Object image size is a fundamental coding dimension in human vision: new insights and model

Abstract

In previous psychophysical work we found that luminance contrast is integrated over retinal area subject to contrast gain control. If different mechanisms perform this operation for a range of superimposed retinal regions of different sizes, this could provide the basis for size-coding. To test this idea we included two novel features in a standard adaptation paradigm to discount more pedestrian accounts of repulsive size-aftereffects. First, we used spatially jittering luminance-contrast adaptors to avoid simple contour displacement aftereffects. Second, we decoupled adaptor and target spatial frequency to avoid the well-known spatial frequency shift aftereffect. Empirical results indicated strong evidence of a bidirectional size adaptation aftereffect. We show that the textbook population model is inappropriate for our results, and develop our existing model of contrast perception to include multiple size mechanisms with divisive surround-suppression from the largest mechanism. For a given stimulus patch, this delivers a blurred step-function of responses across the population, with contrast and size encoded by the height and lateral position of the step. Unlike for textbook population coding schemes, our human results (*N*=4 male, *N*=4 female) displayed two asymmetries: (i) size aftereffects were greatest for targets smaller than the adaptor, and (ii) on that side of the function, results did not return to baseline, even when targets were 25% of adaptor diameter. Our results and emergent model properties provide evidence for a novel dimension of visual coding (size) and a novel strategy for that coding, consistent with previous results on contrast detection and discrimination for various stimulus sizes.

Keywords: adaptation; size perception; gain control; visual psychophysics; computational model

Introduction

A common view of early vision is that it performs local analyses of the retinal image to encode basic image features. Physiological evidence points to cells in early visual areas with appropriate selectivity, and psychophysical adaptation aftereffects suggest a population basis for the coding. This includes, motion (Mather et al., 2008; Sutherland, 1961), colour (Webster and Leonard, 2008), orientation (Coltheart, 1971; Gibson and Radner, 1937), spatial frequency (Blakemore and Sutton, 1969), binocular disparity (Blakemore and Julesz, 1971; Tsai and Victor, 2003) and duration (Heron et al., 2012; Walker et al., 1981). Adaptation aftereffects are also found in other sensory domains including those for pitch (Christman, 1954), loudness (Marks, 1994), temperature (Arnold et al., 1982), odour (Lawless, 1991) and audio-visual timing (Roach et al., 2011).

These perceptual phenomena have been used to develop our understanding of the sensory coding dimensions used by the human brain, but one obvious candidate missing from the list above is that of size. There are probably several reasons for this. First, retinal image size is often lumped in with retinal spatial frequency (e.g. Webster and Leonard, 2008) but as we shall show, these are two separate adaptable stimulus dimensions with different underlying mechanisms. Second, there is a view that size mechanisms have already been identified psychophysically by using adapt and test stimuli with different diameters, for example. However, as we review in this discussion, results from those studies might derive from perceptual contour displacement,

