Public review process

The senior author of this study is a strong advocate of making the *content* of all peer reviews public (see http://wp.me/p5UOkc-bh for details). Such a process is not yet commonplace. As an example how this might work, I decided to make the reviews of our recently rejected manuscript public. It is not feasible to do this for every review under the current system – but considering the substantial changes to our manuscript, I feel that this is a good place to start. Although we are not able to submit our manuscript to the same journal, here we provide a response to each of the reviewers' concerns. This response helped us with planning our revisions. It may also prove useful to future reviewers of our study. Most importantly, it should enable readers to understand the evolution of the manuscript.

D. Sam Schwarzkopf

Response to Reviews

We thank all reviewers and the editor for the time and thought that went into their comments. Based on their comments we have restructured our manuscript fundamentally. We believe this greatly improved its clarity and strengthened our conclusions. Please note that at this particular journal, reviews are not the actual reviewers' comments but a summary written by an editor. Therefore, we cannot be sure that we directly addressed the points made by the reviewers or to what extent the comments reflect an informal third review by the editor. Nonetheless, it is a useful summary of points that should be improved.

In the following and in our revised manuscript we now use 'pRF spread' to describe the spatial extent of pRFs. This helps to distinguish our neuroimaging results from apparent 'size', that is, the stimulus dimension observers were asked to judge in our behavioral experiments. The review comments are in *grey italics* and our responses in normal font:

However, both reviewers have substantial reservations about the manuscript in its current form. Some of the writing is unclear and hard to follow, including key methods, […] Overall, there is much detail about the new MAPS task and little detail about the MRI measures. The main results, particularly figure 2, are potentially of broad interest, but the ms would entail major revamping in order to make this a convincing study.

We intended to present a thorough and detailed study that does not leave many fundamental questions unanswered. We originally felt it important to include a range of behavioral experiments (in particular the artificial bias experiments and the 2AFC experiment) simply because the MAPS task is a methodological novelty in our study and thus it seemed appropriate to provide evidence of its validity and robustness. **In contrast, the fMRI experiments in contrast used already established pRF mapping procedures** that we and others described in previous studies (Alvarez et al., 2015; Dumoulin and Wandell, 2008; de Haas et al., 2014; Schwarzkopf et al., 2014). Thus we felt they required a less detailed description.

Considering these reviews, we however now realize that a different balance may be appropriate. **We have therefore restructured our manuscript considerably by adding a more detailed description of the fMRI/pRF methods and removing the more tangential behavioral experiments**. That is, we removed the artificial bias, the attentional cueing, the MCS, and the 2AFC experiments. These experiments are interesting in their own right but they are probably not an essential part of this study. We will publish these findings separately after we have conducted additional experiments that seek to understand some of the more complex results (such as why bias estimates are smaller than in the MCS experiment and why behavioral performance in the 2AFC experiment at bias factor

1 is at, rather than below, chance levels). Either way, they have now already been presented publicly as part of the previous pre-print manuscript.

…the logic of some of the analyses (in particular estimates of bias), interpretation of the brain-behavior links, and even the main conclusion. There is also insufficient consideration of the relationship between pRF [spread] and cortical surface area, particularly in relation to other literature. In addition, choices about dependent measures are inadequately justified.

These points are discussed specifically below. Briefly, with regard to our main conclusion, we believe that the reviewers misunderstood what our main claim actually is. We have addressed this by improving the clarity of our manuscript, by discussing theoretical models of how size perception could relate to brain function, and by making the conclusions more nuanced.

*** CONCLUSION ** The claim that 'size is inferred directly from stimulus representations in V1' is vague. One reviewer indicates: Does this mean that size estimates do not depend on any visual processing beyond V1? That seems unlikely. In order for this to occur, presumably V1 would need to be able to find boundaries, do object recognition, and integrate information across the object, and so on.*

We did not mean to imply that higher areas are not involved in size judgments at all. In fact, we explicitly stated that this cannot be the case in the original manuscript (e.g. line 104-5). One of the authors has also discussed this in a recent review (Schwarzkopf, 2015). **Higher brain areas must be involved in the perceptual decision-making process** – even if size is inferred from signals in early visual cortex, some neurons must integrate these signals and guide behavior. We now elaborate on this issue in more detail our revised manuscript.

It also seems possible that many contextual modulations of apparent size, in particular those induced by three dimensional context (Fang et al., 2008; He et al., 2015; Murray et al., 2006; Ni et al., 2014), involve processing in higher regions (but also see Pooresmaeili et al., 2013; Sperandio et al., 2012). **Critically, however, all of these studies implicate modulation of the position representation of the object's edges in early visual cortex** (though not necessarily only in V1). This suggests that early retinotopic stimulus representations encode apparent size. So while higher visual regions obviously process more complex information than V1, such as detecting curved boundaries, segmenting surfaces, or recognizing objects, **numerous studies suggest that whenever this higher processing modulates apparent size it seems to do so by modulating activity in early visual cortex**. The extent to which these modulations are required for perceptual judgments is an issue that will need to be addressed in future experiments, e.g. through the use of brain stimulation methods, and is outside the scope of our study. **In the present manuscript, we now show that even simple perceptual biases – without any contextual modulation and presumably without the involvement of feedback connections from higher regions – are also linked to the encoding of stimuli in early visual cortex**. More on this issue in response to the next point.

On the other hand, if one accepts that visual information in extrastriate areas comes mostly through V1, then any visual judgments is likely to rely on V1 signals in some way, a weaker claim, and one that would not be novel.

We believe our claim is stronger than merely saying '*any visual judgments [are] likely to rely on V1 signals in some way,*' but it is also not the simplistic interpretation that these reviews imply. We hypothesize that modulations of apparent size are encoded by V1 (and other early visual areas) and that these representations are then used for making perceptual decisions. Previous reports have

only suggested that contextual modulations of apparent size are reflected in V1 signals. This left open the possibility that these modulations are irrelevant for the perceptual decision. **Our finding that basic perceptual biases are correlated with the spatial tuning of neuronal populations in V1 provides stronger evidence that signals in early visual cortex are indeed used for the decision making process.** The perceptual biases cannot be explained by any contextual effect that relies on higher level processing. **Instead we show that cortical properties, measured independently using a basic retinotopic mapping design, correlate with perception measured outside the scanner and many months apart.** At the very least, this suggests that these cortical measures are a stable trait that is relevant for biases in size perception.

Moreover, absence of evidence is not evidence of absence – therefore the possibility naturally always remains that some apparent size modulations are encoded only by higher brain regions. However, at least to our knowledge every experiment that has tested the neural correlates of apparent size found these in V1. We now discuss this issue in more detail in our revised manuscript.

Both reviewers consider that the main claim - that object size is read out directly from V1 - is unlikely to be true and not supported by the reported results. The authors need to formulate a more convincing main conclusion.

As we discussed above, we show that even without any contextual illusions consistent biases in size estimates are related to the stimulus representation in V1. There is no tenable explanation for how top-down effects could cause this unless one posits that this pattern is extremely stable across many months and completely different testing environments (psychophysical testing room vs inside the MRI scanner). This strikes us as extremely unlikely. **Our main conclusion is that the decision-making processes underlying fine size judgments use representations in V1 to infer size (and position) because they are affected by idiosyncrasies in V1 functional architecture in highly consistent ways** (i.e. consistent underestimates of size in the periphery).

We now suggest several alternative models for how this result could arise: the simplest, the **impoverishment hypothesis**, was already discussed in earlier psychophysical work (Newsome, 1972) and suggests that impoverished stimulus encoding (due to larger receptive fields and thus poorer acuity) somehow causes consistent underestimation of stimulus size. Newsome regarded this hypothesis as unlikely because dimming the stimuli in an attempt to mimic impoverished stimuli did not have the same effect. The causal relationship is also unclear: Do observers simply judge a stimulus to be smaller because its representation is impoverished, or does an impoverished stimulus truly *appear* smaller?

The second model, the **read-out hypothesis** which we already discussed previously, argues for the latter possibility. Especially for our small stimuli it is possible that coarse stimulus encoding results in a contraction of the activity evoked by the two edges of the stimulus. If size is inferred based on this activity profile in the retinotopic map, this will cause an underestimation of stimulus size. **This model predicts the relationship between the raw perceptual biases we observed for both our stimulus types and the sharpness of neuronal tuning** (and thus also eccentricity and pRF spread).

The third model is the **calibration error hypothesis**. This posits that higher decision-making centers are grossly calibrated to the decrease of cortical magnification across the visual field but that there are residual errors in this calibration. These calibration errors then result in biased perceptual judgments of stimulus size (Anstis, 1998; Helmholtz, 1867). The residual calibration errors are precisely the cortical idiosyncrasies we describe, that is, local variations in surface area and average pRF spread. We cannot entirely rule out this model but it **fails to account for several observations**: first, it is unclear why these calibration errors should consistently result in underestimation of stimulus size. Presumably residuals should manifest in both directions. Second, at least in our data

individual differences in cortical magnification are most pronounced in the central visual field but magnification factors level off in the periphery to comparable levels across observers and visual field locations. Thus it would seem more logical if the largest perceptual biases were observed at parafoveal eccentricities. One remaining possibility (that we also already discussed in the previous manuscript) is that the calibration error interacts with pRF spread. As pRFs become larger, the magnitude of the residual error may be enhanced. However, this still fails to explain why the sign of biases is consistent.

Note that both the second and third model assume that higher-level brain areas must use V1 (and perhaps other early visual cortex) signals to infer stimulus position and size. The actual mechanism could also be a hybrid of the two. In fact, this is probable because the *magnitude* of the predictions the basic read-out model makes requires adjustment. We apologize for the lack of detail on these ideas in our previous manuscript. We now present our conclusions in a more nuanced way and also discuss the need for more investigations of the decisions-making processes and the causal relationships underlying size judgments.

Another reviewer indicates: The authors explore the claim that the apparent size of objects can be inferred from their cortical representation in area V1. If it were true, an object in the periphery would appear many times smaller than the same object at the fovea. Clearly we achieve a much more veridical sense of what is out there so that the effects of cortical topography are for the most part removed by some process of calibration. The size effect the authors describe may then be some residual error in this calibration.

We again apologize if our reasoning was unclear but this was not our argument. **In our previous manuscript we already discussed the notion of calibration error** (please see the response to the previous point for a detailed explanation). Specifically, we already cited previous publications that make this very argument (Anstis, 1998; Helmholtz, 1867) and we ourselves mentioned this possibility (lines 104-6): '**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**.' However, as we already discussed above, one would then suspect that the most severe calibration errors occur near the center of gaze, as this is also where individual differences in cortical magnification are most pronounced. This is not consistent with the stronger biases in the periphery but we cannot rule out this possibility entirely. As we now also discuss, it is unclear why biases should consistently result in underestimates of object size. We therefore think the calibration error hypothesis on its own is not a likely explanation for these perceptual biases. However, it is likely that a combination of our population read-out model and a calibration mechanism produces these biases. We now discuss this issue in our revised manuscript.

*** INTERPRETATION ** The reasoning of how pRF [spread] and/or surface area interacts with perception of stimulus size is hard to follow, for example p. 4 (lines 94-106). It would be far clearer and would make the results more valuable if the authors proposed a specific computation that could relate the measures of brain to the measures of behavior.*

We completely agree that a more theoretical account of how coarser pRFs may result in underestimation of stimulus size will be a valuable addition to this study. We had discussed adding this previously but – perhaps unwisely – we decided not to include this, in part because the manuscript was already very long and complex. Since we have now removed considerable sections, there is more space for a thorough discussion of several models that could account for these findings (see above for a detailed discussion of these models) and we include a basic model for how object size may be inferred from population activity. **This read-out hypothesis closely predicts the relationship between neuronal tuning and perceptual biases we observed in our empirical results**.

In several places, the authors refer to the double annulus results and the single annulus results as indicating different mechanisms, for example p. 5 (lines 129-136). While the pattern of results appears to differ for the two kinds of stimuli, the stimuli themselves also differ! Without a proposed computation, one cannot tell whether the same mechanism would make different predictions for the two kinds of stimuli, or whether different mechanisms are indeed needed to explain the results.

In the section of the manuscript this comment refers to **we in fact stated that both kinds of stimuli produce comparable perceptual biases relative to the reference stimulus**. The difference in magnitude of this bias may very well be due to physical differences between these stimuli. However, **we hypothesize that there are different neural mechanisms for the raw bias (which is due to the stimulus) and the contextual modulation between the inducer and target circles.** This hypothesis is based on the different relationships between the various cortical measures (pRF spread and surface area, respectively) and the raw biases (for either stimulus type) on the one hand and the difference in biases (relative illusion strength) on the other.

The reviewer however still makes a good point. We previously assumed that both the inducer annulus and the target circle were equally affected by raw perceptual biases and this assumption may not hold. As mentioned above, we now include a formal model of how cortical measures relate to perceptual biases and we discuss these possibilities in more detail. This suggests that because the inducer and target circles are spatially very close (below the resolution of V1 pRFs) they indeed are similarly affected by pRF spread. Moreover, **the activity pattern that the read-out hypothesis predicts is entirely consistent with the raw perceptual biases that we observe. That is, size estimates for the Delboeuf stimuli should also become smaller as pRF spread and eccentricity increase.** As we discuss in more detail below, we also posit that in addition to the raw perceptual biases there may however be additional contextual interactions that are mediated by a different factor, presumably the lateral circuitry in V1.

