with MAPS or MCS, multiplied by a factor ('bias factor'). Participants judged which interval contained the unbiased stimulus. Unsurprisingly, when biases were doubled, tripled or inverted, participants could reliably detect that difference (Figure 7B). Conversely, when both stimuli were unbiased (bias factor 0) participants performed at chance. However, when participants saw a stimulus constructed from the actual bias estimates (bias factor 1) they were able to distinguish the biased and unbiased stimuli derived from MCS better than those derived from MAPS ($t(7)=-3.43$, $p=0.011$).

Importantly, this was not trivially explained by greater task difficulty due to the smaller bias estimates from MAPS: Theoretically, the differences between MAPS and MCS in the 2AFC experiment could have arisen because the task was more difficult since biases estimated by MAPS were smaller than those estimated by MCS. However, greater difficulty should result in performance symmetric around a bias factor of zero. This was not the case as performance at negative bias factors was consistently above chance. Most notably, participants were able to distinguish inverted

MAPS biases (bias factor -1) significantly above chance levels ($t(7)=2.55$, $p=0.038$) while performance for the actual MAPS biases (bias factor 1) was at chance ($t(7)=0.88$, $p=0.41$), and the difference between these two points was also significant ($t(7)=2.46$, $p=0.043$). Thus MAPS provided a somewhat closer estimate of participants' subjective experience of the circles in the array and suggests that the array configuration indeed reduced perceptual biases as predicted by our attentional cueing experiment and previous findings (Ariely, 2001; Parkes et al., 2001; Chong and Treisman, 2003; Walker and Vul, 2014).

Heterogeneity in perceptual biases has central origin

Naturally, the spatial heterogeneity in perceptual biases could possibly arise from factors prior to visual cortical processing, like small corneal aberrations or inhomogeneity in retinal organization. We tested this possibility in a final control experiment in which we measured perceptual biases while we presented the stimuli either binocularly or dichoptically to the left and right eye. There was a close correspondence between biases measured with either eye (Figure 8). Thus at least a large part of the variance in perceptual biases must arise at a more central stage of visual processing where the input from both eyes has converged, such as the binocular cells in V1.

Conclusions

Our experiments show that when the spatial tuning of neuronal populations in V1 is coarser, visual objects are experienced as *smaller*. These findings support the hypothesis that object size is inferred directly from retinotopic representations in V1 (Schwarzkopf, 2015) and thus provides an explanation for previous reports of a neural signature of apparent size in V1 responses (Murray et al., 2006; Fang et al., 2008; Sperandio et al., 2012; Pooresmaeili et al., 2013). Taking advantage of our unique task design, we further demonstrate that processes related to basic perceptual biases are dissociable from contextual effects, like the Delboeuf illusion: While raw perceptual biases of object size are explained by pRF size, V1 surface area (as a proxy for cortical distance) explains the contextual modulation of apparent size in these illusions. Lastly, our findings imply that measurements of functional brain architecture in sensory areas can predict individual differences in our subjective experience of the world.

Materials and Methods

Participants

The authors and several naïve observers participated in these experiments. All participants were healthy and had normal or corrected-to-normal visual acuity. All participants gave written informed consent and procedures were approved by the UCL Research Ethics Committee.

Ten participants (4 authors; 3 female; 2 left-handed; ages 24-37) took part in both the first behavioral experiment measuring perceptual biases at 3 eccentricities and in the fMRI retinotopic mapping experiment (henceforth, *size eccentricity bias* experiment). An additional 3 participants (1 female; all right-handed; ages 20-25) took part in behavioral experiments only but could not be recruited for the fMRI sessions. These fMRI data form also part of a different study investigating the inter-session reliability of pRF analysis that we are preparing for a separate publication.

Fifteen participants (4 authors; 7 female; 1 left-handed; ages 20-37) took part in an additional behavioral experiment measuring perceptual biases in size perception with or without artificially induced biases (*artificial size bias*). 10 participants (1 author; 4 female; 2 left-handed; ages 21-32) took part in the artificial bias experiment using orientation discrimination (*artificial orientation bias*). Eight participants (4 authors; 2 female; 1 left-handed) participated in the experiment comparing perceptual biases measured with different tasks (*comparing tasks*). Eighteen participants (6 authors; 8 female; 1 left-handed; ages 21-42) participated in the attentional cueing experiment (*attentional cueing*). Finally, 6 participants (5 authors; 2 female; all right-handed; ages 21-36) participated in the dichoptic control experiment (*dichoptic bias*).

General psychophysical procedure

Participants were seated in a dark, noise-shielded room in front of a computer screen (Samsung 2233RZ) using its native resolution of 1680 x 1050 pixels and a refresh rate of 120Hz. Minimum and maximum luminance values were 0.25 and 230cd/m^2. Head position was held at 48cm from the screen with a chinrest. Participants used both hands to indicate responses by pressing buttons on a keyboard.

The dichoptic control experiment took place in a different testing room, using an Asus VG278 27" LCD monitor running its native resolution of 1920 x 1080 pixels and a refresh rate of 120Hz. Minimum and maximum luminance values were 0.16 and 100cd/m^2, with a viewing distance of 60 cm ensured with a chinrest. To produce dichoptic stimulation participants wore nVidia 3D Vision 2 shutter goggles synchronized with the refresh rate of the monitor. Frames for left and right eye stimulation thus alternated at 120Hz.

249 Multiple Alternatives Perceptual Search (MAPS) procedure

250 To estimate perceptual biases efficiently at four visual field locations we developed the MAPS
251 procedure. This is a matching paradigm using analyses related to reverse correlation or classification
252 image approaches (Abbey and Eckstein, 2002; Li et al., 2004) that seeks to directly estimate the
253 points of subjective equality, whilst also allowing an inference of discrimination sensitivity.

254 *Stimuli*

255 All stimuli were generated and displayed using MATLAB (The MathWorks Inc., Natick, MA) and the
256 Psychophysics Toolbox version 3 (Brainard, 1997). The stimuli in all the size discrimination
257 experiments comprised light grey (54cd/m^2) circle outlines presented on a black background. Each
258 stimulus array consisted of five circles (Figure 1B). One, the *reference*, was presented in the center of
259 the screen and was always constant in size (diameter: 0.98° visual angle). The remaining four, the
260 *targets*, varied in size from trial to trial and independently from each other. They were presented at
261 the four diagonal polar angles and at a distance of 3.92° visual angle from the reference, except for
262 the size eccentricity bias experiment where target eccentricity could be 1.96°, 3.92°, or 7.84° visual
263 angle. To measure the bias under the Delboeuf illusion, a larger inducer circle (diameter: 2.35°)
264 surrounded each of the four target circles (but not the reference) to produce a contextual
265 modulation of apparent size.