not size-mechanisms per se. Finally, there is a view that the problem of retinal image size and position is solved in the primary visual cortex because it is retinotopically mapped. While local signs might at least be involved in the encoding of relative position on the retina (Morgan et al., 1990), it is no more suitable as a route for encoding size as it is for orientation, a feature for which dedicated orientation-tuned mechanisms are well-known and celebrated.
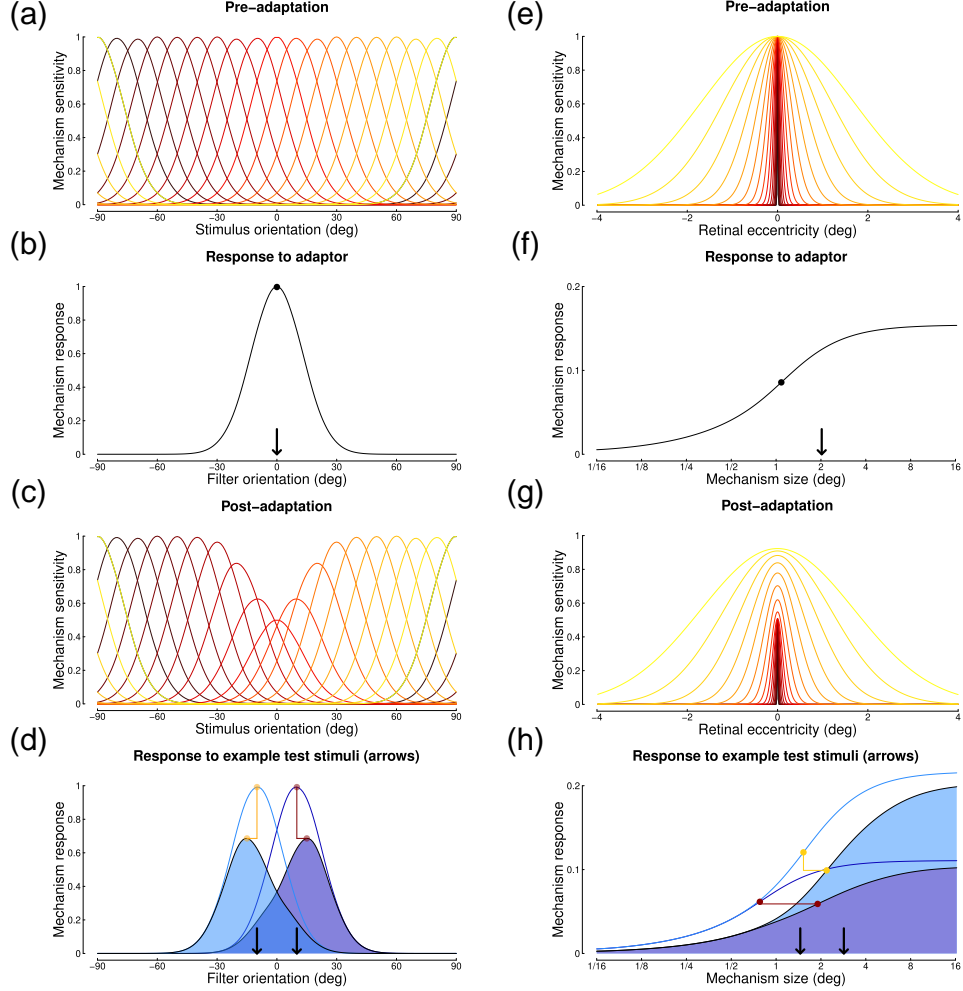

Figure 1: The standard 'textbook' model for orientation-coding and adaptation (the tilt aftereffect) (a-d) applied to size-coding and adaptation (e-h). A bank of orientation-tuned spatial filters (a) responds selectively (b) to stimulus orientation (black arrow) encoded by the peak of the response distribution. Following adaptation at this orientation, the sensitivities of nearby mechanisms are attenuated in proportion to their response to the adaptor (c), distorting the population codes (d) for nearby stimuli (black arrows) and resulting in a repulsive tilt aftereffect. (In (d) the thin curves show the response distributions for two stimuli (black arrows) pre-adaptation, and the filled curves are for those same stimuli post-adaptation.) For a population of size mechanisms, the arrangement of mechanism sensitivities along the coding dimension is different (e) from the orientation case (a). The population response to a stimulus of a given size (e.g. an adaptor; black arrow) resembles a blurred step-edge (f). For this situation the population code is not summarised by the location of its peak, as in (b), but its gradient maximum (black dot). (This happens to be to the left of the nominal stimulus size (black arrow), but this is unimportant for our purposes here). This blurred step-edge distribution means that adaptation desensitises all of the mechanisms smaller than the adapter and, to a lesser extent, some of those that are larger (g). It follows that this model predicts that adaptation will increase the perceived size of stimuli both larger and smaller than the adaptor (note the rightward shifts in gradient maxima (dots) in panel h)—i.e. it is not generally repulsive. The purpose of (e-h) is to illustrate the details of a basic size adaptation model, but thus is not the model we advocate. See text for further details, including the problems involved in trying to achieve an arrangement like that in (a-d) for size.