*** PRF [SPREAD] AND SURFACE AREA ** The main results are in figure 2, showing that larger pRFs are correlated with smaller size estimates, and figure 3 (D-F), showing that larger surface area is correlated with smaller size estimates at some eccentricities.*

The reviewer correctly points out that at least for the isolated circles at the innermost eccentricity there is also a positive correlation between raw bias and V1 surface area. This could suggest that the factors linking perceptual biases to pRF spread are distinct from those linking perception to V1 surface area. If smaller V1 areas were accompanied by smaller perceptual size estimates, this could simply follow from the established relationship between pRF spread and V1 area – our results show that this is clearly not the case (more on this below).

In figure 3, the vertical axes are given as V1 surface area as percent of whole cortex. The reported values (0.4% to 1% of the total surface) is smaller than previous reports, perhaps because of incomplete mapping of the visual field. This needs clarification and critical evaluation with regard to the robustness of these estimates.

There is a very simple explanation for this discrepancy that demonstrates there is no issue with the robustness of our measurements. **As we had explicitly stated in both the figure caption and the methods, these surface area measurements correspond to the size of quarter field maps**. Previous reports by one of us reporting V1 surface area relative to the area of the whole cortex (Schwarzkopf and Rees, 2013) instead measured hemifield maps. (Note that this study used similar visual field coverage as we did here, that is, up to 9 deg eccentricity compared to 8.5 deg in the present study). **This means that our estimates of relative V1 surface area should be approximately half of previous reports – precisely what we found** (0.4-1% compared to 1-2% in the previous study).

Furthermore, our surface area estimates may also be reduced subtly compared to previous work as our present analysis used pRF mapping, which has been shown to produce less biased eccentricity estimates than the traditional phase-encoded mapping used in that previous study (Alvarez et al., 2015; Dumoulin and Wandell, 2008). Here we also specifically excluded the central 1 deg from our quantitative analyses. This foveal region is likely to contain artifactual data that may distort estimates of surface area between individuals. Because of the biased eccentricity estimates, this is not trivial with traditional phase-encoded methods; it is however straightforward when using pRF mapping, as this procedure estimates the actual eccentricity of the pRF. We now clarify both these issues in our revised manuscript.

It is hard to reconcile these results with prior literature showing that pRF [spread] is inversely correlated with surface area /cortical magnification (Harvey et al, 2011), or that cortical magnification is correlated with acuity thresholds (Duncan and Boynton, 2003). The logic of Duncan and Boynton, for example, was that greater cortical magnification (surface area) means that more or cortex can analyze a stimulus, and that distinct parts of the stimulus are represented further away from each other on the cortical surface. Harvey et al show that greater CM means smaller pRFs, predicting that greater CM should result in finer spatial resolution, consistent with Duncan and Boynton. Here, one might infer indirectly that greater pRF [spread] correlates with greater SA, since both correlate positively with illusion strength. But the authors do not report the relationship between pRF [spread] and surface area in their subjects, making it hard to compare their results with prior results.

Critically, the claim that our present findings contradict the relationship between pRF spread and surface area is based on a common misconception: Simply because variable A correlates with B, and B correlates with C, it therefore does not follow that A correlates significantly with C**. This assumption only holds when the correlation is either extremely strong or when the analysis has close to maximal power**. This is clearly not the case for typical biological data where correlations usually only explain a fraction of the variance. C may very well also correlate with additional factors that are not linked with A. This will in turn weaken and can even reverse the relationship between A and C.

We now conducted principal component analysis (PCA) to explore the separable components in our data. **This supports our previous conclusion that pRF spread and V1 surface area explain distinct parts of the variance in perceptual biases**. There are separable components that explain the link between V1 area, pRF spread, and perceptual biases. Moreover, while we replicate previous reports of an inverse relationship between pRF spread and V1 area, this relationship is still characterized by considerable unexplained variance and is strongly reduced when expressing V1 area relative to the whole cortex as we did in our previous manuscript. We now present these PCA results in our revised manuscript.

Moreover, the quantification of surface area differs from the quantification of pRF [spread] in that the former is summarized as one number per quadrant, whereas the latter is assigned a different number at each eccentricity (based on a linear fit). Given that the authors correlate each of these measures with behavior separately at each eccentricity, the analysis would be more compelling if they estimated surface area / cortical magnification at each eccentricity, just as they do for pRF [spread].

This is a very good point. We quantified surface area for quadrant field maps rather than for only the narrow eccentricity range of the individual stimuli. There were three reasons for this: First, this enabled us to make a more direct comparison with our previous findings (Schwarzkopf and Rees,

2013; Schwarzkopf et al., 2011). Second, we assumed that surface area estimates are more robust for whole quadrants than for narrow eccentricity bands. While such bin-wise estimates of other measures (like pRF spread) tend to be fairly similar for neighboring voxels, the quantification of the surface area of small patches of cortex is very susceptible to small errors in the cortical reconstruction. Third, since we found that bias estimates are very closely related between different eccentricities within each quadrant (Figure 1D and supplements) we decided that the surface area of the whole quadrant map was reasonable measure.

However, the reviewers are of course correct that this is inconsistent with our analysis of pRF spread. **We therefore now instead calculate the surface area for the individual stimulus locations in a comparable way to what we did for pRF spread**. That is we calculate the surface area of 1 deg wide eccentricity bands (within a range of 1 to 9 deg). We fit a second order polynomial function to these binned data and then extrapolate the surface area at the three stimulus eccentricities. This is done separately for each visual field quadrant in every observer. As suspected, these surface area estimates are a little more variable than those for the whole quadrants. The curves for each individual quadrant are noisier than those for pRF spread. Generally, pRF spread should increase approximately linearly with eccentricity while surface area decreases non-linearly. While we confirm the linear relationship with pRF spread in all quadrants of all individual observers, for surface area there are some quadrants that do not show a non-linear decrease. The validity of these data points is thus not as great for local surface area measures as it is for pRF spread.

However, the general pattern of results we reported previously still holds in this analysis. Perhaps unsurprisingly, these local surface area estimates are also strongly correlated with those for whole quadrants (r between 0.65 and 0.87). This suggests that it probably does not matter which measure is used. We now relegate the analysis of whole quadrant surface area to the supplementary information and report results for local surface area as our main analysis.

Moreover, there is little explanation of how surface was computed, for example over what range of eccentricities, how these values compare to published reports of V1 surface area, and so forth. In fact, the entire description of the surface area calculation is a single sentence at the end of the methods, providing no details, making it difficult for others to replicate the methods.

We apologize for this oversight. This was the unfortunate result of previous edits of the methods to make the text fit into journal restrictions, which meant that these important details were accidentally omitted. We now greatly expanded the detail of the fMRI methods and explain how these calculations were performed.

Overall, the paper provides a lot of detail on the behavior but very little on the brain measurements. Since the primary point of the paper is to demonstrate the relationship between brain measures (pRF [spread] and surface area) and behavior, it is important to have more details on the brain measures, such as a table with the surface area and pRF [spread] functions of each subject in each quadrant, perhaps with a confidence interval on the pRF [spread] fits, as these can be noisy.

This is a very good suggestion. While matrices containing these parameters were previously already included in our publicly available data set, it makes sense to include these findings also in the manuscript. We now show the underlying data plots for all observers and visual field quadrants in a supplementary data file. For pRF spread, these plots include bootstrapped 95% confidence intervals. Note that these confidence intervals are actually fairly narrow for most eccentricity bins.

*** BIAS ** The authors spend considerable effort to discuss the issue of bias, and they try to untangle perceptual bias from decision bias. This section of the paper is confusing, including what is meant by 'artificial bias' or by 'ground truth'.*

As we stated above, we removed these experiments from the revised manuscript but we will nonetheless address these points. We consider 'perceptual bias' to be a different perceptual experience of the size compared to the actual physical stimulus. When a circle shown in the periphery appears as smaller than a physically identical circle in the center of gaze, this is a perceptual bias.

We used 'ground truth' only in the context of the artificial bias experiments and we now realize that this may have been confusing. We used 'artificial bias' to describe a physical difference in the stimuli and 'ground truth' to denote the known magnitude of this artificial bias. For example, a stimulus may be slightly larger than the reference but as far as the stimulus protocol is concerned this difference counts as zero. The ground truth in this situation would be that this stimulus was artificially biased. The rationale here was to mimic the situation where an observer might have a perceptual bias (i.e. that they saw the stimulus as slightly larger) and we conducted this manipulation specifically to test the reliability of the method. We will reconsider this terminology and clarify the rationale of these experiments when we write them up as a separate manuscript.

Apparently, every trial of every size experiment, including the artificial bias experiments, includes 4 peripheral stimuli of various sizes. The 'ground truth' - which of the 4 stimuli is closest in size to the central disc - is always known to the experimenter (at least given a choice of size metric, such as log scale or linear scale). Puzzlingly, the authors write as though the artificial bias experiments have a ground truth and other experiments do not. It is thus unclear how these experiments address the issue of perceptual bias v decision bias.

The MAPS task should minimize decision bias for the same reason other methods developed for this purpose do (Jogan and Stocker, 2014; Morgan et al., 2013): **the influence of decision bias on perceptual estimates is minimized through the presentation of multiple alternative candidate stimuli**. When using the traditional method of constant stimuli (MCS) one would instruct observers to compare a constant reference to a variable test stimulus, for example by asking which of the two stimuli is larger. In this situation an observer might simply choose the test stimulus more/less often whenever they are uncertain about the correct response, that is, when the two stimuli appear identical to them. Their motivation for doing so could be because they can guess the experimenter's hypothesis or other cognitive factors. This can have a dramatic effect on the psychometric function used to infer the perceptual bias (Morgan et al., 2012, 2013). **Our experiments (and similar ones) circumvent this problem because the correct choice, the target with same/similar size as the reference, is not fixed.** Unlike in traditional MCS experiments, all targets are viable test stimuli while the constant reference is never a choice. Thus simply choosing one location whenever one is uncertain would not overestimate perceptual bias. In fact, such a strategy would instead result in a bias estimate near zero as choosing stimuli of opposite signs would cancel out.

Please see the previous comment regarding our definition of 'ground truth'. This specifically referred to the artificial biases – but as the reviewers correctly point out there is a ground truth in all of these experiments and this corresponds to a perceptual bias estimate of zero.

Also unclear is the logic of the 2AFC experiment (figure 7). The task in this experiment is to indicate which of two stimulus pairs contains a biased set of discs (one in which all discs are the same size), in order to test whether the measured biases could be nulled.

As we removed the 2AFC experiment and it is probably tangential to the main claim of the manuscript, we will limit ourselves to very brief responses here. The purpose of this experiment was to test whether biases measured with MCS or MAPS were a closer match to the observer's actual perceptual experience. This allowed us to test our claims regarding the minimization of decision bias (as above).

Hence, when the discs are shown with sizes opposite the subjects' biases (bias factor +1), shouldn't the subjects reliably choose the biased stimulus as the answer, rather than the unbiased stimulus? In other words, shouldn't the percent correct go systematically below 50? The authors write as though with a bias factor of +1, the stimulus set ought to be indistinguishable from a bias of 0. The justification for this prediction and the reasoning about this experiment are unclear.

This is exactly correct and we agree that this point requires more experimental investigation. We already discussed in our previous manuscript that this cannot be explained by the fact that MAPS biases are weaker and that therefore the task was simply more difficult in this condition. If that were the case, performance for the opposite bias factor (-1) should have also been at chance.

However, as the reviewer correctly points out, if the MAPS bias factor +1 perfectly nulled the perceptual bias then performance should be consistently below chance. The reason we did not find this may be because naturally all the four bias estimates are contaminated by some residual variability (predominantly measurement error). Thus, even if the MAPS biases were closer to the observer's actual biases the additive effect of this error would result in making the target (unbiased) stimulus indistinguishable from the biased one. We will further explore this issue in future experiments and add this to a new, separate manuscript.

*** NORMALIZED BIAS ** The authors summarize size judgments in terms of 'normalized perceptual bias', that is, the estimate of the size judgment divided by the standard deviation. They do this because otherwise "raw biases are confounded by discrimination ability". But in fact, the normalized measure confounds bias with discrimination ability.*

Consider plot 1C. The Delboeuf stimulus appears to get smaller with greater eccentricity according to the plot. However, it may just be that the size estimate gets noisier with greater eccentricity.

We chose to normalize biases by dividing the raw biases by the dispersion. This expresses biases as units of just noticeable differences and is standard psychophysical practice. Taking the dispersion (that is, the inverse of acuity) into account is particularly important because these estimates are semi-independent in the four locations. However, the reviewer nonetheless has a very good point here. It could theoretically be the case that by this normalization we actually confound biases and acuities. We have therefore now changed our analysis approach as suggested by the reviewers by using raw perceptual biases only. We have also used weighted statistics for these group data and present this in our supplementary information. **These analyses rule out the trivial explanation that the pattern of results is due to acuity differences: the relationship between bias and eccentricity is qualitatively the same when using actual raw biases or averages weighted by the behavioral precision** (i.e. the reciprocal of the dispersion).