266 In the artificial orientation bias experiment, the stimuli were five small Gabor patches (sinusoidal
267 grating with wavelength: 0.62°; contrast: 50% convolved with a Gaussian envelope with standard
268 deviation: 0.49°) presented on a grey background (54cd/m^2). The central patch was the reference
269 and was oriented either at 30° or 120° counterclockwise from horizontal. The reference orientation
270 was counterbalanced across participants. From trial to trial the target orientations varied relative to
271 the reference.

272 In all experiments on size perception, the independent variable (the stimulus dimension used to
273 manipulate each of the targets) was the binary logarithm of the ratio of diameters for the target
274 relative to the reference circles. In the size eccentricity bias experiments the sizes of the four targets
275 were drawn without replacement from a set of fixed sizes (0, ±.05, ±.1, ±.15, ±.2, ±.25, ±.5, ±.75, or
276 ±1 log units). Thus, frequently there was no "correct" target to choose from. Because this made the
277 task feel quite difficult for many participants, in all the other experiments we decided to select a
278 random subset of three targets from a Gaussian noise distribution centered on 0 (the
279 size/orientation of the reference) where one target was correct, i.e. it was set to 0. The standard
280 deviation of the Gaussian noise was 0.3 log units for size discrimination experiments and 15° for the

artificial orientation bias experiment. A subset of 9 participants were tested both in the size eccentricity bias experiment and the artificial size bias experiment (at 3.92° eccentricity) confirming that these minor methodological differences did not affect results substantially despite the fact that these experiments were separated by approximately one year.

Tasks

Each trial started with 500ms during which only a fixation dot (diameter: 0.2°) was visible in the middle of the screen. This was followed by presentation of the stimulus array for 200ms after which the screen returned to the fixation-only screen. Participants were instructed to make their response by pressing the F, V, K, or M button on the keyboard corresponding to which of the four targets appeared most similar to the reference. After their response a "ripple" effect over the target they had chosen provided feedback about their response. In the size discrimination experiments this constituted three 50ms frames in which a circle increased in diameter from 0.49° in steps of 0.33° and in luminance. In the orientation discrimination experiments this was a non-Cartesian grating, a concentric sinusoidal grating (wavelength: 2.48°) convolved with a Gaussian envelope increasing from 0.49° in steps of 0.33° and decreasing in contrast from 50% in steps of 10%.

Moreover, the color of the fixation dot also changed during these 150ms to provide feedback about whether the behavioral response was correct. In the size eccentricity bias experiment, the color was green and slightly larger (0.33°) for correct trials and red for incorrect trials. In all later experiments, we only provided feedback on *correct* trials. This helped to reduce the anxiety associated with large numbers of incorrect trials that are common in this task: Accuracy was typically around 45-50% correct. Even though this is well in excess of chance performance of 25% it means that participants would frequently make mistakes.

Experimental runs were broken up into blocks of 20 trials. After each block there was a resting break. A message on the screen reminded participants of the task and indicated how many blocks they had already completed. Participants initiated blocks with a button press.

Size eccentricity bias experiment: Participants were recruited for two sessions on separate days. In each session they performed six experimental runs, three with only circles and three with the Delboeuf stimuli. Each run tested one of the three target eccentricities. Trials with different eccentricities were run in separate blocks to avoid confounding these measurements with differences in attentional deployment across different eccentricities. There were 10 blocks per experimental run.

Artificial size/orientation bias experiments: Half of the experimental runs participants performed measured their baseline biases. The other half of the runs contained artificial biases: two of the four targets were altered subtly, one by adding and one by subtracting 0.1 log units for the artificial size or 5° for the artificial orientation bias. Which two targets were altered was counterbalanced across participants, as was the order of experimental runs. In the artificial orientation bias experiment participants were recruited for two sessions on separate days comprising four runs each (half with artificial biases). In the artificial size bias experiment participants were recruited for only one session comprising four runs (two with artificial bias) plus an additional run measuring the Delboeuf stimuli. There were 10 blocks per experimental run.

Attentional cueing experiment: In this experiment presentation of the stimulus array was preceded by a brief attentional cue. This comprised two small dots (diameter: 0.49°) presented at 120% and 83% of the eccentricity along the radial axis between fixation and the cued target location. The duration of the cue was 60ms followed by 40ms of fixation. There were five experimental conditions: one for cueing each of the four targets and a baseline condition during which there was no attentional cue. These conditions were randomly interleaved over the course of the experiment. Participants were recruited for two sessions on separate days. In each session they performed one experimental run comprising 50 blocks.

Dichoptic bias experiment: There were three experimental conditions in this experiment. By means of shutter goggles the stimulus arrays could be presented dichoptically, either binocularly or monocularly to either eye. To aid stereoscopic fusion we additionally added 5 concentric squares surrounding the stimulus arrays (side length: 18.9-98.9° in equal steps). The three experimental conditions were randomly interleaved within each experimental run. There were 34 blocks per run; however, in this experiment each block comprised only 12 trials. Participants performed two such runs within a single session.

Analysis

To estimate perceptual biases we fit a model to predict a given participant's behavioral response in each trial (Figure 1 – figure supplement 1). For each target stimulus location a Gaussian tuning curve denoted the output of a "neural detector". The detector producing the strongest output determined the predicted choice. The model fitted the peak location (μ) and dispersion (σ) parameters of the Gaussian tuning curves that minimized the prediction error across all trials. Model fitting employed the Nelder-Mead simplex search optimization procedure (Lagarias et al., 1998). We initialized the μ parameter as the mean stimulus value (size or orientation offset) whenever a given target location was chosen *incorrectly*. We initialized the σ parameter as the standard deviation across all stimulus

values when a given target location was chosen. The final model fitting procedure however always used *all* trials, correct and incorrect.

To estimate the perceptual bias at each of the four stimulus locations we divided the raw peak location (μ) by the dispersion (σ). Thus the perceptual bias is a standardized score in units of the just noticeable difference. This transformation is important because otherwise raw biases are confounded by discrimination ability and the analysis puts undue weight on biases for locations where discrimination is poor. We also quantified the choice probabilities for each location as a measure of response bias.

The MAPS analysis could predict participants' behavioral responses well above chance levels (25%). Model prediction accuracy in the size eccentricity experiment ranged between 43.2±2.3% and 53.4±1.7%. Importantly, the model fitting procedure afforded on average a subtle (1-1.8%) but highly significant improvement (least significant difference: $t(12)=2.7$, $p=0.018$) of prediction accuracy over a basic model using only the seed parameters for the tuning functions.

<u>Comparing tasks experiment</u>

We compared perceptual biases measured with MAPS to those measured with the more traditional method of constant stimuli. Participants first completed a short version of the size bias measurement using MAPS, comprising two experimental runs of MAPS with 10 blocks each and at a target eccentricity of 3.92°.