On the face of it, one might suppose that the size of a patch of grating, for example, could be encoded by a population of size-selective mechanisms along the lines of the standard population model. Figure 1a-d shows the textbook idea for orientation coding and the tilt aftereffect (see caption for details). A similar arrangement can be envisaged for spatial frequency. Note that the narrow tuning of the mechanisms derives from the difference of Gaussian-type arrangement of the underlying receptive fields (i.e. elongated excitatory centres for luminance with subtractive inhibition from flanking surrounds; not shown).

Modelers of size-adaptation might proceed by simply relabeling the orientation axes in Figures 1a-d. However, this would leave the problem of how to get from retinal image size to a standard population code with the appropriate characteristics. For orientation and spatial frequency, this is straightforward, but the situation is not so simple for retinal image size. Figure 1e shows that (unlike for orientation and spatial frequency), a population of mechanisms that integrate luminance contrast over area (Meese and Summers, 2012; Richard et al., 2019) do not sample the coding dimension at regular intervals, but with considerable overlap and a common centre, which means the population response for a given stimulus is not the familiar bell-shaped curve (Figure 1b), but a blurred step-edge (Figure 1f), where the rising region owes to the benefit of contrast integration within mechanism. The arrangement might be revised by supposing difference of Gaussian-type second-order mechanisms with contrast integration windows of various sizes and inhibitory contributions from contrast in the surrounds. However, this is not the remedy it might seem. First, while the arrangement benefits mechanism selectivity, in that small mechanisms are quashed by large stimuli, it fails the other way around, in that small stimuli will excite equally all the mechanisms that are greater than or equal to the stimulus size. A fix for this is to normalise each mechanism's sensitivity by its integration area, but that arrangement kills the benefit of stimulus area at detection threshold for which evidence shows a signal integration process (Meese, 2010; Meese and Summers, 2012, 2009). Second, surround suppression from outside the classical receptive field is not subtractive, but divisive. This is true for cortical cells (Cavanaugh et al., 2002) and for the mechanisms derived from visual psychophysics (Foley, 1994; Meese, 2004; Snowden and Hammett, 1998) and is inconsistent with the assumptions underlying the standard model. Third, regardless of the arrangement for suppression, these mechanisms will confound an increase in stimulus contrast area (within the excitatory part of its mechanisms) with an increase in image contrast. Similarly, as stimulus size increases, larger mechanisms will respond more strongly than the smaller mechanisms did for the same level of contrast (e.g. compare the two solid pre-adapt curves in Figure 1h). This model feature is consistent with contrast detection thresholds (Meese and Summers, 2012) but not with the perception and discrimination of suprathreshold contrast (Cannon and Fullenkamp, 1988; Meese et al., 2005) and is all very different from the population behaviour of the standard model (Figures 1a-d). Of course, one might appeal to the process of contrast gain control (as we have done elsewhere and will do here) but this raises questions about the method of implementation and still leaves the problem of how to extract an estimate of stimulus properties from the population code which is not as simple as it is for the tidier standard model (see Figure 1). We suggest that the first derivative of the response distribution (Figures 1f, h) is a useful approach, and return to this in a later section.

With the above in mind, our motives were two-fold. First, we wanted to know whether we could find evidence for size mechanisms in human vision. Second, should they be evidenced by repulsive adaptation aftereffects, the standard population model would need considerable development for it to provide a plausible account of this result while remaining consistent with contrast perception. However, we were far from daunted. Our previous work on the detection and discrimination of luminance contrast provided evidence for the spatial integration of contrast signals across the retinal image, a sound starting point for the construction of size mechanisms. This contrast integration was found in both the excitatory and suppressive pathways of a contrast gain control network (Baker et al., 2013); Meese and Baker, 2013; Meese and Baker, 2011; Meese and Summers, 2012; Meese and Summers, 2007). For example, by manipulating both stimulus size and extrinsic uncertainty, Meese and Summers (2012) rejected contemporary (signal detection theory) accounts of probability summation for area summation of grating contrast at detection threshold. Instead, those results provided evidence for a process that sums (exponentiated) contrast and internal noise over contiguous regions of the retina. Other experiments using novel spatial stimulus designs tailored to the issue came to the same conclusion (Baldwin and Meese, 2015; Meese, 2010; Meese and Baker, 2011; Meese and Summers, 2009; Meese and Summers, 2007). Furthermore, strong evidence for this spatial integration of contrast was found when the enquiry was extended above threshold using pedestal masking (Meese and Baker, 2011;