If the authors wish to avoid giving larger weight to more imprecise measures when they compute regressions, then they could compute weighted regressions, while still using the actual size estimates.

We agree that this is a very good suggestion. We now report the correlations for raw biases as our main analyses. **The pattern of results is qualitatively the same as in our previous manuscript.** We also calculated correlations on the actual raw biases weighted by the acuity or when the variance of acuity is removed by partial correlation. These analyses confirm that it is not critical for our main findings whether the acuity is taken into account at all. Because the results for raw biases are the most conservative results, we now only report these in the revised manuscript.

The subtraction of the two normalized curves in figure 1 is also odd. Why should the SNR measure (mean divided by standard deviation) be additive rather than the raw values being additive?

This is indeed a very interesting point that we did not consider previously. While this does not seem to affect the pattern of results in any substantial way, the normalized biases may not be additive. As described above, we have now dropped the normalization of biases. We now calculate the illusion measure as the difference of raw biases for both stimulus types. **The result is qualitatively the same as for normalized biases.**

Note that the authors do not normalize other variables. For example the confidence interval on pRF [spread] varies with voxels, perhaps systematically varying with eccentricity. Yet the authors do not divide the pRF [spread] by the CI in the measurement. The justification for normalizing the bias measures is unclear and this should probably be dropped.

It would certainly be possible to also weight the cortical measures (pRF spread and surface area). We decided not to do so because – unlike for the perceptual biases – there does not appear to be a consistent pattern of differences in their reliability. Please also note that our method for extracting the cortical measures at the different stimulus locations does not simply extract the measure from that location but takes the whole quadrant data into account. As described earlier, we fit a curve (first order polynomial for pRF spread and, in this new version, also second order polynomial for surface area) to the data from individual eccentricity bands (see Figure 2B). We then extrapolate the pRF spread and surface area for the three stimulus eccentricity from these curves. Using the confidence interval for a single eccentricity band would therefore not really be appropriate. Rather we could use the goodness-of-fit for these polynomials for weighting but these are all generally very high, especially for pRF spread (as can be seen by the plots for individual observers we now include). **More importantly, we now dropped the normalized perceptual bias analysis as suggested by the reviewers so these analyses are now consistent.**

*** FIGURE 2 v FIGURE 3 ** Why do the authors plot two behavioral results in figure 2 and three results in figure 3? Why not plot all 3 in each figure? And why not in the same order?*

To avoid confusion, we chose the arrangement of these plots to follow the narrative of the text. However, the reviewer is probably correct that a consistent arrangement of these figures (now Figures 2C-H and Supplementary Figures S3-S5) is more accessible for the reader.

Moreover, it would also be helpful to show a single combined plot of all three eccentricities for each type of stimulus, to ask whether perceived size can be predicted from pRF [spread] irrespective of eccentricity.

We had initially produced such plots. They appear quite cluttered but since the reviewers suggest this could be useful we decided to include these now and relegated the plots separated by eccentricity to the supplementary information.

*** TASKS ** It is not clear what the purpose of the orientation task is in this paper. It is never compared to brain measures. It reads as though this were meant to be a psychophysical methods paper about MAPS for different stimulus types, and then the brain measurements were tacked on at the end. Either the orientation experiments should be dropped or else their purpose clarified in this paper.*

We now removed this experiment from the manuscript. Both artificial bias experiments were conducted during the development of our MAPS procedure. Their main aim was to address the question if MAPS could actually detect an objective bias. We chose orientation as a comparison because the (subjective) perceptual biases for orientation are much weaker than those for stimulus size. In particular, the orientation biases do not show any clear relationship with eccentricity. As such this experiment was a "cleaner" version than the artificial size bias experiment. We will clarify this in our future manuscript containing these behavioral experiments.

*** CLASSICAL ILLUSION ** The phrase 'Classical Illusion Strength' is unclear. The authors did not measure this. They might call it 'Classical Illusion Prediction', as this value is in fact a prediction made from combining the other two measures, not an independent measurement. The prediction may or may not were the authors to actually measure the classical illusion directly. It is not clear how the direct size estimates - the most straightforward measurements - and the illusory size effects are related.*

This is an interesting point. We cannot know that the illusion strength we calculated would be the same as the one we would have obtained, if we had measured the Delboeuf illusion directly (by asking observers to compare the circle with the annulus to an isolated circle at the same location). It is very likely that this measure would be highly similar but we will address this question empirically in future experiments.

We now clarify this issue in the manuscript. However, instead of calling it 'predicted illusion' as the reviewer suggested, **we now use the term '*Relative* illusion strength'** throughout the revised manuscript. We feel using 'predicted' would also be inaccurate because the biases upon which this measure is based were *measured* not predicted. Potentially, 'predicted' would also be confusing because we already report perceptual biases predicted by the read-out model.

The illusion data again obscures the importance of the basic size effects. The interaction between the pRF and V1 measures on basic size and illusion effects is hard to follow.

It seems logical that the illusion effect is somehow related to the raw biases for the two stimulus types. We now emphasize the importance of these raw biases and how they could relate to the illusion. In general we further clarified the discussion of these two potential mechanisms.

Minor: the word 'subjective' in the title is redundant, at least if the authors concur that all experience is by definition subjective.)

The term 'subjective experience' is common parlance for phenomenological consciousness, or qualia. However, this is a reasonable point and we have therefore changed the title to '*Cortical idiosyncrasies predict the perception of object size*'.

References

Alvarez, I., De Haas, B.A., Clark, C.A., Rees, G., and Schwarzkopf, D.S. (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.

Dumoulin, S.O., and Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. NeuroImage *39*, 647–660.

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

de Haas, B., Schwarzkopf, D.S., Anderson, E.J., and Rees, G. (2014). Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. Curr. Biol. CB *24*, R66–R67.

He, D., Mo, C., Wang, Y., and Fang, F. (2015). Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. Exp. Brain Res.

Helmholtz, H. (1867). Handbuch der Physiologischen Optik.

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

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

Morgan, M.J., Melmoth, D., and Solomon, J.A. (2013). Linking hypotheses underlying Class A and Class B methods. Vis. Neurosci. *30*, 197–206.

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

Newsome, L.R. (1972). Visual angle and apparent size of objects in peripheral vision. Percept. Psychophys. *12*, 300–304.

Ni, A.M., Murray, S.O., and Horwitz, G.D. (2014). Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. Curr. Biol. CB.

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

Schwarzkopf, D.S. (2015). Where Is Size in the Brain of the Beholder? Multisensory Res. *28*, 285–296.

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

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

Schwarzkopf, D.S., Anderson, E.J., Haas, B. de, White, S.J., and Rees, G. (2014). Larger Extrastriate Population Receptive Fields in Autism Spectrum Disorders. J. Neurosci. *34*, 2713–2724.

Song, C., Schwarzkopf, D.S., Kanai, R., and Rees, G. (2015). Neural Population Tuning Links Visual Cortical Anatomy to Human Visual Perception. Neuron.

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

Cortical idiosyncrasies predict the perception 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. The way in which the visual system determines the size of objects remains unclear, however. We hypothesize that object size is inferred from neuronal population activity in V1 and predict that idiosyncrasies in cortical functional architecture should therefore explain individual differences in size judgments. Indeed, using novel behavioral methods and functional magnetic resonance imaging (fMRI) we demonstrate that biases in size perception are correlated with the spatial tuning of neuronal populations in healthy volunteers. To explain this relationship, we formulate a population read-out model that directly links the spatial distribution of V1 representations to our perceptual experience of visual size. Altogether, we suggest that the individual perception of simple stimuli is warped by idiosyncrasies in visual cortical organization.

How do we perceive the size of an object? A range of recent observations have lent support to the hypothesis that the visual system generates the perceived size of an object from its cortical representation in early visual cortex[1]. In particular, the spatial spread of neural activity in visual cortex is related to apparent size under a range of contextual modulations[2–7]. The strength of contextual size illusions has further been linked to the cortical territory in V1 that represents the central visual field[8,9]. These findings suggest that lateral connections in V1 may play a central role in size judgements because these interactions are reduced when V1 surface area is larger. Indeed, similar interactions have been argued to underlie the strength of the tilt illusion[10,11], perceptual alternations in binocular rivalry[12], the influence of distractors in visual search tasks[13], and visual working memory capacity[14]. Even the precision of mental imagery co-varies with V1 area[15] suggesting V1 may be used as a 'workspace' for storing mental images whose resolution is better when surface area is larger.

However, the hypothesis that V1 signals are relevant for *subjective* judgments of object size also predicts that the functional architecture of V1 should explain *idiosyncratic biases in basic size perception* (i.e. size judgements that occur in the absence of any contextual/illusory effects). To date this prediction is untested. Previous neuroimaging experiments have focused on modulations of apparent size that must involve additional processing, either due to local interactions between adjacent stimuli in V1 or by a context that likely involves processing in higher visual areas. Others have shown that the *objective* ability to discriminate subtle differences between stimuli is related to cortical magnification and spatial tuning in early visual cortex[11,16,17]. However, no experiment to date showed a relationship between V1 and *subjective* perceptual biases.

Size judgments for small, simple stimuli vary substantially between observers and even across the visual field within the same observer. Previous behavioral research has shown that small visual stimuli appear smaller when they are presented in the periphery[18–20]. A simple explanation for this could be the impoverished encoding of stimuli in peripheral vision. However, dimming stimuli artificially to mimic this effect does not produce the same biases[19]. Another explanation could be that higher brain regions that integrate the perceptual input to V1 into a behavioral decision are poorly calibrated against the decrease in cortical magnification when moving from central to peripheral vision. Small errors in this calibration would cause a residual misestimation of stimulus size based on V1 representations and in turn lead to perceptual misestimation[20]. However, neither of these models can explain why these perceptual biases are consistent underestimates of stimulus size. Impoverished stimulus encoding alone should only result in poorer acuity while residual errors in calibration would be expected to show both under- and overestimation. Furthermore, recent research has also demonstrated reliable heterogeneity in size judgments across the visual field within individual observers at iso-eccentric locations[9,21]. We can consider these variations as a 'perceptual fingerprint' that is unique to each observer. The neural basis of these individual differences however remains unknown.

Here we used fMRI to compare individual functional architecture in V1, specifically the population receptive field (pRF) spread and local cortical surface area, with perceptual biases in size judgments. To do so we developed the Multiple Alternative Perceptual Search (MAPS) task. This approach combines a matching task with analyses similar to reverse correlation[22,23]. Observers search a peripheral array of multiple candidate stimuli for the one whose subjective appearance matches that of a centrally presented reference. This task allows measurement of subjective appearance whilst minimizing the decision confounds that plague more traditional tasks like stimulus adjustment or the method of constant stimuli[24–26]. It further estimates perceptual biases and discrimination acuity while several stimuli are presented simultaneously. We also consider this a more naturalistic task, akin to our daily perceptual judgments (Figure 1A), compared with traditional psychophysical tasks involving single, isolated objects.

Results

Thirteen normal, healthy observers 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. Observers reported which of the four target circles appeared most similar in size to the reference. We fit a model to explain each observer's behavioral responses, with each of the four target locations modeled via the output of a detector tuned to stimulus size. In each trial the detector showing the strongest response was used to predict the observer's behavioral choice. This procedure allowed the estimation of both raw perceptual bias and uncertainty (dispersion) at each location (Figure 1C).

Apparent size depends on eccentricity

Peripheral stimuli appeared smaller on average than the central reference, confirming earlier reports[18,20,27]. This reduction in apparent size increased with stimulus eccentricity (Figure 1D, black curve). 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[28]) so that targets appeared on average larger than the reference (Figure 1D, red curve). This illusory effect interacted with the effect of eccentricity on apparent size, leading again to a gradual reduction in (illusory) size as 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 *Relative 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 1D blue curve; but note that since observers never compared the stimuli directly at iso-eccentric locations this may not fully account for the classical Delboeuf illusion). To summarize, objects appear increasingly smaller as they move into peripheral vision, where the magnitude of size illusions also has an increasing effect (here with the Delboeuf illusion to make them appear larger).

These results cannot be trivially explained by differences in discrimination acuity. Because spatial resolution decreases in peripheral vision, it would theoretically be possible that bias estimates are noisier at greater eccentricities and thus produce this pattern of results, in particular for the Delboeuf stimuli where bias magnitude decreases. To rule out this confound we also calculated mean bias estimates weighted by the precision of observers' size estimates (i.e. the reciprocal of dispersion) at the corresponding locations. The pattern of results is very similar to the one for raw biases (Supplementary Figure S1).

Idiosyncratic biases in size perception

Critically, we next analyzed the *idiosyncratic patterns* of perceptual biases within each observer, both for isolated circles and for Delboeuf stimuli. Biases were strongly correlated for both conditions and across the three eccentricities tested (Figure 1E and Supplementary Figure S2). That is, if observers perceived a strong reduction in the apparent size of a stimulus at a given location, they tended to show strong reductions for the same stimulus type within the same visual field quadrant, regardless of eccentricity.