Method of constant stimuli (MCS)

The two runs of MAPS were followed by an MCS run. The stimuli and trial sequence were similar to the MAPS experiments. However, the task was to compare the size of *one* of the four targets with the reference and indicate which one was *larger* by means of a button press. In separate blocks participants were instructed which of the targets they had to judge. The size of the current target was chosen to be either 0, ±.05, ±.1, ±.2, ±.3, ±.5, or ±1 log units and there were 80 trials for each these 13 possible sizes. The sizes of the three remaining circles (distracters) were chosen from a Gaussian distribution with a standard deviation of 0.3 log units. There were 80 blocks in the run and the whole range of possible target sizes appeared in each block in a pseudo-randomized order.

We then estimated the perceptual bias by calculating the proportion of trials for each target stimulus size that the target was reportedly seen as larger than the reference. We fit a cumulative Gaussian function with three free parameters to these data: 1. The peak of the Gaussian, μ, to estimate the point of subjective equality. 2. The standard deviation of the Gaussian, σ, to quantify the discrimination uncertainty. 3. The amplitude of the Gaussian, β, to take lapse rates into account.

This curve fitting was done using the same optimization procedure as the MAPS model fitting procedure (Lagarias et al., 1998) to minimize the squared residuals of the fitted model. As with MAPS, the raw perceptual biases were normalized by dividing the raw point of subjective equality by the uncertainty.

Two-alternative forced choice (2AFC) experiment

Finally, we validated the perceptual biases estimated by MAPS and MCS in a 2AFC experiment. In each trial of this experiment participants saw two consecutive stimulus arrays. In one array all targets were the same size as the reference (*unbiased*), while target sizes in the other array deviated from the reference (*biased*). Participants were asked to indicate which of the two arrays was unbiased.

For each participant, we constructed individual biased stimulus arrays using the raw peak locations of the psychometric curves from the two previous experiments. In the MAPS and MCS conditions biases were based on the estimates derived with the respective methods and trials of these conditions were randomly interleaved. We further manipulated the 'bias factor,' that is, in different trials the biases of the target stimuli were multiplied by either 0, ±1, ±2, or ±3. Because a bias factor of 1 represents the most critical condition, the actual measured biases from MAPS and MCS, there were 120 trials for this conditions, twice as many as for the other conditions (although note that bias factor 0 is physically identical for MAPS and MCS so there were also 120 trials of that condition overall).

Each trial began with 500ms during which only a blue fixation dot was presented against a dark screen. This was followed by the presentation of the two (biased and unbiased) stimulus arrays for 200ms, interleaved with another 500ms fixation interval. Which interval contained the unbiased array was counterbalanced across trials.

There were 32 trials per block, one for each combination of condition (MAPS vs MCS), bias factor, and which interval contained the unbiased array. There were 30 blocks in the whole experiment.

<u>Retinotopic mapping experiment</u>

Ten individuals participated in two sessions of retinotopic mapping in a Siemens Avanto 1.5T MRI scanner using a 32-channel head coil. The front half of the coil was removed to allow unrestricted field of view leaving 20 channels. Participants lay supine and watched the mapping stimuli, which were projected onto a screen (resolution: 1920 x 1080) at the back of the bore, via a mirror mounted on the head coil. The viewing distance was 68cm.

408 We used a T2*-weighted multiband 2D echo-planar sequence (Breuer et al., 2005) to acquire 235
409 functional volumes per run (2.3mm isotropic voxels, 36 slices, FOV=96x96 voxels, TR=1000ms,
410 TE=55ms, flip angle=75°, acceleration=4). In addition, we collected a T1-weighted anatomical
411 magnetization-prepared rapid acquisition with gradient echo (MPRAGE) scan with 1 mm isotropic
412 voxels (TR=2730ms, TE=3.57ms) using the full 32-channel head coil.

413 *Stimuli and task*

414 The procedure for retinotopic mapping was similar to a previous study (Alvarez et al., 2015). In short,
415 we used a combined wedge and ring stimulus. A polar wedge subtending a polar angle of 12° rotated
416 in 60 discrete steps (one per second) around the fixation dot (diameter: 0.13° surrounded by a 0.25°
417 annulus where contrast ramped up linearly). A ring expanded or contracted, both in width and
418 overall diameter, in 36 logarithmic steps. The maximal eccentricity of the wedge and ring was 8.5°.
419 There were 3 cycles of wedge rotation and 5 cycles of ring expansion/contraction. Each mapping run
420 concluded with 45s of a fixation-only period. At all times a low contrast 'spider web' pattern was
421 superimposed on the screen to aid fixation compliance.

422 The wedge and ring parts contained colorful natural images from Google Image search, which
423 changed every 500ms. They depicted outdoor scenes (tropical beaches, forests, mountains, and rural
424 landscapes), faces, various animals, and pictures of written script (228 images in total). One picture
425 depicted the 'Modern Anderson' clan tartan. These pictures were always rotated in accordance with
426 the current orientation of the wedge. Participants were asked to fixate a fixation dot at all times.
427 With a probability of 0.03 every 200ms the black fixation dot would change color for 200ms to one
428 of the primary and complementary colors or white followed by another 200ms of black. Participants
429 were asked to tap their finger when the dot turned red. To also maintain attention on the mapping
430 stimulus they were asked to tap their finger whenever they saw the tartan image.

431 In alternating runs the wedge rotated in clockwise and counterclockwise directions, while the ring
432 expanded and contracted, respectively. In each session we collected six such mapping runs and an
433 additional run to estimate the hemodynamic response function. The latter contained 10 trials each
434 of which started with a 2s sequence of four natural images from the same set used for mapping.
435 These were presented in a circular aperture centered on fixation with radius 8.5° visual angle. This
436 was followed by 28s of the blank screen (fixation and radar screen only).

437 *Population receptive field (pRF) analysis*

438 The method we used for analyzing pRF (Dumoulin and Wandell, 2008) data has been described
439 previously (Schwarzkopf et al., 2014; Alvarez et al., 2015). The MATLAB toolbox (available at

http://dx.doi.org/10.6084/m9.figshare.1344765) models the pRF of each voxel as a two-dimensional Gaussian in visual space and incorporates the hemodynamic response function measured for each individual participant. It determines the visual field location (in Cartesian coordinates) and the size (standard deviation) of the pRF plus an overall response amplitude. We delineated the early visual regions (specifically V1) manually and then extracted the pRF parameter data from each visual field quadrant. Data were divided into eccentricity bands and we calculated mean pRF size for each band. Finally, we fit a linear function to the relationship between pRF size and eccentricity to extrapolate each person's pRF size at the target locations in the behavioral experiments. We also quantified the macroscopic surface area of each visual quadrant in V1 and normalized it relative to the whole cortical surface area.

References

Abbey CK, Eckstein MP (2002) Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. J Vis 2:66–78.

Alvarez I, De Haas BA, Clark CA, Rees G, Schwarzkopf DS (2015) Comparing different stimulus configurations for population receptive field mapping in human fMRI. Front Hum Neurosci 9:96.