Meese and Summers, 2007) and contrast matching (Meese et al., 2017). Crucially, in those experiments, the suprathreshold integration process was seen only with help of 'Battenberg' or 'Swiss cheese' stimuli (essentially, these are gratings containing evenly spaced holes) because under normal circumstances, when the area of both target and pedestal increase together, the benefit of signal integration is hidden by a concomitant process of suppression. This provides the solution to one of the problems outlined above: the visual system does not confound stimulus area with contrast because of the opposing operations of spatial integration and spatial suppression. These findings prompted a cartoon model involving a population of size-mechanisms each subject to suppressive gain control from the largest spatial region in the pool (Meese and Baker, 2011). From this, we developed the hypothesis that spatial pooling of luminance contrast might provide the basis of a population code for retinal image size in visual cortex (Meese and Baker, 2011) (e.g. for a patch of grating), and that adapting this population would distort the perception of stimulus size.

To test this prediction, we conducted a series of novel adaptation experiments designed to isolate the putative size-coding mechanisms predicted by our hypothesis. The outcome is supported by our formalisation of a computational model developed from the earlier work on contrast perception referred to above, where surprising asymmetric features in our size-adaptation data are emergent properties of the model.

Experimental Procedures

Participants

Participants were eight undergraduate students in their early twenties, who participated for course credit. They had no known visual abnormalities and gave written informed consent. The second author also served as a participant in a follow-up experiment. Any prescribed optical correction typically used for near work was also worn during testing. The work was conducted with the ethical approval of Aston's School of Life and Health Sciences (now the College of Health and Life Sciences).

Equipment

Stimuli were displayed on a Phillips MGD403 19-inch greyscale monitor, running at a resolution of 1280×1025 pixels, with a refresh rate of 80Hz. The mean luminance of the monitor was $150 cd/m^2$. The monitor was viewed from 103cm, at which one degree of visual angle subtended 64 pixels on the display. Participants placed their heads in a head-and-chin rest positioned at the appropriate distance and were instructed to fixate on a central cross throughout. We generated stimuli using Matlab and rendered them using a VSG2/5 framestore device (Cambridge Research Systems Ltd., Kent, UK). Participants made responses using a Kensington Expert Mouse Pro trackball device – this is a stationary device housing a ball that could be rotated to provide dynamic adjustment of the stimulus on the screen.

Stimuli

Stimuli were sinusoidal gratings, windowed by a raised cosine envelope with a blur width of 4 pixels added to each side of the plateau of the envelope. (The width of the plateau was the nominal width of the stimulus). All stimuli had a Michelson contrast of 50%. Target stimuli were always horizontal, whereas adapting stimuli were either horizontal or vertical depending on the experiment. In experiments where size was manipulated, all stimuli had a spatial frequency of 4c/deg, and adaptors and targets varied in size (adaptors either 1 or 2 degrees, and targets from 0.25 to 8 degrees at seven levels in proportion to the adaptor; see Figure 3a for examples). In experiments where spatial frequency was manipulated, target stimuli were the same size as the adaptor (either 1 or 2 degrees) but varied in spatial frequency between 1 and 16 c/deg (see Figure 3b for examples).