We further confirmed the reliability of these bias estimates by comparing estimates from two sessions conducted on different days (Supplementary Figure S2C). Moreover, 9 observers who had taken part in the size eccentricity bias experiment also participated in another experiment using the MAPS task approximately one year later. Despite the long time between these experiments and the

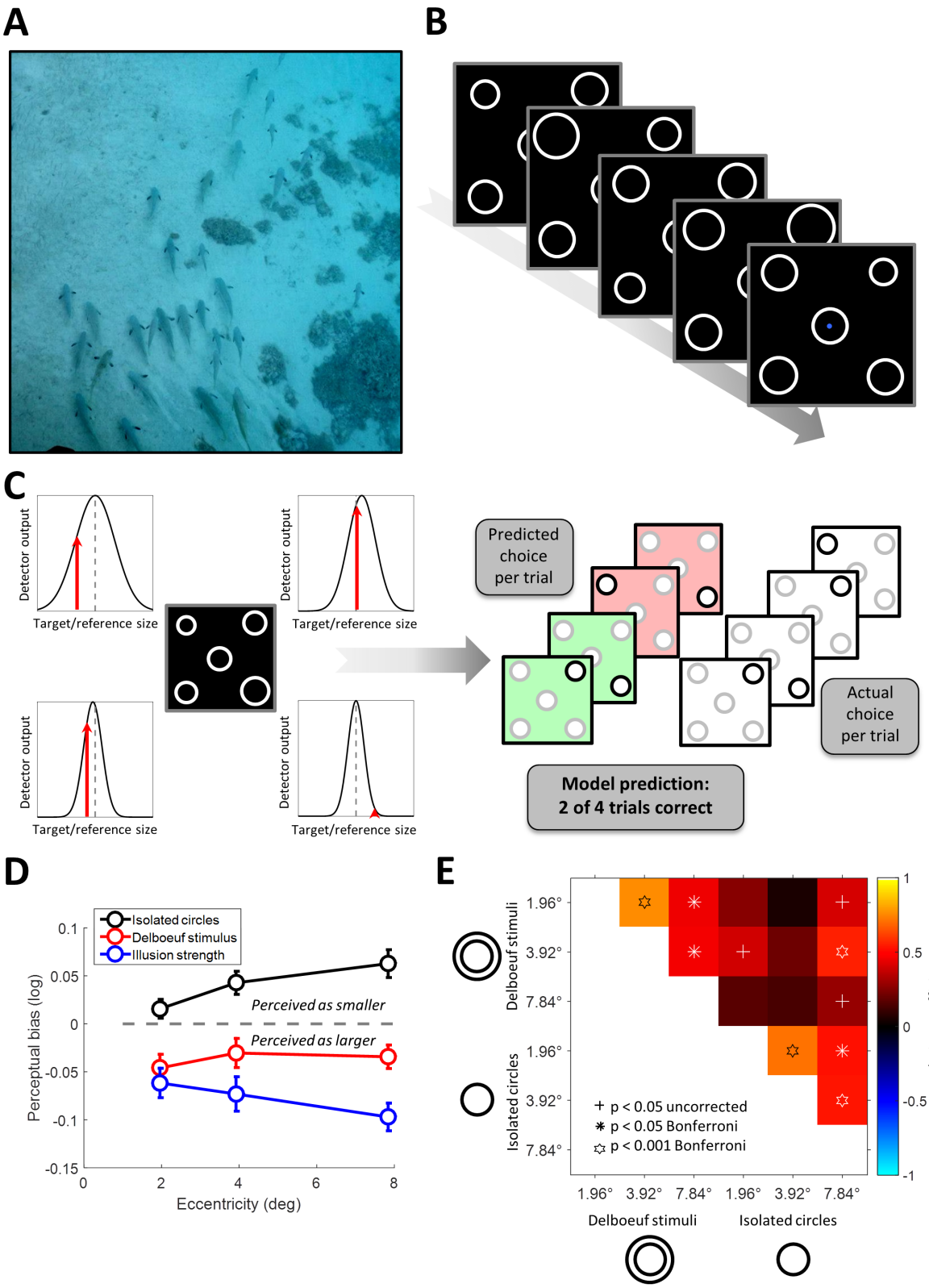


Figure 1. Idiosyncratic biases in size perception. **A.** Visual objects often appear in the presence of similar objects. For example, a spearfisherman may be searching this school for the largest fish. What is the neural basis for this judgment? **B.** The MAPS task. In each trial, observers fixated 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. Each frame here represents the stimulus from one trial. The arrow denotes the flow of time. Observers judged which of the circles in the four corners appeared most similar in size to the central one. **C.** 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. **D.** 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 relative illusion strength (blue), that is, the difference in biases measured for the two stimulus conditions. Error bars denote ±1 standard error of the mean. **E.** 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.

fact that stimulus values were sampled differently (see Methods) the estimates of perceptual biases at target eccentricity 3.92° (which was common to both experiments) were correlated ($r=0.35$, $p=0.0373$). This correlation was largely driven by the within-subject variance. It was considerably greater after subtracting the mean across the four target locations 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 ($r=0.18$, $p=0.6483$).

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[29–31]). Importantly, these neuroimaging experiments were independent from the behavioral experiments and conducted many months later in a different testing environment (MRI scanner vs behavioral testing room).

Interestingly, this analysis revealed a systematic relationship between pRF spread and perceptual biases. With data averaged across observers, increasing eccentricity gives both an increase in pRF spread[29] (see Supplementary Data File 1) and a decrease in apparent size (Figure 1D). We then considered individual data by calculating the correlation between pRF spread and perceptual biases. Both isolated circles (pooled across eccentricity: $r=0.43$, $p<0.0001$; Figure 2C and Supplementary Figure S3A) and Delboeuf stimuli ($r=0.21$, $p=0.0223$; Figure 2D and Supplementary Figure S3B) were perceived as smaller when they were presented at visual field locations covered by voxels with larger pRFs. These individual differences demonstrate that there is correlation between pRFs and apparent size for *idiosyncratic variations* at a *fixed* eccentricity. Our relative illusion strength also showed a negative correlation with pRF spread ($r=-0.22$, $p=0.0166$; Figure 2E), indicating that larger pRFs were associated with smaller differences between raw biases for Delboeuf stimuli and isolated circles. This result was however largely driven by the results for the largest eccentricity (Supplementary Figure S3C).

Basic read-out model of size perception

Why should the apparent size of our circle stimuli be smaller when pRFs are larger? While this result is consistent with the simple impoverishment hypothesis, which states that perceptual biases depend on the precision of the stimulus representation, this alone does not explain why biases are consistent underestimates of stimulus size[19]. To understand this better we conducted a series of simulations that assume that higher brain regions involved in integrating sensory inputs into a perceptual decision about object size read out signals from V1 neurons[32]. We simulated the neuronal activity inside the retinotopic map by passing the actual spatial position of the two edges of the stimulus through a Gaussian filter bank covering that stimulus location. The stimuli were simulated as a binary vector representing 1050 pixels (corresponding to the height of the screen) where the

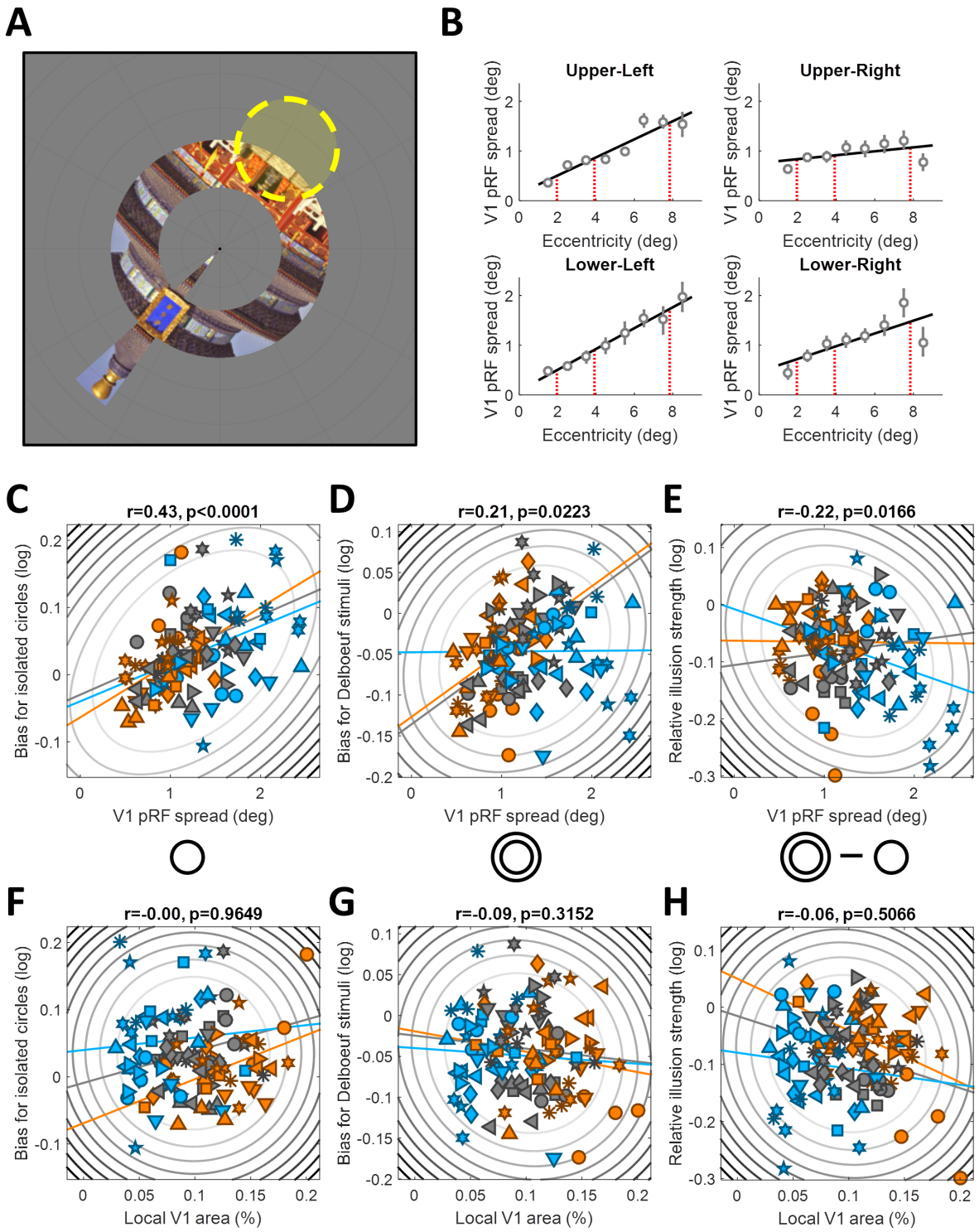


Figure 2. Neural correlates of size perception. **A.** Population receptive field (pRF) mapping with fMRI. Observers viewed natural images presented every 500ms within a combined wedge-and-ring aperture. In alternating runs the wedge rotated clockwise/counterclockwise in discrete steps (1Hz) around the 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 spread corresponding to each target location in the behavioral experiment by fitting a first-order polynomial function (solid black line) to pRF spreads averaged within each eccentricity band for each visual quadrant and extrapolating the pRF spread at the target eccentricities. Grey symbols in the four plots show the pRF spread by eccentricity plots for the four target locations (see insets) in one observer. Grey error bars denote bootstrapped 95% confidence intervals. The solid black line shows the polynomial fit used to estimate pRF spread at each target location. The vertical red dashed lines denote the three stimulus eccentricities at which we extrapolated pRF spread and surface area from the fitted polynomial functions. Data from other observers and V1 surface area measurements are included as Supplementary Information. **C-E.** Perceptual biases for isolated circles (**C**), Delboeuf stimuli (**D**), and the relative

illusion strength (**E**) plotted against pRF spread for each stimulus location and observer. **F-H.** Perceptual biases for isolated circles (**F**), Delboeuf stimuli (**G**), and the relative illusion strength (**H**) plotted against V1 surface area (as percentage of the area of the whole cortical hemisphere) for each stimulus location and observer. In **C-I** Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The colored, straight lines denote the linear regression separately for each eccentricity. Colors denote stimuli at 1.96° (orange), 3.92° (grey), or 7.84° (light blue) eccentricity.

edges were set to 1 while the background was set to 0. The filter bank assumed a Gaussian tuning curve whose width was parameterized at each pixel along this vector. We calculated the response of each filter, to give rise to a population activity profile. Subsequently, higher brain areas then sample this activity to infer stimulus size. This basic model only assumes two layers – however, it is principally the same if activity is submitted from V1 across multiple stages along the visual hierarchy with each layer applying similar filtering.

With this approach, stimulus size can be inferred from the distance between the activity peaks corresponding to the two edges (Figure 3). When the spatial tuning of visual neurons (i.e. neuronal receptive field size) is narrow, the peaks can accurately localize the actual stimulus edges. However, as tuning width increases, the activity profile becomes blurrier. Critically, the space between the two peaks also becomes *narrower* because activity from the two edges is conflated (Figure 3A). It follows that with wider tuning (at greater eccentricity and larger pRF spread), estimates of the separation between peaks decreases until eventually the two peaks merge. This scenario presumably corresponds to far peripheral vision. For the Delboeuf stimuli, the separation of peaks is wider than that of the actual stimulus edges when tuning width is narrow because activity corresponding to the inducer and the target blurs together. However, as tuning width increases the separation also becomes narrower just as for isolated circles (Figure 3B).