Anstis S (1998) Picturing peripheral acuity. Perception 27:817–825.

Ariely D (2001) Seeing Sets: Representation by Statistical Properties. Psychol Sci 12:157–162.

Bedell HE, Johnson CA (1984) The perceived size of targets in the peripheral and central visual fields. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom 4:123–131.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Breuer FA, Blaimer M, Heidemann RM, Mueller MF, Griswold MA, Jakob PM (2005) Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. Magn Reson Med 53:684–691.

Chong SC, Treisman A (2003) Representation of statistical properties. Vision Res 43:393–404.

Delboeuf J (1892) Sur une nouvelle illusion d'optique. Acade ¨ Mie R Sci Lett B-arts Belg Bull 24:545–558.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. NeuroImage 39:647–660.

Fang F, Boyaci H, Kersten D, Murray SO (2008) Attention-dependent representation of a size illusion in human V1. Curr Biol 18:1707–1712.

Helmholtz H (1867) Handbuch der Physiologischen Optik.

Jogan M, Stocker AA (2014) A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. J Vis 14:20.

Lagarias J, Reeds J, Wright M, Wright P (1998) Convergence properties of the Nelder—Mead simplex method in low dimensions. SIAM J Optim 9:112–147.

Li RW, Levi DM, Klein SA (2004) Perceptual learning improves efficiency by re-tuning the decision "template" for position discrimination. Nat Neurosci 7:178–183.

Morgan M, Dillenburger B, Raphael S, Solomon JA (2012) Observers can voluntarily shift their psychometric functions without losing sensitivity. Atten Percept Psychophys 74:185–193.

Morgan MJ, Melmoth D, Solomon JA (2013) Linking hypotheses underlying Class A and Class B methods. Vis Neurosci 30:197–206.

Murray SO, Boyaci H, Kersten D (2006) The representation of perceived angular size in human primary visual cortex. Nat Neurosci 9:429–434.

Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M (2001) Compulsory averaging of crowded orientation signals in human vision. Nat Neurosci 4:739–744.

Pooresmaeili A, Arrighi R, Biagi L, Morrone MC (2013) Blood Oxygen Level-Dependent Activation of the Primary Visual Cortex Predicts Size Adaptation Illusion. J Neurosci 33:15999–16008.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125–132.

Schwarzkopf DS (2015) Where Is Size in the Brain of the Beholder? Multisensory Res 28:285–296.

Schwarzkopf DS, Anderson EJ, Haas B de, White SJ, Rees G (2014) Larger Extrastriate Population Receptive Fields in Autism Spectrum Disorders. J Neurosci 34:2713–2724.

Schwarzkopf DS, Rees G (2013) Subjective size perception depends on central visual cortical magnification in human v1. PloS One 8:e60550.

Schwarzkopf DS, Song C, Rees G (2011) The surface area of human V1 predicts the subjective experience of object size. Nat Neurosci 14:28–30.

Song C, Schwarzkopf DS, Lutti A, Li B, Kanai R, Rees G (2013a) Effective connectivity within human primary visual cortex predicts interindividual diversity in illusory perception. J Neurosci Off J Soc Neurosci 33:18781–18791.

Song C, Schwarzkopf DS, Rees G (2013b) Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. Nat Commun 4:2201.

Sperandio I, Chouinard PA, Goodale MA (2012) Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. Nat Neurosci 15:540–542.

Walker D, Vul E (2014) Hierarchical encoding makes individuals in a group seem more attractive. Psychol Sci 25:230–235.