Procedure

Figure 2 outlines the main procedure. Trials were blocked by adaptation condition throughout. The adaptor was presented to the left of a central fixation cross for 60 seconds at the start of a block (initial adaptation), and 10 seconds before each trial (top-up adaptation). The adaptor was always a sinusoidal grating with

a spatial frequency of 4c/deg and a diameter of either 1 degree (4 cycles) or 2 degrees (8 cycles). During adaptation, the adaptor moved to a random spatial position every 250 ms, with the range of possible positions constrained to lie within the boundary of an invisible circle with a diameter of 4 degrees (for the small adaptor) or 8 degrees (for the large adaptor; see upper panel of Figure 2). After adaptor offset, there was a blank period of 500 ms, followed by a target presentation of 250 ms, accompanied by a beep. The target was offset to the left of fixation by 2.5 degrees (for conditions with a 1 degree adaptor), or 5 degrees (for conditions with a 2 degree adaptor). The target then disappeared, and a matching stimulus was presented to the right of the fixation cross, offset by the same distance. In some blocks this was a thin grey ring, the diameter of which was adjusted by moving the trackball. In the remaining blocks it was another sine-wave grating, and moving the trackball adjusted the spatial frequency. The starting size or spatial frequency of the match had equal probability of being larger or smaller (or lower or higher in frequency) than the target. Once the participant was satisfied with their judgement, they clicked a button on the trackball to initiate the next trial.
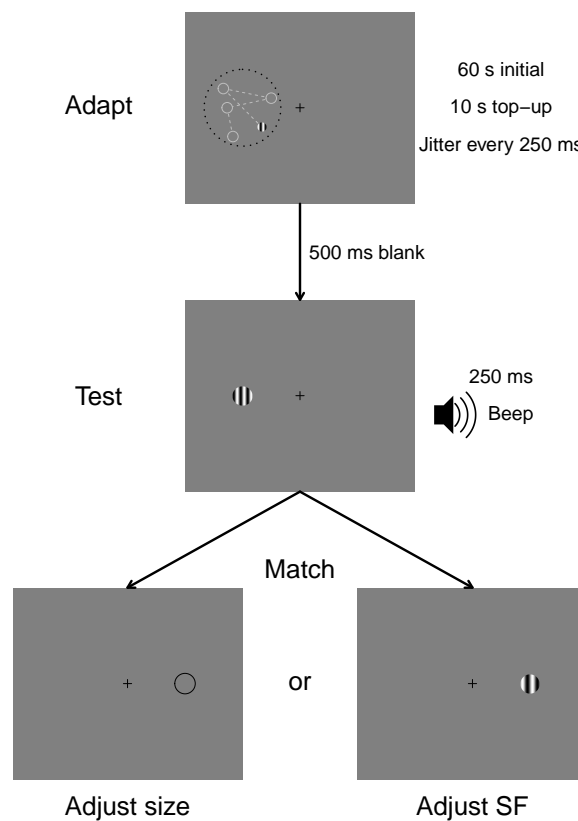

Figure 2: Schematic illustration of the stimulus arrangement and adaptation regime. Adapting stimuli (top panel) moved to random spatial positions to the left of a central fixation cross (the dotted circle shows the region covered by the adaptors) for a 60 second adaptation period at the start of each block of trials, and a 10 second top-up period between trials. The target stimulus was then presented to the left of fixation for 250 ms, indicated by a beep (middle panel). In different experiments, participants adjusted either the diameter of a black ring or the spatial frequency of a grating patch, both shown to the right of fixation in the unadapted region of the visual field, until it matched their perception of the target.

Baseline conditions were as described above, except that there was no adaptation sequence. We collected baseline data for each session before any adaptation took place, and each participant adapted to only one

orientation on a given day. For conditions in the main experiment where the target size varied and spatial frequency was judged, the matching stimulus always had the same diameter as the target. In the control experiment, the matching stimulus had a fixed diameter of either 1 or 2 degrees. Each target condition (target size or spatial frequency) was repeated four times in each block, and each adaptation condition was repeated twice by each participant. Overall, participants completed 1,344 settings each across 168 distinct conditions (7 target levels × 2 adaptor types × 2 adaptation conditions × 6 sub-experiments) in the main experiment, and a further 112 settings each across 14 conditions (7 target sizes × 2 match sizes) in the control experiment.