To quantify the perceptual biases predicted by this model, we simulated perceptual judgments for both stimulus types and across a range of neuronal tuning widths. The relationship between increasing tuning width and perceptual biases parallels that between empirically observed perceptual biases and stimulus eccentricity (Figure 3C). Size estimates at very small receptive fields (and thus lower eccentricity) are largely accurate but apparent size becomes increasingly smaller than the physical stimulus as tuning width increases. Estimates for the Delboeuf stimuli are generally larger than the physical target. However, as tuning width increases estimates again become smaller. Thus, a large part of the difference in perceptual quality between these two stimulus types may be simply due to the physical difference between them, and the corresponding representation within a population of receptive fields, rather than a more complex interaction between the target and the surrounding annulus. Finally, our *relative* illusion strength is the difference between biases for the two stimulus types. As tuning width increases, this measure in turn becomes larger, just as it does for the empirical data in Figure 2D.

One caveat to this model is that the magnitude of simulated biases is a lot larger than those we observed empirically. This may indicate additional processes involved in calibrating the size judgment. However, it may also be simplistic to infer size from the actual activity peaks. The actual read-out process may instead calculate a confidence range that accounts for the whole function describing the activity profile[33]. The exact relationship between pRF spread and neuronal receptive field size is also unknown. Estimates of pRF spread from fMRI data must aggregate the actual *sizes* of neuronal receptive fields, but also the *range* of center positions of all the receptive fields in the voxel, and their *local* positional scatter within this range. In addition, extra-classical receptive field interactions, response nonlinearities[34], and non-neuronal factors like hemodynamic effects, fixation stability, and head motion must also contribute to some extent. While the simulated tuning widths in our model probably roughly correspond to *neuronal* receptive field size in V1 within our eccentricity range (see e.g. Figure 9B in ref.[29]), an aggregate of the different factors contributing to pRF spread may thus be more appropriate. However, at least qualitatively the relationship between

perceptual biases and tuning width parallels the empirical pattern of perceptual biases and pRF spread estimates that we found.

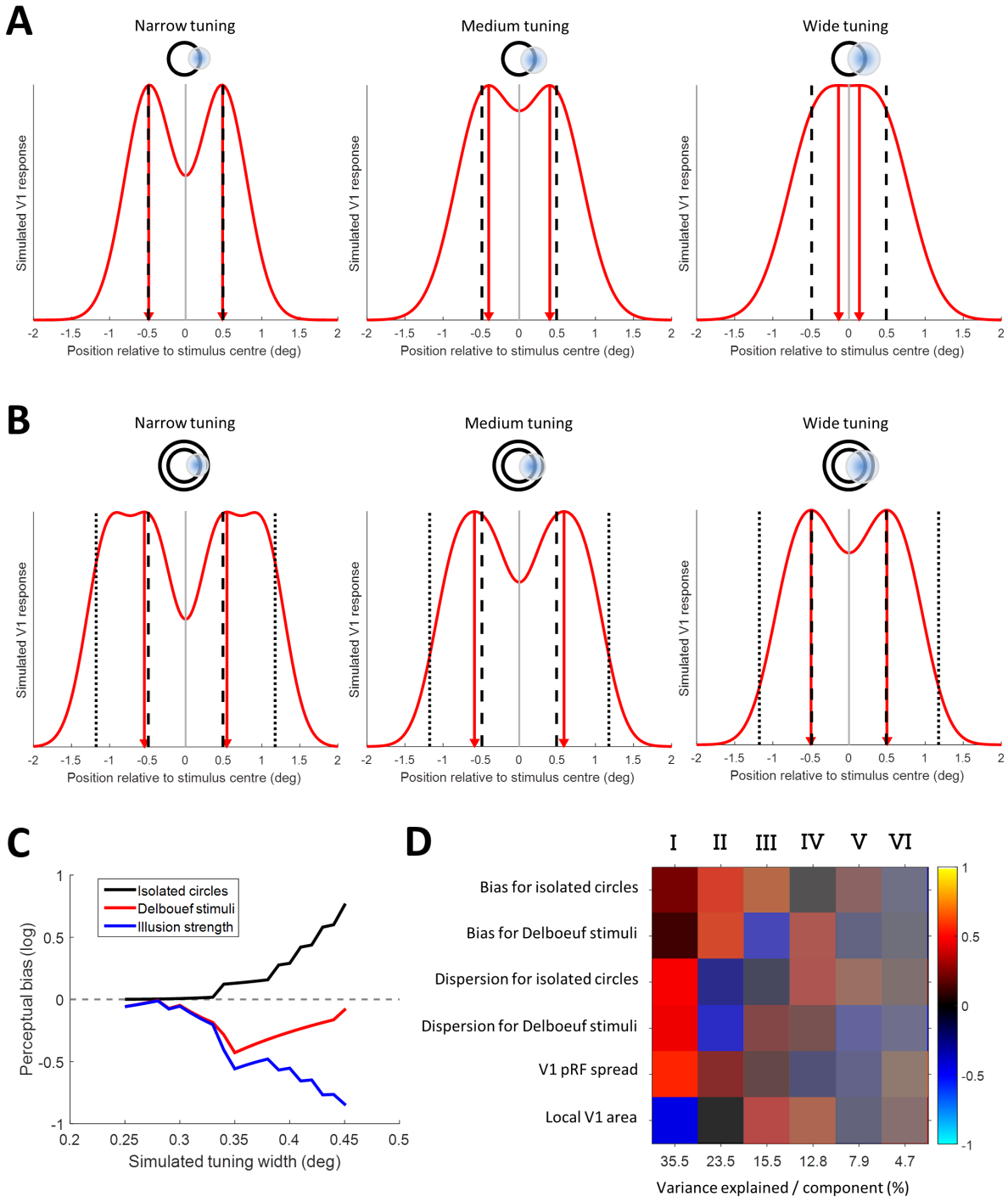


Figure 3. Basic read-out model. **A-B.** Simulated activity profiles for isolated circle (**A**) and Delboeuf stimuli (**B**) were generated by passing the actual location of the stimulus edges through a bank of Gaussian filters covering the stimulus locations. The red curves indicate the output of the filter bank as an simulation of stimulus-related population activity in V1. The separation between the two peaks is an estimate of stimulus size (red triangles and dashed red lines). The vertical black lines denote the actual position of the edges of the target stimulus (dashed) and the inducer annulus in the Delboeuf stimuli (dotted). Simulated neuronal tuning width increases from left to right. This is also illustrated by the schematic diagram above each graph representing the stimulus and an example receptive field. **C.** The simulated perceptual biases for the two stimulus types and the relative illusion strength plotting for a range of tuning widths. Black: isolated circles. Red: Delboeuf stimuli. Blue: Relative illusion strength. See Figure 1D for comparison. **D.** Principal component analysis on a data set combining the biases for isolated circles and Delboeuf stimuli, their respective dispersions (i.e. discrimination acuity), V1 pRF spread, and local V1 surface area (relative to the area of the whole cortical hemisphere). Columns indicate

the six principal components. The numbers on the x-axis show the percentage of variance explained by each component. Each row is one of the six variables. The color of each cell denotes the sign (red: positive, blue: negative) and magnitude of how much each variable contributes to each principal component. The saturation denotes the amount of variance explained by each principal component.

Dissociation between basic perceptual bias and contextual illusions

At the smallest eccentricity of 1.96°, raw perceptual biases for isolated circles also correlated with local V1 area but this pattern was not evident when data were pooled across eccentricity (r=-0.0, p=0.9649; Figure 2F and Supplementary Figure S4A). With Delboeuf stimuli, raw perceptual biases (relative to the central reference) did not correlate with local V1 area at any eccentricity (r=-0.09, p=0.3152; Figure 2G and Supplementary Figure S4B). Because all previous research has compared perception to the macroscopic surface area of the entire central portion of V1[8,9,11–15,17], we further calculated the overall surface area of V1, representing each visual field quadrant between an eccentricity of 1° and 9°. This showed a similar relationship with perceptual biases as local V1 area at the innermost eccentricity (Supplementary Figure S5; isolated circles: r=0.27, p=0.0029; Delboeuf stimuli: r=-0.08, p=0.3968). These results suggest that the variability in perceptual biases is largely driven by differences in cortical magnification for the central visual field: For our innermost eccentricity the relationship between surface area and perceptual measures was always strongest. This variability in central V1 area may thus dominate measurements of the whole quadrant. However, the macroscopic surface area should also be a more stable measure than the area of small local cortical patches. The local surface area and overall area of quadrant maps were very strongly correlated (area relative to whole cortex: r=0.54, p<0.0001; absolute surface area: r=0.54, p<0.0001; see Supplementary Figure S6 for plots separated by eccentricity). Therefore, the macroscopic V1 surface area is a close proxy for local variations in cortical magnification.

In an indirect replication of our earlier findings[8,9,11], we also observed an inverse relationship between the *relative* strength of the Delboeuf illusion (the difference in perceptual bias measured for the two stimulus types) and V1 surface area. Again this was only significant at the smallest eccentricity and not when data were pooled across eccentricities (r=-0.06, p=0.5066; Figure 2H and Supplementary Figure S5C) but it was significant for the overall area of the quadrant map (r=-0.28, p=0.0017). The relative illusion strength (and thus presumably the Delboeuf illusion itself) is the *difference* in apparent size between these stimuli at the *same location*. This measure could be partially independent of pRF spread as it may instead be related to long-range horizontal connections that exceed the voxel size and that mediate the contextual interaction between target and surround.

Under the hypothesis that surface area predicts illusion strength, the bias induced by the illusion differs mechanistically from basic perceptual biases. Both isolated circles (Figure 2C) and Delboeuf stimuli (Figure 2D) were perceived as smaller when pRFs were larger. However, *at the same location* Delboeuf stimuli were nonetheless seen as larger than isolated circles. Even though the apparent size of both isolated circles and Delboeuf stimuli was linked to pRF spread – consistent with the basic read-out model – the *difference* between these biases was also modestly correlated with the area of central V1. The illusion effect may be modulated by cortical distance, possibly via lateral intra-cortical connections[1,10], rather than pRF spread. We conjectured previously that the illusion could arise due to long-range 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[8–11]. In contrast, basic perceptual biases for any stimulus seem to be linked to the coarseness (pRF spread) of the retinotopic stimulus representation itself, which relates to neuronal receptive field sizes and their local positional scatter.

This interpretation may seem to contradict previous findings that pRF spread is inversely related to V1 surface area[17,35]. However, there is considerable additional unexplained variance to this

relationship. To further disentangle the potential underlying factors, we conducted a principal component analysis on a multivariate data set, including z-standardized raw biases for isolated circles and Delboeuf stimuli, the respective dispersions of these distributions (as an indicator of discrimination thresholds), and pRF spread estimates as well as local surface area at corresponding locations in V1. The first four components explained over 87% of the variance (Figure 3D). The first component suggests a positive relationship between pRF spread and dispersion and a negative relationship with V1 area. This supports earlier findings linking pRF spread and cortical magnification to acuity[16,17]. There is however little relation between these measures and perceptual biases. The second component shows a positive relationship between biases for both stimulus types and pRF spread, which reflects our present results (Figure 2C-D). In contrast, the third component involves a negative correlation between biases for the two stimulus types and a positive link between raw biases for isolated circles and V1 area. This may explain the negative correlation between relative illusion strength and V1 area (Supplementary Figure S4C). The fourth component involves a positive link between dispersion for isolated circles and biases for Delboeuf stimuli and also with V1 area. This resembles our earlier findings for orientation processing that also suggest a link between discrimination thresholds for isolated grating stimuli, the strength of the contextual tilt illusion, and V1 surface area[11].

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, as predicted by the read-out model. However, variability in cortical surface area (and thus the scale required of intra-cortical connections) also seems to be an important factor in the illusory modulation of apparent 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.

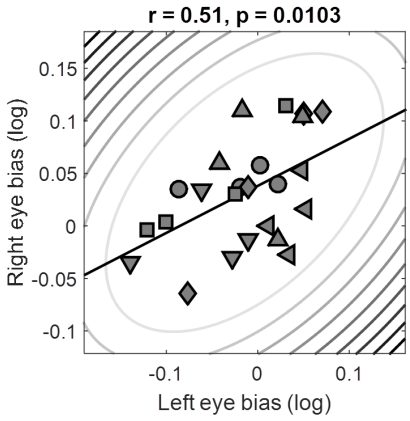


Figure 4. Perceptual biases measured under dichoptic presentation. Biases measured with stimuli in the left eye plotted against those measured in the right eye. Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The straight, black lines denote the linear regression.

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, inhomogeneity in retinal organization, or the morphology of retinotopic maps in the lateral geniculate nucleus. We tested this possibility in a behavioral 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 (r=0.51, p=0.0103; Figure 4). Thus at least a large part of

the variance in perceptual biases must arise at a higher stage of visual processing where the input from both eyes has converged, such as the binocular cells in V1.