Figure Captions

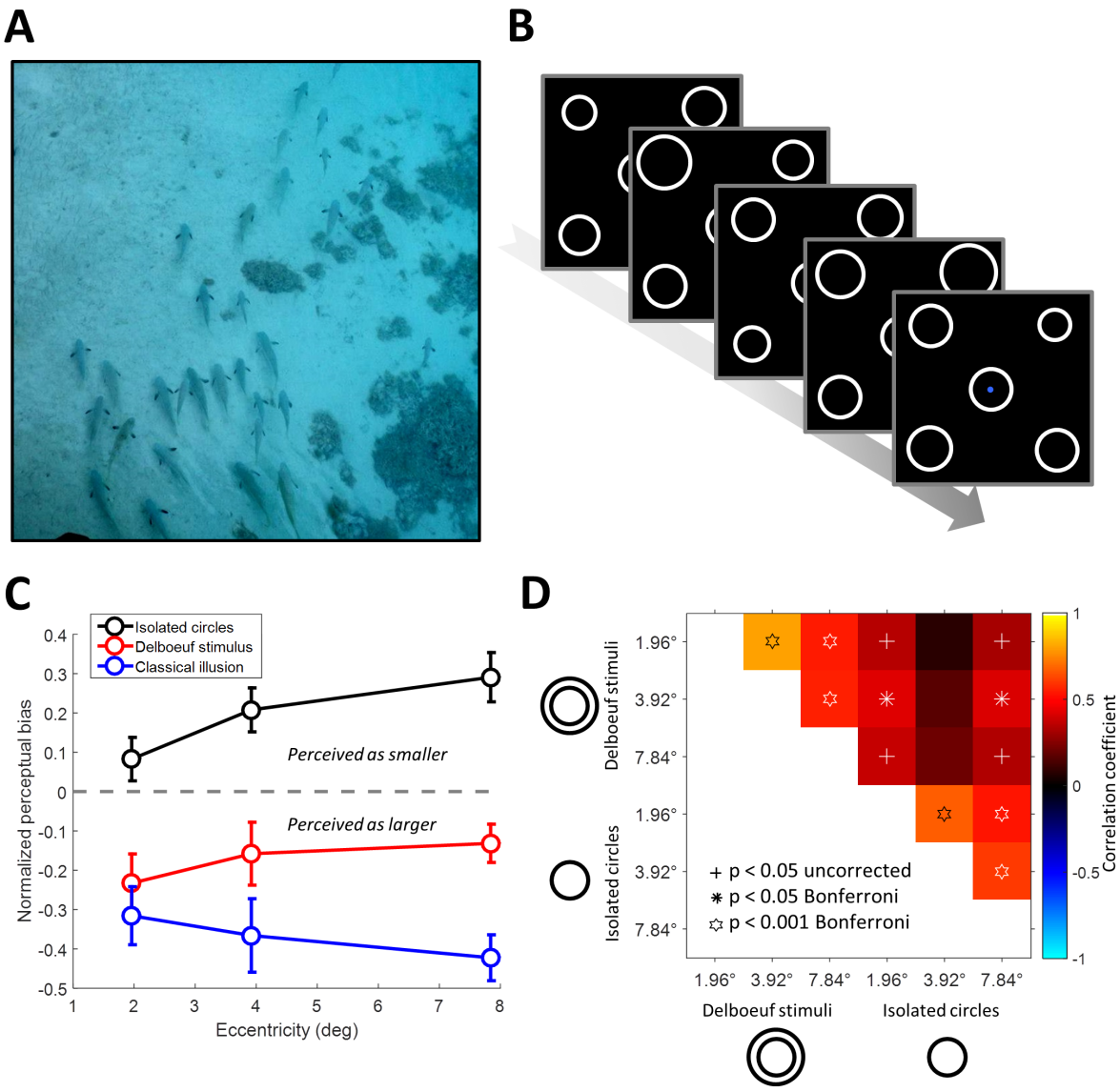

Figure 1. Idiosyncratic biases in size perception. **A.** Visual objects often appear in the presence of similar objects. For example, a spearfisherman may be searching this school for the largest fish. What is the neural basis for this judgment? **B.** The MAPS task. In each trial, participants fixated on the center of the screen and viewed an array of five circles for 200ms. The central circle was constant in size, while the others varied across trials. Participants judged which of the circles in the four corners appeared most similar in size to the central one. **C.** Average perceptual bias (positive and negative: target appears smaller or larger than reference, respectively), across individuals plotted against target eccentricity for simple isolated circles (black), contextual Delboeuf stimuli (red), and classical illusion strength (blue), that is, the difference in biases measured for the two stimulus conditions. Error bars denote ±1 standard error of the mean. **D.** Correlation matrix showing the relationship of unique patterns of perceptual biases in the two conditions (isolated circles and Delboeuf stimuli) and at the three target eccentricities. Color code denotes the correlation

coefficient. Symbols denote statistical significance. Crosses: p<0.05 uncorrected. Asterisks: p<0.05 Bonferroni corrected. Hexagrams: p<0.001 Bonferroni corrected.

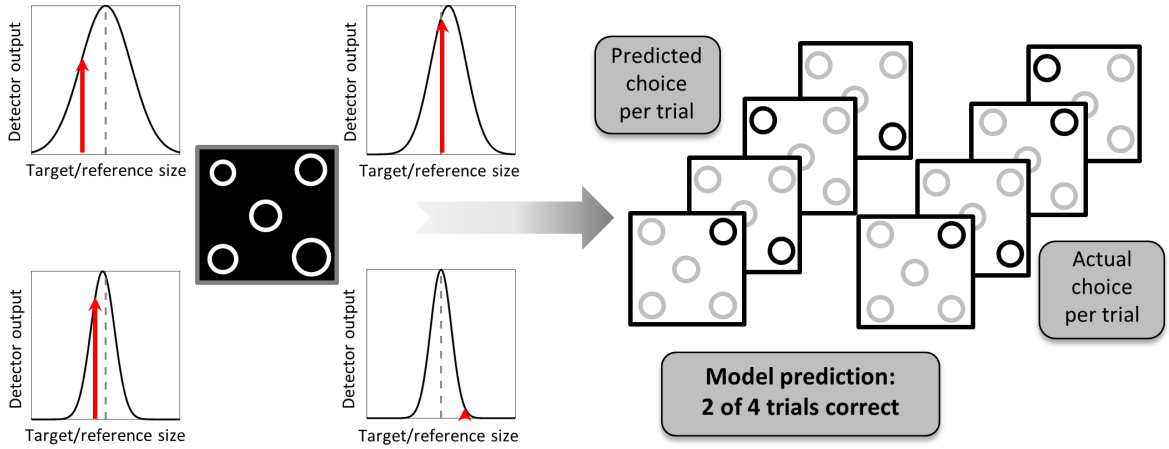

Figure 1 – figure supplement 1. Analysis of behavioral data from MAPS task. The behavioral responses in each trial were modeled by an array of four "neural detectors" tuned to stimulus size (expressed as the binary logarithm of the ratio between the target and the reference circle diameters). Tuning was modeled as a Gaussian curve. The detector showing the strongest output to the stimulus (indicated by the red arrows) determined the predicted behavioral response in each trial (here, the top-right detector would win). Model fitting minimized the prediction error (in this example the model predicted the actual behavioral choice correctly for 50% of trials) across the experimental run by adapting the mean and dispersion of each detector.

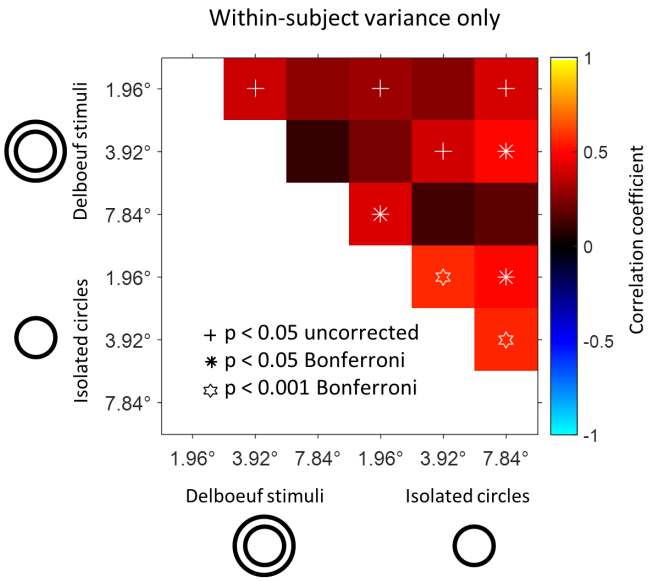

Figure 1 – figure supplement 2. Correlation matrices as in Figure 1D showing the relationship between the perceptual biases in the two conditions (isolated circles and Delboeuf stimuli) and at the three target eccentricities *after removing between-subject variance*, i.e. the mean across the biases for the four targets was *subtracted* from each condition.

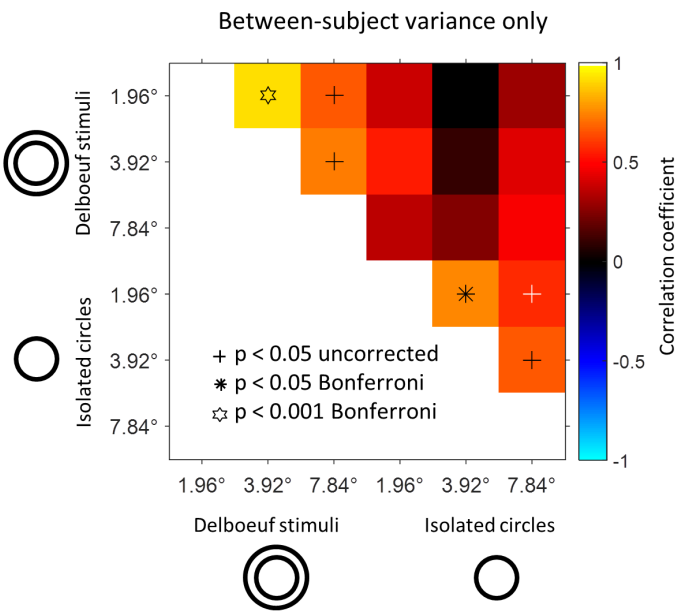

Figure 1 – figure supplement 3. Correlation matrices as in Figure 1D showing the relationship between the perceptual biases in the two conditions (isolated circles and Delboeuf stimuli) and at the three target eccentricities *after removing the within-subject variance*, i.e. biases were *averaged* across the four targets in each condition. Note that statistical power is lower relative to the other correlation matrices, because after averaging there is only a quarter of the number of observations.