Quantification and statistical analysis

All analyses were conducted in R. The mean setting (in log units) for each participant and condition was calculated across all 8 repetitions. We calculated the percentage shift in perceived size (or spatial frequency) relative to the baseline settings. The mean and standard error of this adaptation effect were then calculated across participants. We performed a frequentist 2 (adaptor size) by 7 (target size or spatial frequency) repeated measures ANOVA for each experiment using the log values. All degrees of freedom were Greenhouse-Geisser corrected where Mauchly's test of sphericity was significant, and we report generalised eta-squared as a measure of effect size (Olejnik and Algina, 2003). We also conducted a Bayesian ANOVA (Rouder et al., 2012) (using the BayesFactor R package), allowing us to report Bayes factor scores corresponding to the main effects and interactions.
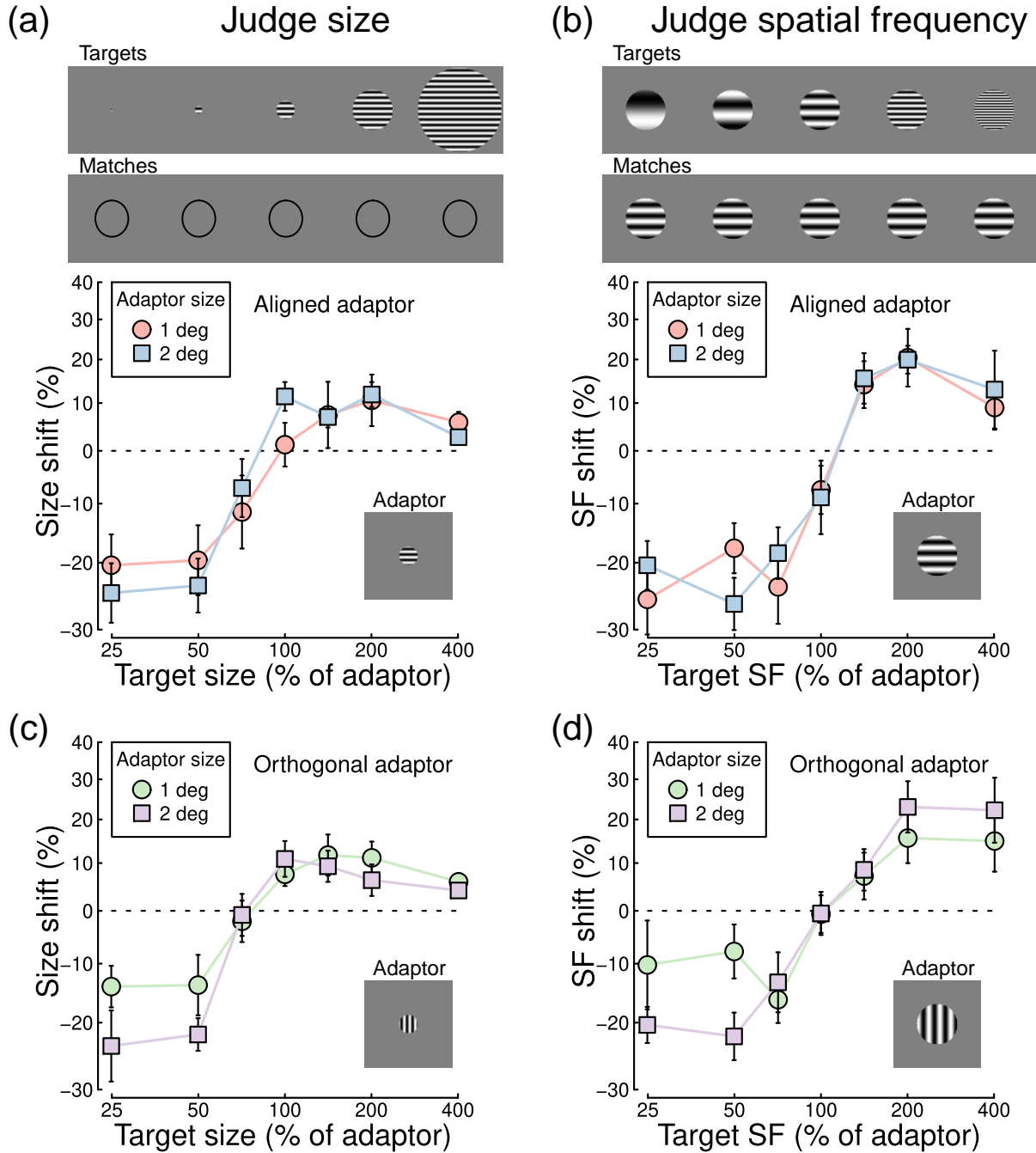
Data and software availability