Discussion

Our experiments show that when the spatial tuning of neuronal populations in V1 is coarse, visual objects are experienced as *smaller*. These findings support the hypothesis that object size is inferred by decision-making processes from the retinotopic representations in V1[1] and possibly other early visual regions. Our results are therefore consistent with previous reports of a neural signature of apparent size in V1 responses[2–7]. Here we demonstrate that raw perceptual biases are correlated with spatial tuning of neuronal populations in V1. This provides strong evidence that the representation in early visual cortex is indeed used for perceptual decisions about stimulus size, because the biases we observed were independent from contextual or top-down modulation of early visual cortex. Considering that perceptual biases correlate with cortical measures acquired a year later and under completely independent conditions (MRI scanner vs behavioral testing room) we posit that this link between cortical measures and perception is a stable feature of the human visual system.

We have formulated a basic read-out model by which higher-level brain regions sample the activity in early retinotopic maps to infer stimulus size. This model predicts the relationship we observed between eccentricity, pRF spread, and raw perceptual biases measured behaviorally. This was true both for simple, isolated circle stimuli and the Delboeuf stimuli in which the target was surrounded by an annulus. 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 spread, the local surface area (a measure of cortical distance) of at least the part of V1 representing the very central visual field also explains some variance in contextual modulation of apparent size in these illusions. This underlines the need for a greater understanding of how cortical distance relates to pRF spread. Note however that we only calculated the *relative* strength of the Delboeuf illusion based on the biases measured for the two stimulus types, isolated circles and the contextual stimulus including an annulus. It is possible that this prediction does not fully account for the illusion strength one would measure experimentally.

An alternative explanation to our read-out model is that higher-level decoding mechanisms misestimate the size of the stimulus because they are inadequately calibrated to idiosyncratic differences in cortical magnification[20]. This would cause a residual error between the grossly calibrated read-out that may be reflected in perceptual judgments. This explanation however does not explain why perceptual biases are largely consistent reductions in apparent size. In contrast, our basic read-out model fully accounts for this pattern of results. However, we do not wish to rule out the calibration error hypothesis entirely and in fact a hybrid of the two is certainly possible. In particular, in foveal vision, where individual differences in cortical magnification are far more pronounced[30], calibration errors are likely. Moreover, the perceptual biases predicted by our basic model are considerably larger than those we observed empirically (even though the relationship with eccentricity parallels the observed data). This finding is consistent with a calibration mechanism that compensates for the incorrect estimation based on basic read-out of the activity in V1.

Naturally, absence of evidence is not evidence of absence. It is entirely possible that some modulations of apparent size are mediated solely by higher-level brain regions but are not represented in early visual cortex. Moreover, we do not mean to imply that other visual areas are not involved even in our experiments. Differences in stimulus representations caused by

idiosyncratic spatial tuning should be inherited by areas downstream the visual hierarchy, such as V2 and V3. Signals in these regions may still be used for perceptual judgments. Future research must explore the neural substrate of size judgments, in particular with regard to where in the brain the sensory input is integrated into a perceptual decision[1]. Brain stimulation techniques may help to understand the causal link between early visual cortex and higher decision-making centers. Regardless, our present findings imply that measurements of functional architecture in early sensory cortex can predict individual differences not only of objective discrimination abilities but also our subjective experience of the world.

Acknowledgements

This research was supported by an ERC Starting Grant (310829) to DSS, a Deutsche Forschungsgemeinschaft (Ha 7574/1-1) to BdH, a UCL Graduate School Bridging Fund and a Wellcome Trust Institutional Strategic Support Fund to CM, and a UK MRC Career Development Award (MR/K024817/1) to JAG.

Methods

Observers

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

Ten observers (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 observers (1 female; all right-handed; ages 20-25) took part in behavioral experiments only but could not be recruited for the fMRI sessions (which commenced several months later and were conducted over the course of a year). 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. Nine of the observers from the size eccentricity bias experiment (3 authors; 3 female; 1 left-handed; ages 25-37 at second test) took part in an additional behavioral experiment approximately one year after the first measuring perceptual biases in size perception (*long-term bias reliability*). Six observers (5 authors; 2 female; all right-handed; ages 21-36) participated in the dichoptic control experiment (*dichoptic bias*).

General psychophysical procedure

Observers 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. Observers 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 observers 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.

Multiple Alternatives Perceptual Search (MAPS) procedure

To estimate perceptual biases efficiently at four visual field locations we developed the MAPS procedure. This is a matching paradigm using analyses related to reverse correlation or classification image approaches[22,23] that seeks to directly estimate the points of subjective equality, whilst also allowing an inference of discrimination ability.

Stimuli

All stimuli were generated and displayed using MATLAB (The MathWorks Inc., Natick, MA) and the Psychophysics Toolbox version 3[36]. The stimuli in all the size discrimination experiments comprised light grey (54cd/m^2) circle outlines presented on a black background. Each stimulus array consisted of five circles (Figure 1B). One, the *reference*, was presented in the center of the screen and was always constant in size (diameter: 0.98° visual angle). The remaining four, the *targets*, varied in size from trial to trial and independently from each other. They were presented at the four diagonal polar angles and at a distance of 3.92° visual angle from the reference, except for the size eccentricity bias experiment where target eccentricity could be 1.96°, 3.92°, or 7.84° visual angle. To measure the bias under the Delboeuf illusion, a larger inducer circle (diameter: 2.35°) surrounded each of the four target circles (but not the reference) to produce a contextual modulation of apparent size.

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

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. Observers 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.

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 observers would frequently make mistakes. See Supplementary Information for further details on the task procedure.

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

Size eccentricity bias experiment: Observers 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.

Long-term bias reliability experiment: Half of the experimental runs observers performed measured their baseline biases. The other half of the runs contained artificially induced biases: two of the four targets were altered subtly: one by adding and one by subtracting 0.1 log units. Which two targets were altered was counterbalanced across observers, as was the order of experimental runs. Observers were recruited for only one session comprising four runs (two with artificial bias) plus an additional run measuring biases for the Delboeuf stimuli. There were 10 blocks per experimental run. Only the results of the baseline biases (i.e. without artificially induced bias) are presented in the present study. The remainder of these experiments form part of another study and will be presented elsewhere.

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: 8.1-10.5° 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. Observers performed two such runs within a single session.

Analysis

To estimate perceptual biases we fit a model to predict a given observer's behavioral response in each trial (Figure 1C). 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[37]. We initialized the μ parameter as the mean stimulus value (offset in logarithmic size ratio from 0) 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.

Retinotopic mapping experiment

The same ten observers from the size eccentricity bias experiment participated in two sessions of retinotopic mapping in a Siemens Avanto 1.5T MRI scanner using a 32-channel head coil located at the Birkbeck-UCL Centre for Neuroimaging. The front half of the coil was removed to allow

unrestricted field of view leaving 20 channels. Observers 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.

We used a T2*-weighted multiband 2D echo-planar sequence[38] to acquire 235 functional volumes per pRF mapping run and 310 volumes for a run to estimate the hemodynamic response function (HRF). In addition, we collected a T1-weighted anatomical magnetization-prepared rapid acquisition with gradient echo (MPRAGE) scan with 1 mm isotropic voxels (TR=2730ms, TE=3.57ms) using the full 32-channel head coil.

The method we used for analyzing pRF[29] data has been described previously[30,31]. We used a combined wedge and ring stimulus that contained natural images that changed twice a second (see Supplementary Information for further details on the design of the mapping experiments). 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 observer. It determines the visual field location (x and y in Cartesian coordinates) and the spread (standard deviation) of the pRF plus an overall response amplitude.

Stimuli and task

A polar wedge subtending a polar angle of 12° rotated in 60 discrete steps (one per second) around the fixation dot (diameter: 0.13° surrounded by a 0.25° annulus where contrast ramped up linearly). A ring expanded or contracted, both in width and overall diameter, in 36 logarithmic steps. The maximal eccentricity of the wedge and ring was 8.5°. There were 3 cycles of wedge rotation and 5 cycles of ring expansion/contraction. Each mapping run concluded with 45s of a fixation-only period. At all times a low contrast 'radar screen' pattern (Figure 2A) was superimposed on the screen to aid fixation compliance.

The wedge and ring parts contained colorful natural images (Figure 2A) from Google Image search, which changed every 500ms. They depicted outdoor scenes (tropical beaches, forests, mountains, and rural landscapes), faces, various animals, and pictures of written script (228 images in total). One picture depicted the 'Modern Anderson' clan tartan. These pictures were always rotated in accordance with the current orientation of the wedge. Observers were asked to fixate a fixation dot at all times. With a probability of 0.03 every 200ms the black fixation dot would change color for 200ms to one of the primary and complementary colors or white followed by another 200ms of black. Observers were asked to tap their finger when the dot turned red. To also maintain attention on the mapping stimulus they were asked to tap their finger whenever they saw the tartan image.

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

Preprocessing and pRF modeling

Functional MRI data were first preprocessed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, http://www.fil.ion.ucl.ac.uk/spm/software/spm8). The first 10 volumes were removed to allow the signal to reach equilibrium. We performed intensity bias correction, realignment and unwarping, and coregistration of the functional data to the structural scan, all using default

parameters. We used FreeSurfer (https://surfer.nmr.mgh.harvard.edu/fswiki) for automatic segmentation and reconstruction to create a three-dimensional inflated model of the cortical surfaces for the grey-white matter boundary and the pial surface, respectively[39,40]. We then projected the functional data to the cortical surface by finding for each vertex in the surface mesh the median position between the grey-white matter and pial surfaces in the functional volume. All further analyses were performed in surface space.

We applied linear detrending to the time series from each vertex in each run and then z-standardized them. Alternating pRF mapping runs (i.e. those sharing the same stimulus directions – clockwise/expanding and counterclockwise/contracting) were averaged. These two average runs were then concatenated. We further divided the HRF run into the 10 epochs and averaged them. Only vertexes for which the average response minus the standard error in the first half of the trial was larger than zero were included. The HRFs for these vertices were than averaged and we fit a two-gamma function with four free parameters: the amplitude, the peak latency, the undershoot latency, and the ratio amplitude between peak and undershoot.

Population receptive field analysis was conducted in a two-stage procedure. First, a coarse fit was performed on data smoothed with a large kernel on the spherical surface (FWHM=5mm). We performed an extensive grid search on every permutation of 15 plausible values for x and y, respectively, and a range of pRF spreads from 0.18° to 17° in 34 logarithmic steps (0.2 in binary logarithm). For each permutation we generated a predicted time series by calculating the overlap between a two-dimensional Gaussian pRF profile and a binary aperture of the mapping stimulus for every volume. This time series was then z-standardized and convolved with the subject-specific HRF. The grid search is a very fast operation that computes the set of three pRF parameters that produce the maximal Pearson correlation between the predicted and observed time series for the whole set of search grid parameters and all vertices. This was followed by the slow fine fit. Here we used the parameters identified by the coarse fit to seed an optimization algorithm[37,41] on a vertex by vertex basis to refine the parameter estimates by minimizing the squared residuals between the predicted and observed time series. This stage used the unsmoothed functional data and also included a fourth amplitude parameter to estimate response strength. Finally, the estimates parameter maps were also smoothed on the spherical surface with a modest kernel (FWHM=3mm).

Analysis of functional cortical architecture

We next delineated the early visual regions (specifically V1) manually based on reversals in the polar angle map and the extent of the activated portion of visual cortex along the anterior-posterior axis. We then extracted the pRF parameter data separately from each visual field quadrant represented in V1. Data were divided into eccentricity bands 1° in width starting from 1° eccentricity up to 9°. For each eccentricity band we then calculated mean pRF spread and the sum of surface area estimates. For pRF spread we used the raw, unsmoothed pRF spread estimates produced by our fine-fitting procedure. However, the quantification of surface area requires a smooth gradient in the eccentricity map without any gaps in the map and with minimal position scatter in pRF positions. Therefore, we used the final smoothed parameter maps for this analysis. The results for pRF spread are very consistent when using smoothed parameter maps but we reasoned that unsmoothed data make fewer assumptions.

To extract the parameters for each stimulus location we fit polynomial functions to the relationship between these binned parameters and eccentricity. For pRF spread we used a first order polynomial (i.e. a linear relationship). For surface area we used a second order polynomial. We then extrapolated each person's pRF spread/surface area at the 12 target locations in the behavioral experiments (that is, 4 stimulus locations and 3 eccentricities. Individual plots for each observer and visual field quadrant are included in the Supplementary Information. We also quantified the

macroscopic surface area of each *visual field quadrant* in V1 by summing the surface area between 1° and 9°. This range ensured that artifactual, noisy estimates in the foveal confluence or edge effects well beyond the stimulated region did not introduce spurious differences between individuals. In our main analyses we normalized all surface area measures relative to the whole cortical surface area. However, results are also very consistent for using the square root of the absolute surface area for this analysis.