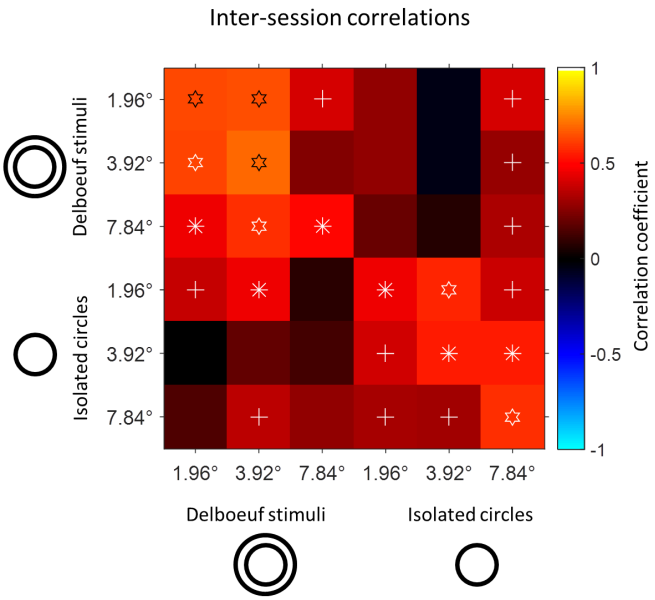

Figure 1 – figure supplement 4. Correlation matrices as in Figure 1D showing the relationship between the perceptual biases in the two conditions (isolated circles and Delboeuf stimuli) and at the three target eccentricities. Here data from the first and second session of the experiment conducted on different days were correlated.

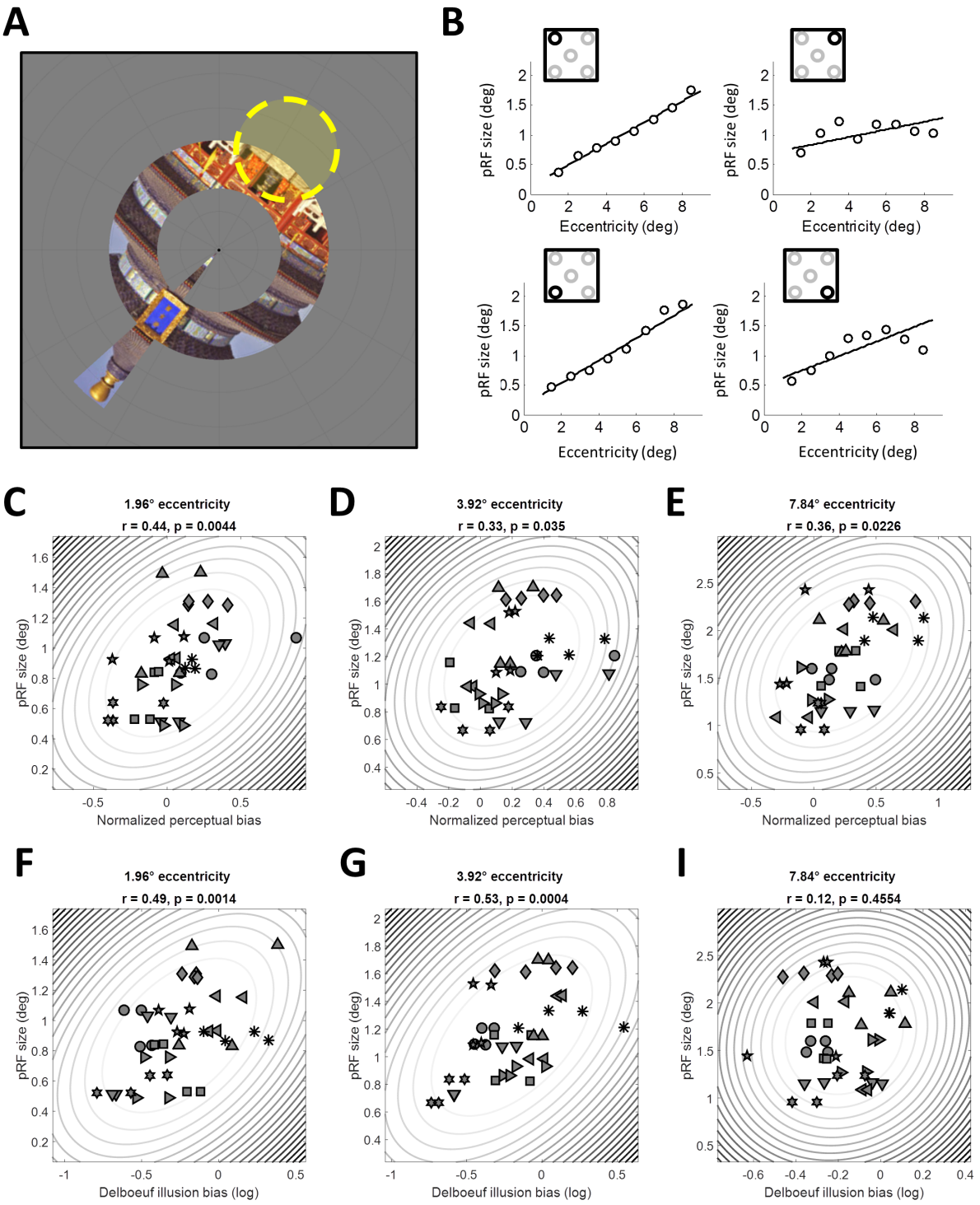

Figure 2. Neural correlates of size perception. **A.** Population receptive field (pRF) mapping with fMRI. Participants viewed natural images presented every 500ms within a combined wedge-and-ring aperture. In alternating runs the wedge rotated clockwise/counterclockwise in discrete steps (1Hz) around the fixation dot while the ring either expanded or contracted. A forward model estimated the position and size of the pRF (indicated by yellow circle) that best explained the fMRI response to the mapping stimulus. **B.** We estimated the pRF size corresponding to each target location in the behavioral experiment by fitting a linear function to pRF sizes averaged within each eccentricity band for each visual quadrant and extrapolating the pRF size at the target eccentricities. The four

plots show the pRF-eccentricity plots for the four target locations (see insets) in one participant. **C-E.** PRF size for each stimulus location and participant plotted against their perceptual bias. **F-I.** PRF size for each stimulus location and participant plotted against their perceptual bias for Delboeuf stimuli. In **C-I** Symbols denote individual participants. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The panels show results for stimuli at 1.96° (**C,F**), 3.92° (**D,G**), or 7.84° (**E,I**) eccentricity.