Data and materials

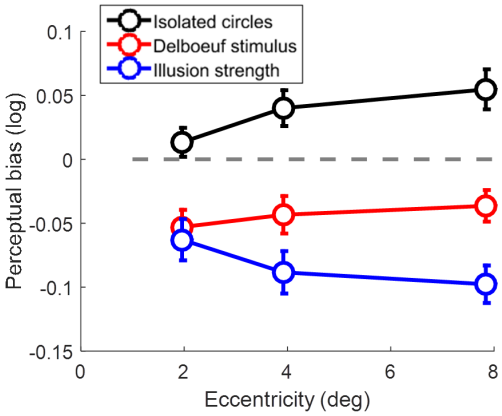
Materials and behavioral data: http://dx.doi.org/10.6084/m9.figshare.1579442
Processed pRF data per observer: http://dx.doi.org/10.5281/zenodo.19150

References

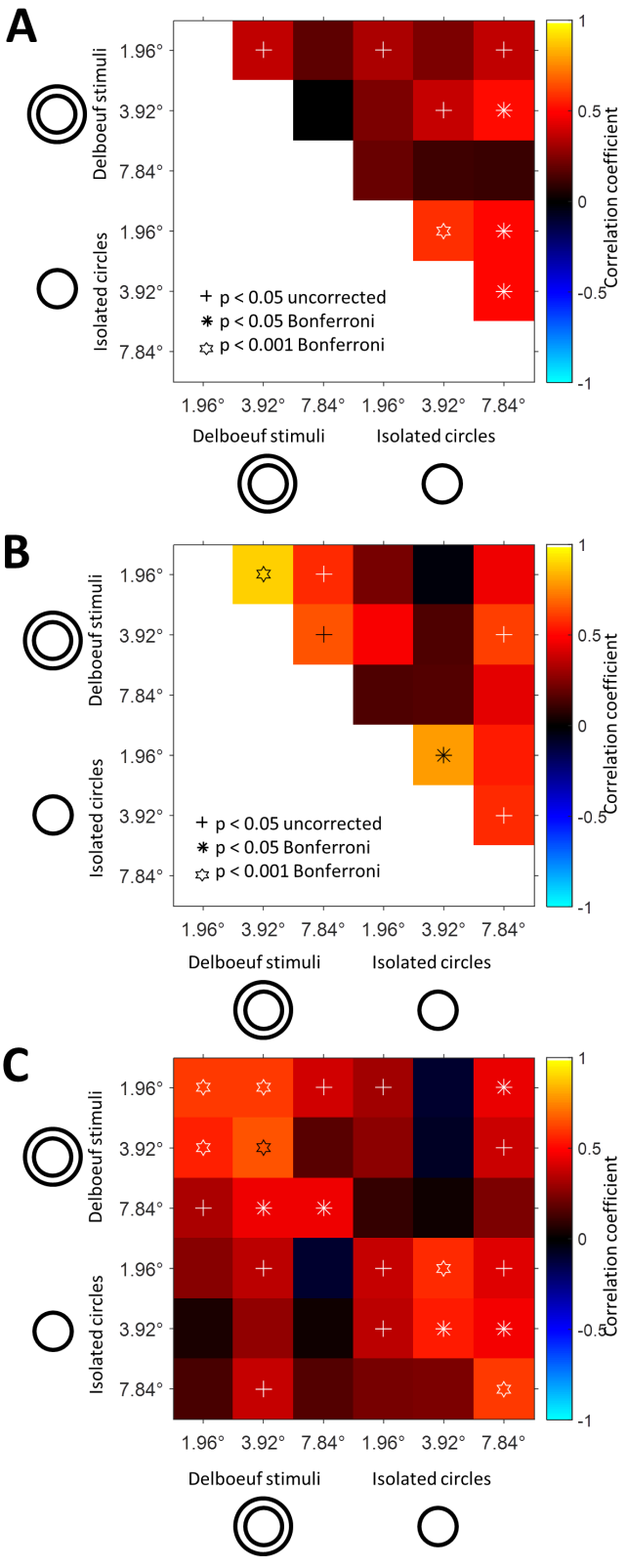
1. Schwarzkopf, D. S. Where Is Size in the Brain of the Beholder? *Multisensory Res.* **28,** 285–296 (2015).
2. Murray, S. O., Boyaci, H. & Kersten, D. The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.* **9,** 429–434 (2006).
3. Fang, F., Boyaci, H., Kersten, D. & Murray, S. O. Attention-dependent representation of a size illusion in human V1. *Curr. Biol.* **18,** 1707–1712 (2008).
4. Sperandio, I., Chouinard, P. A. & Goodale, M. A. Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nat. Neurosci.* **15,** 540–542 (2012).
5. Pooresmaeili, A., Arrighi, R., Biagi, L. & Morrone, M. C. Blood Oxygen Level-Dependent Activation of the Primary Visual Cortex Predicts Size Adaptation Illusion. *J. Neurosci.* **33,** 15999–16008 (2013).
6. Ni, A. M., Murray, S. O. & Horwitz, G. D. Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. *Curr. Biol. CB* (2014). doi:10.1016/j.cub.2014.06.003
7. He, D., Mo, C., Wang, Y. & Fang, F. Position shifts of fMRI-based population receptive fields in human visual cortex induced by Ponzo illusion. *Exp. Brain Res.* (2015). doi:10.1007/s00221-015-4425-3
8. Schwarzkopf, D. S., Song, C. & Rees, G. The surface area of human V1 predicts the subjective experience of object size. *Nat. Neurosci.* **14,** 28–30 (2011).
9. Schwarzkopf, D. S. & Rees, G. Subjective size perception depends on central visual cortical magnification in human v1. *PloS One* **8,** e60550 (2013).
10. Song, C. *et al.* Effective connectivity within human primary visual cortex predicts interindividual diversity in illusory perception. *J. Neurosci. Off. J. Soc. Neurosci.* **33,** 18781–18791 (2013).
11. Song, C., Schwarzkopf, D. S. & Rees, G. Variability in visual cortex size reflects tradeoff between local orientation sensitivity and global orientation modulation. *Nat. Commun.* **4,** 2201 (2013).
12. Genç, E., Bergmann, J., Singer, W. & Kohler, A. Surface Area of Early Visual Cortex Predicts Individual Speed of Traveling Waves During Binocular Rivalry. *Cereb. Cortex N. Y. N 1991* (2014). doi:10.1093/cercor/bht342
13. Verghese, A., Kolbe, S. C., Anderson, A. J., Egan, G. F. & Vidyasagar, T. R. Functional size of human visual area V1: A neural correlate of top-down attention. *NeuroImage* (2014). doi:10.1016/j.neuroimage.2014.02.023
14. Bergmann, J., Genç, E., Kohler, A., Singer, W. & Pearson, J. Neural Anatomy of Primary Visual Cortex Limits Visual Working Memory. *Cereb. Cortex N. Y. N 1991* (2014). doi:10.1093/cercor/bhu168
15. Bergmann, J., Genç, E., Kohler, A., Singer, W. & Pearson, J. Smaller Primary Visual Cortex Is Associated with Stronger, but Less Precise Mental Imagery. *Cereb. Cortex* bhv186 (2015). doi:10.1093/cercor/bhv186
16. Duncan, R. O. & Boynton, G. M. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* **38,** 659–671 (2003).
17. Song, C., Schwarzkopf, D. S., Kanai, R. & Rees, G. Neural Population Tuning Links Visual Cortical Anatomy to Human Visual Perception. *Neuron* (2015). doi:10.1016/j.neuron.2014.12.041
18. Helmholtz, H. Handbuch der Physiologischen Optik. (1867).
19. Newsome, L. R. Visual angle and apparent size of objects in peripheral vision. *Percept. Psychophys.* **12,** 300–304 (1972).
20. Anstis, S. Picturing peripheral acuity. *Perception* **27,** 817–825 (1998).
21. Afraz, A., Pashkam, M. V. & Cavanagh, P. Spatial heterogeneity in the perception of face and form attributes. *Curr. Biol. CB* **20,** 2112–2116 (2010).

22. Abbey, C. K. & Eckstein, M. P. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J. Vis.* **2,** 66–78 (2002).
23. Li, R. W., Levi, D. M. & Klein, S. A. Perceptual learning improves efficiency by re-tuning the decision 'template' for position discrimination. *Nat. Neurosci.* **7,** 178–183 (2004).
24. Morgan, M., Dillenburger, B., Raphael, S. & Solomon, J. A. Observers can voluntarily shift their psychometric functions without losing sensitivity. *Atten. Percept. Psychophys.* **74,** 185–193 (2012).
25. Morgan, M. J., Melmoth, D. & Solomon, J. A. Linking hypotheses underlying Class A and Class B methods. *Vis. Neurosci.* **30,** 197–206 (2013).
26. Jogan, M. & Stocker, A. A. A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *J. Vis.* **14,** 20 (2014).
27. Bedell, H. E. & Johnson, C. A. The perceived size of targets in the peripheral and central visual fields. *Ophthalmic Physiol. Opt. J. Br. Coll. Ophthalmic Opt. Optom.* **4,** 123–131 (1984).
28. Delboeuf, J. Sur une nouvelle illusion d'optique. *Acade ¨ Mie R. Sci. Lett. B.-arts Belg. Bull.* **24,** 545–558 (1892).
29. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *NeuroImage* **39,** 647–660 (2008).
30. Schwarzkopf, D. S., Anderson, E. J., Haas, B. de, White, S. J. & Rees, G. Larger Extrastriate Population Receptive Fields in Autism Spectrum Disorders. *J. Neurosci.* **34,** 2713–2724 (2014).
31. Alvarez, I., De Haas, B. A., Clark, C. A., Rees, G. & Schwarzkopf, D. S. Comparing different stimulus configurations for population receptive field mapping in human fMRI. *Front. Hum. Neurosci.* **9,** 96 (2015).
32. Pouget, A., Dayan, P. & Zemel, R. Information processing with population codes. *Nat. Rev. Neurosci.* **1,** 125–132 (2000).
33. Treue, S., Hol, K. & Rauber, H.-J. Seeing multiple directions of motion—physiology and psychophysics. *Nat. Neurosci.* **3,** 270–276 (2000).
34. Kay, K. N., Winawer, J., Mezer, A. & Wandell, B. A. Compressive spatial summation in human visual cortex. *J. Neurophysiol.* **110,** 481–494 (2013).
35. Harvey, B. M. & Dumoulin, S. O. The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. *J. Neurosci. Off. J. Soc. Neurosci.* **31,** 13604–13612 (2011).
36. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10,** 433–6 (1997).
37. Lagarias, J., Reeds, J., Wright, M. & Wright, P. Convergence properties of the Nelder—Mead simplex method in low dimensions. *SIAM J. Optim.* **9,** 112–147 (1998).
38. Breuer, F. A. *et al.* Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. *Magn. Reson. Med.* **53,** 684–691 (2005).
39. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9,** 179–194 (1999).
40. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* **9,** 195–207 (1999).
41. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **7,** 308–313 (1965).

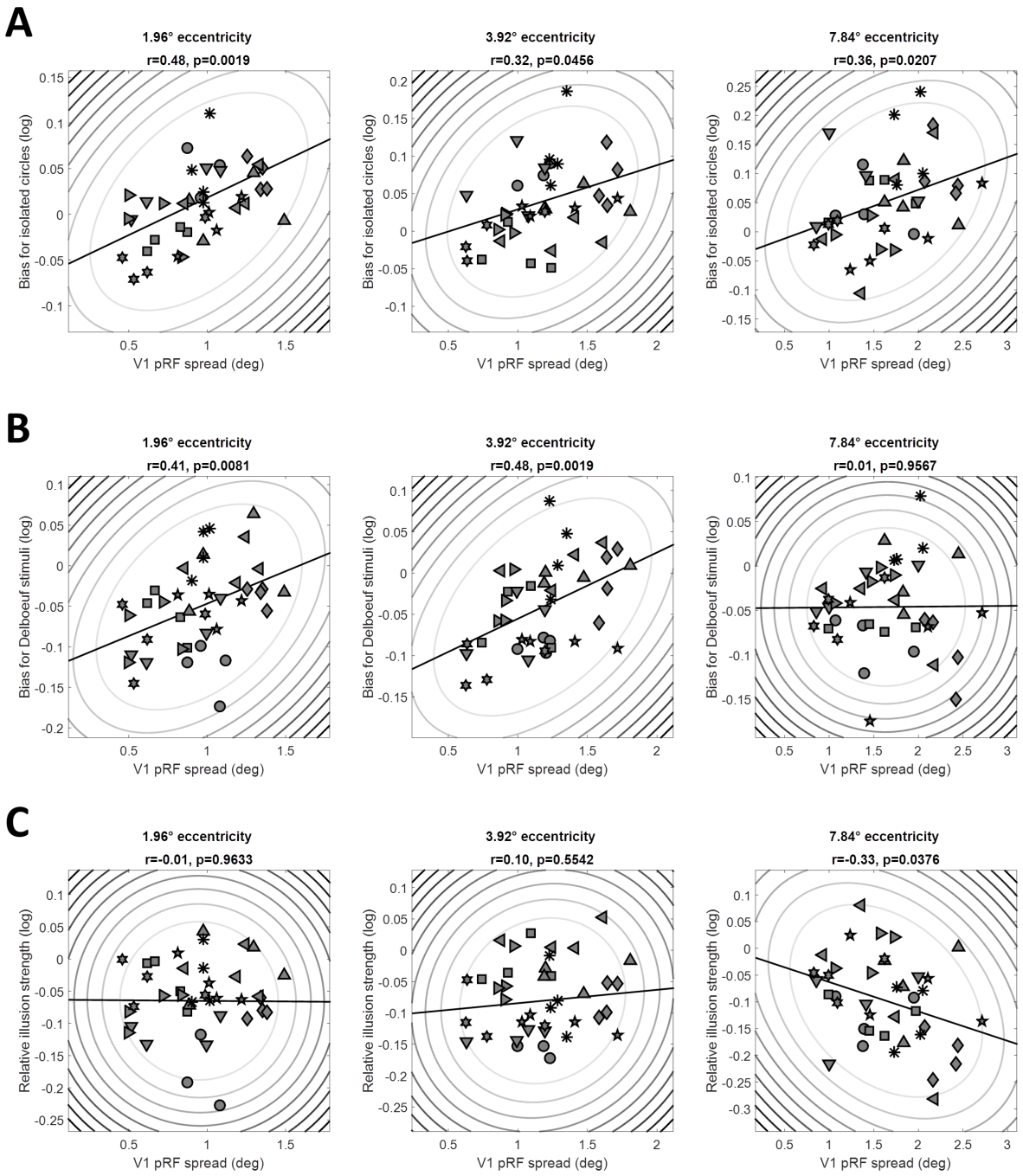
Supplementary Figures



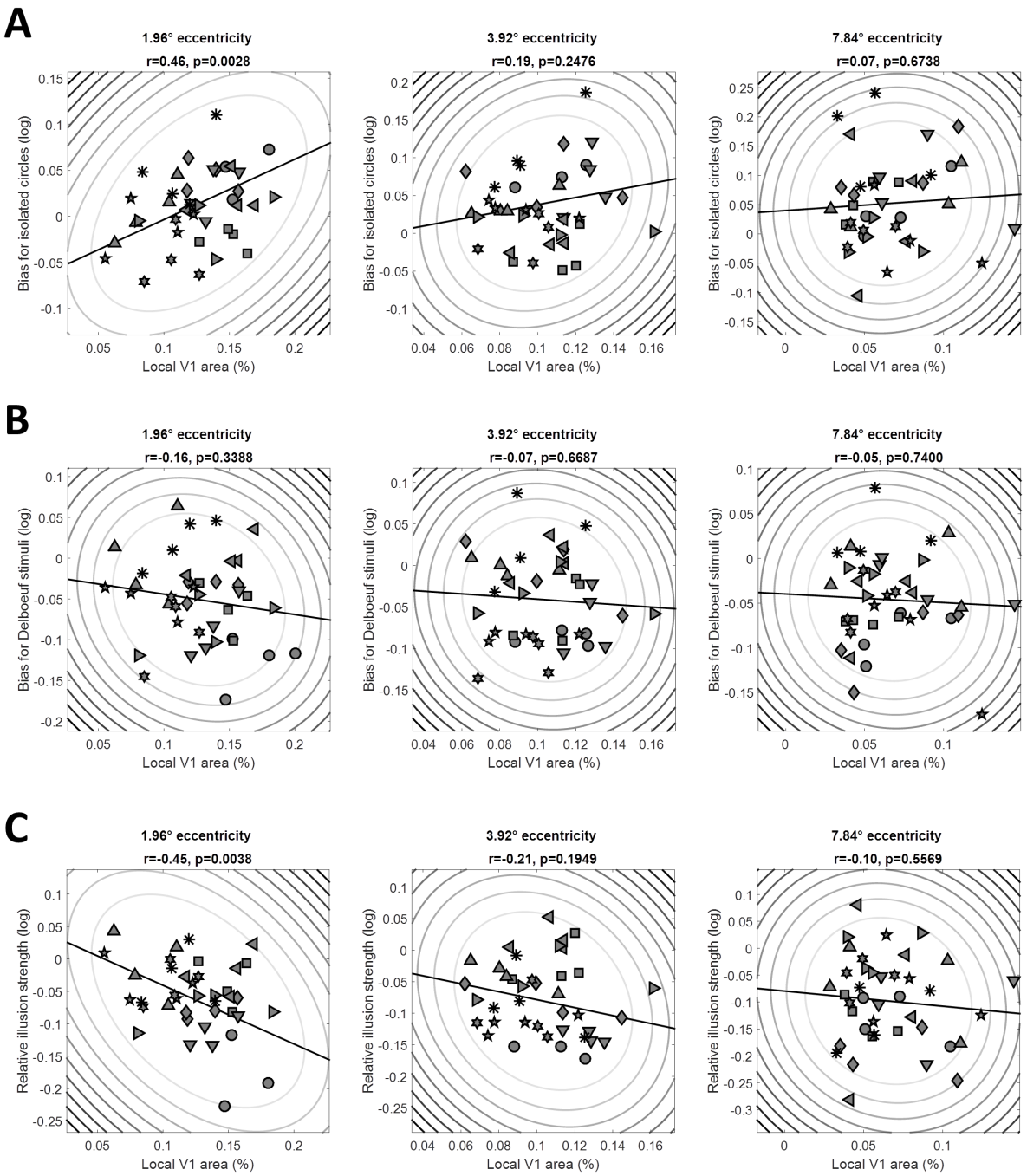
Supplementary Figure S1. Average perceptual bias (positive and negative: target appears smaller or larger than reference, respectively) weighted by the acuity (reciprocal of squared dispersion), across individuals plotted against target eccentricity for simple isolated circles (black), contextual Delboeuf stimuli (red), and relative illusion strength (blue), that is, the difference in biases measured for the two stimulus conditions. Error bars denote ±1 standard error of the mean.



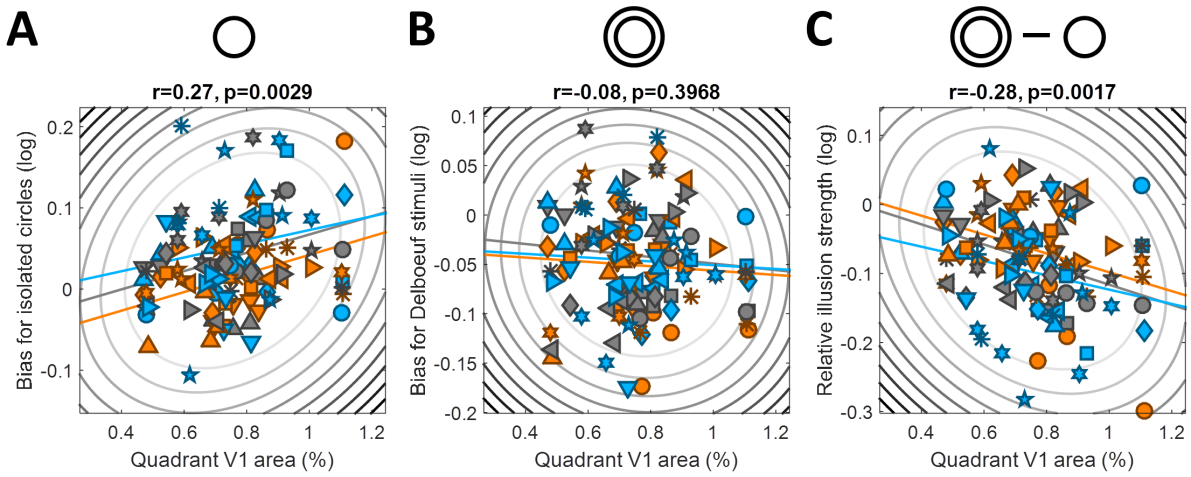
Supplementary Figure S2. Correlation matrices showing the relationship between the perceptual biases in the two conditions (isolated circles and Delboeuf stimuli) and at the three stimulus eccentricities. **A.** Correlations after removing between-subject variance, i.e. the mean across the biases for the four targets was *subtracted* from each condition. **B.** Correlations after removing the within-subject variance, i.e. biases were *averaged* across the four targets in each condition. **C.** Correlations between the first and second session of the experiment conducted on different days. All other conventions as in Figure 1E. Note that statistical power in **B** is lower relative to the other figures, because after averaging there is only a quarter of the number of observations.



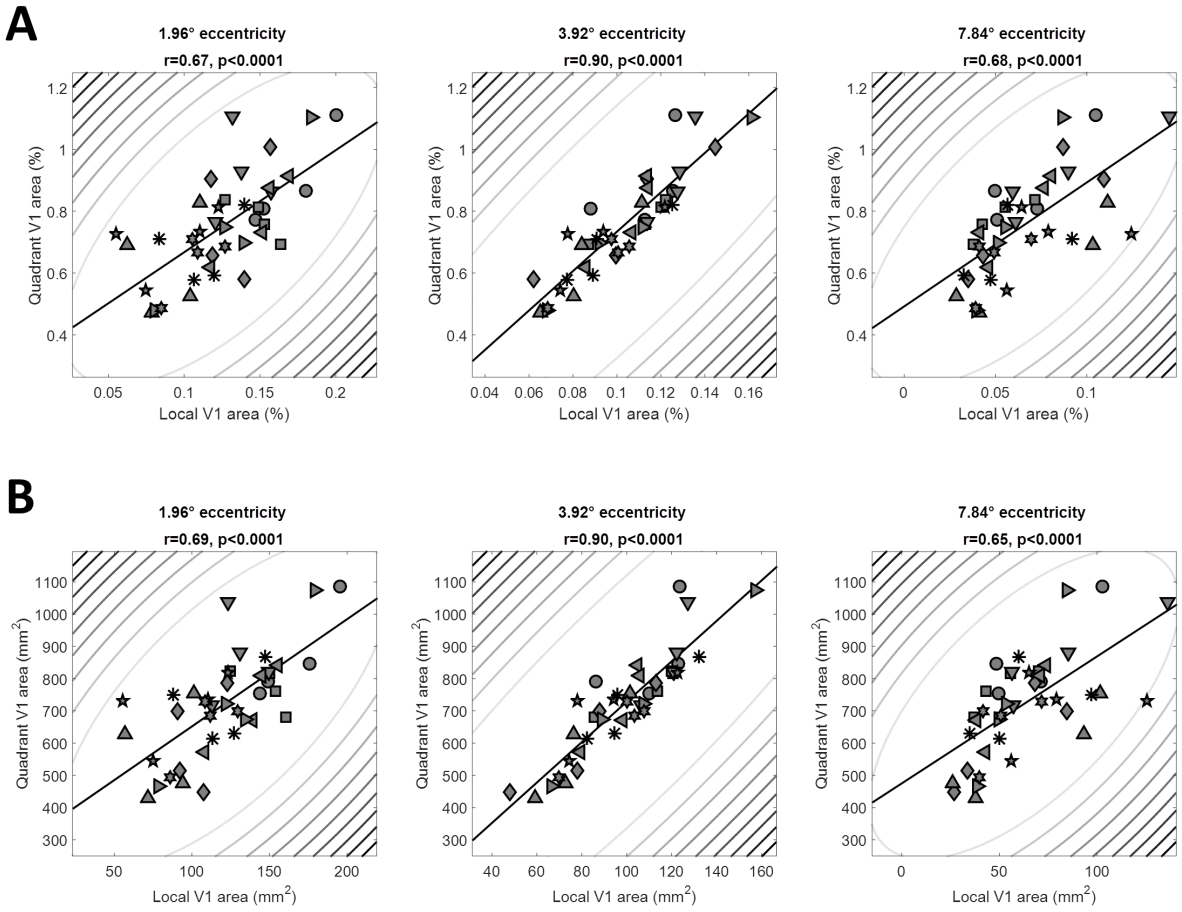
Supplementary Figure S3. Perceptual biases for isolated circles (**A**), for the Delboeuf stimuli (**B**), and the relative illusion strength (**C**), that is, the bias for Delboeuf stimuli minus the bias for isolated circles, plotted against pRF spread at the corresponding location in V1 for each observer and stimulus location. Columns show data for stimuli at 1.96°, 3.92°, or 7.84° eccentricity. Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The straight, black lines denote the linear regression.



Supplementary Figure S4. Perceptual biases for isolated circles (**A**), for the Delboeuf stimuli (**B**), and the relative illusion strength (**C**), that is, the bias for Delboeuf stimuli minus the bias for isolated circles, plotted against the surface area of the corresponding location in V1 for each observer and stimulus location (as percentage of the area of the whole cortical hemisphere). Columns show data for stimuli at 1.96°, 3.92°, or 7.84° eccentricity. Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The straight, black lines denote the linear regression.



Supplementary Figure S5. Perceptual biases for isolated circles (**A**), Delboeuf stimuli (**B**), and the relative illusion strength (**C**) plotted against the surface area (as percentage of the whole cortex) for each quadrant map in V1 between 1° and 9° eccentricity and observer. Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The colored, straight lines denote the linear regression separately for each eccentricity. Colors denote stimuli at 1.96° (orange), 3.92° (grey), or 7.84° (light blue) eccentricity.



Supplementary Figure S6. The surface area of the whole quadrant map in V1 between 1° and 9° eccentricity and each observer plotted against the surface area of the corresponding location in V1 for each observer and target stimulus location. Surface area is expressed either as a percentage of the whole cortical hemisphere (**A**) or as absolute area (**B**). Columns show data for stimuli at 1.96°, 3.92°, or 7.84° eccentricity. Symbols denote individual observers. Elliptic contours denote the Mahalanobis distance from the bivariate mean. The straight, black lines denote the linear regression.

Supplementary Data File caption

Mean V1 pRF spread for the four visual field quadrants (top four panels) and V1 surface area (bottom four panels) plotted for eccentricity bands 1° in width. The vertical dashed red lines indicate the eccentricities of target stimuli in the psychophysical experiments. The solid black lines denote the fitted polynomial functions. Each page shows plots from one observer.