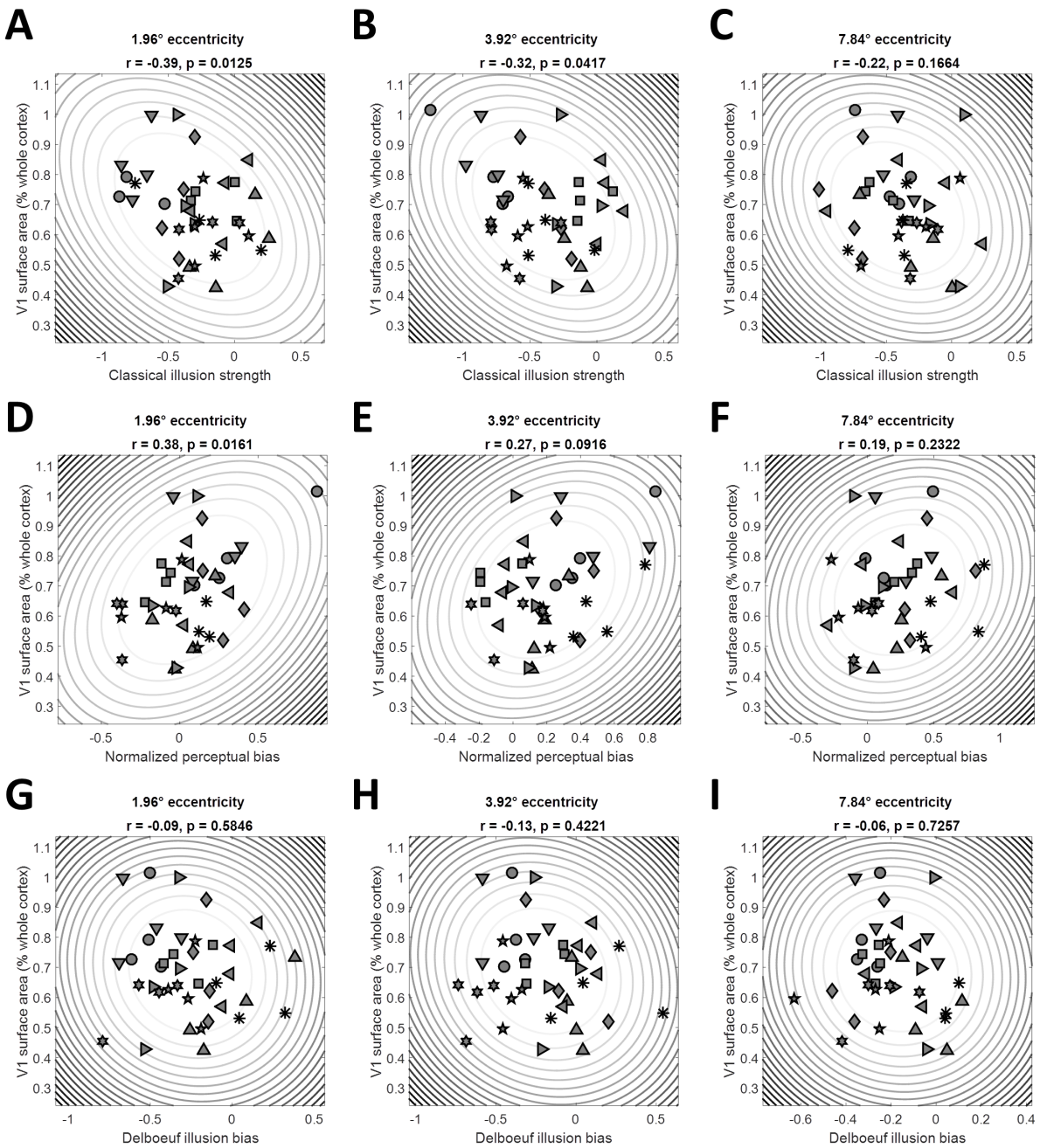

Figure 3. Macroscopic V1 surface area for each participant and each quarter field map in V1 plotted against the classical Delboeuf illusion strength (**A-C**), that is, the bias for Delboeuf stimuli minus the bias for isolated circles. V1 surface area plotted against raw perceptual biases for isolated circles **(D-F)** and for the Delboeuf stimuli **(G-I)**. Columns show data for stimuli at 1.96° (**A,D,G**), 3.92° (**B,E,H**), or 7.84° (**C,F,I**) eccentricity. Symbols denote individual participants. Elliptic contours denote the Mahalanobis distance from the bivariate mean.

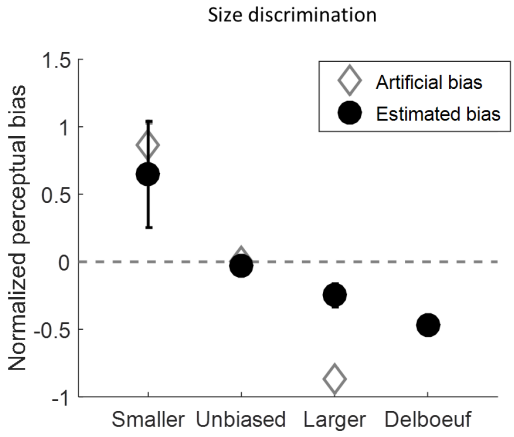

Figure 4. MAPS could reliably identify locations where an artificial bias had been introduced. The labels on the x-axis indicate the target locations with either positive (larger), negative (smaller) or zero bias. In the artificial size bias experiment, there was also one run to test the Delboeuf stimuli. Black circles: Estimated bias averaged across participants. Grey diamonds: Normalized artificial bias for each location averaged across participants (note that the sign has been inverted as the bias estimate compensates for the physically introduced bias). Error bars denote ±1 standard error of the mean. The MAPS estimates of perceptual bias went into the same direction as the artificial biases. However, the magnitude of the estimated bias tended to be smaller than the artificial bias.

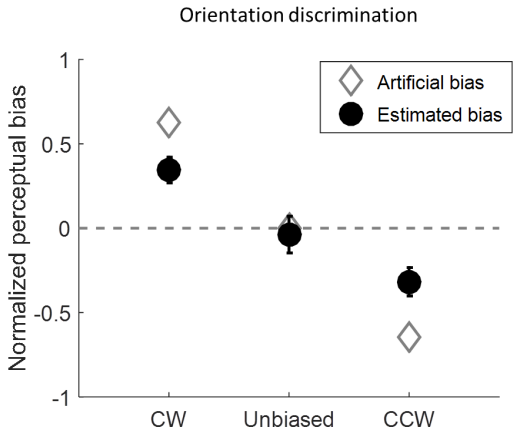

Figure 4 – figure supplement 1. MAPS could reliably identify locations where an artificial bias had been introduced. The labels on the x-axis indicate the target locations with either positive (counterclockwise), negative (clockwise) or zero bias. Black circles: Estimated bias averaged across participants. Grey diamonds: Normalized artificial bias for each location averaged across participants (note that the sign has been inverted as the bias estimate compensates for the physically introduced bias). Error bars denote ±1 standard error of the mean. CW: clockwise. CCW: counterclockwise.

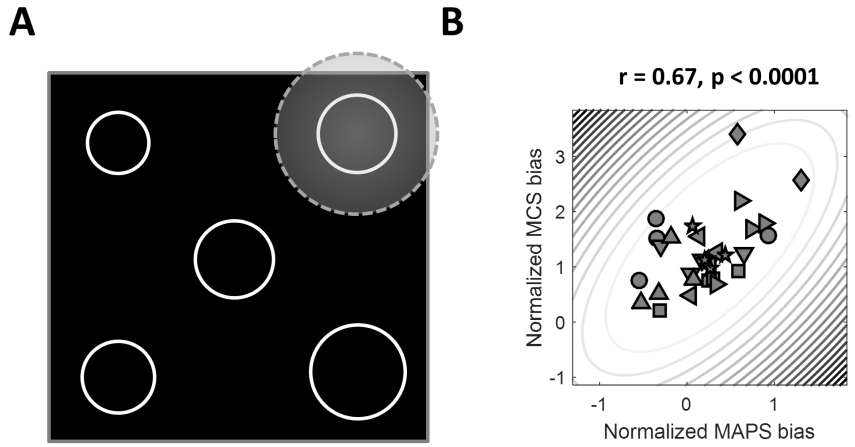

Figure 5. Comparing MAPS with the classical method of constant stimuli (MCS). **A.** In the MCS experiment participants were instructed before each block to attend to only one target (e.g. here the upper-right), while the remaining locations contained random distractors. Participants then judged whether the attended target appeared larger or smaller than the central reference and we extrapolated the point of subjective equality. **B.** Perceptual biases estimated with MAPS and MCS correlated strongly. Symbols denote different participants. Elliptic contours denote the Mahalanobis distance from the bivariate mean.

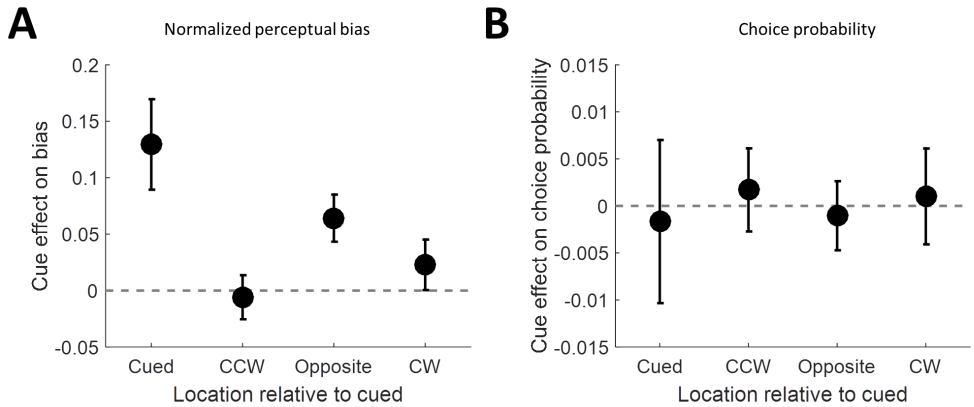

Figure 6. In the attentional cuing experiment, we flashed a brief cue at one of the target locations before stimulus onset. The effects of cuing on bias (**A**) and choice probability (**B**) are plotted for the four target locations relative to the cued location. Error bars denote ±1 standard error of the mean. CCW: counterclockwise. CW: clockwise. **A.** Perceptual bias at the cued location was enhanced subtly (repeated-measures ANOVA on cuing effect for the four locations: $F(3,51)=3.55$, $p=0.021$). **B.** There was no difference in the frequency of the participants' behavioral choices relative to the cued location ($F(3,51)=0.06$, $p=0.982$).

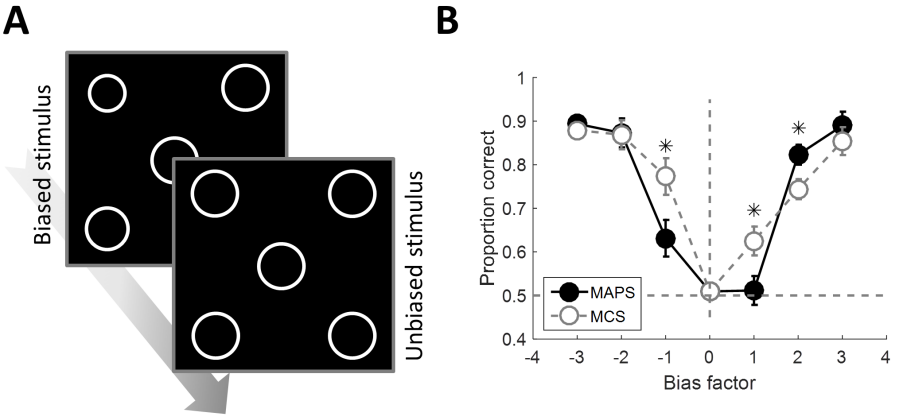

Figure 7. 2AFC experiment. **A.** In each trial, participants viewed two stimulus intervals. One interval contained an unbiased array in which the size of all five circles was identical. The other contained an array where the target sizes were determined by the biases measured either with MAPS or MCS, multiplied by a 'bias factor'. **B.** Performance in 2AFC experiment averaged across participants plotted against the bias factor for biases derived by MAPS (filled circles) or MCS (open circles). At the critical bias factor of 1 participants could not distinguish MAPS biases from the unbiased stimulus but could do so for MCS biases. Error bars denote ±1 standard error of the mean difference between MAPS and MCS at each bias factor. Asterisks denote significant differences at $p<0.05$ in a paired two-tailed t-test between MCS and MAPS at a given bias factor.

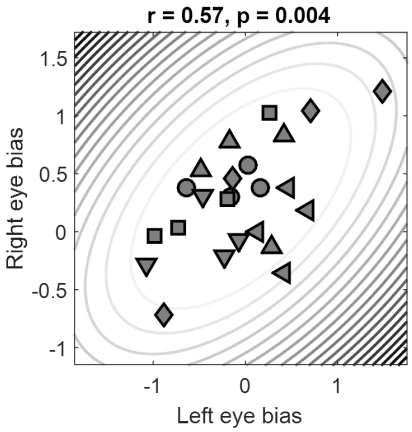

Figure 8. Normalized perceptual biases measured when stimuli were presented dichoptically to the left or right eye were strongly correlated. Symbols denote individual participants. Elliptic contours denote the Mahalanobis distance from the bivariate mean.