Experimental scripts, raw data, analysis scripts and modelling code are available at: https://osf.io/kthg3/ (DOI: 10.17605/OSF.IO/KTHG3)

Results

Main Experiments: Manipulations of size, spatial frequency and orientation

Human participants (N=8) were presented with an adapting stimulus (a patch of sine wave grating) on the left side of a computer monitor, which jittered in position (Baker and Meese, 2012; Storrs and Arnold, 2017) every 250ms to avoid causing retinal afterimages (Köhler and Wallach, 1944). They were then shown a target stimulus in the adapted region of the display and asked to report its perceived size by adjusting the diameter of a ring on the opposite side of the display (see Figure 2). Compared to unadapted baselines (which were approximately veridical as summarised in Supplementary Figure S1), when the target was smaller than the adaptor, its perceived size reduced by around 20%. When the target was larger than the adaptor, its perceived size increased by around 10% (see Figure 3a). These effects were consistent across two different sizes of adaptor and are robust and compelling in demonstrations (see Supplementary Movie S1). A factorial (7 stimulus size × 2 adaptor size) repeated measures ANOVA indicated that the main effect of relative stimulus size was significant ($F_{(2.3,16.3)}$ = 18.64, p <0.001, η_g^2 = 0.54, $log_{10}BF_{10}$ = 15.1), but there was no effect of absolute adaptor size ($F_{(1,7)}$ = 0.005, p = 0.95, η_g^2 <0.01, $log_{10}BF_{10}$ = -0.7), nor any interaction ($F_{(6,42)}$ = 1.44, p = 0.22, η_g^2 = 0.04, $log_{10}BF_{10}$ = -0.8).

Figure 3: Adaptation aftereffects for four experiments ($N = 8$ participants). The stimulus icons (top row) depict examples of the target stimuli relating to the x-axes. (They are for the 1 and 2 deg adaptor conditions on the left and right, respectively, where the adaptors are shown by the lower right inset patches.) Panel (a) shows the percentage change in perceived size as a function of target size, expressed relative to the size of the adaptor. Data are shown for two adaptor sizes (different symbols). The spatial frequency was 4 c/deg. For the 2 deg adaptor, the targets had twice the diameter shown by the icons. Panel (b) shows adaptation aftereffects for perceived spatial frequency (SF). Adaptors, targets and matches were the same size. Panels (c) and (d) show results for orthogonal orientations between target and adaptor. All panels show results relative to baseline measures where there was no adaptation (see Supplementary Figure S1). Error bars in all panels indicate $\pm$1SE across participants (N=8), and horizontal dashed lines show baselines in the absence of adaptation.

We then sought to replicate a classic aftereffect in which the coarseness of the texture (but not the size of the patch) is affected by adaptation (Blakemore and Sutton, 1969) using our jittering adaptor (see Supplementary Movie S2). The method remained the same but this time the targets had a constant size and varied in spatial frequency (i.e. bar width). The perceived target spatial frequency was indicated by adjusting the spatial frequency of a matching patch of grating on the opposite side of the display. This experiment also produced a repulsive aftereffect of around 20% in each direction (see Figure 3b; note the logarithmic y-axis). The effect of target spatial frequency was significant ($F(6,42) = 24.74$, p = <0.001, $\eta_g^2 = 0.599$, $log_{10}BF_{10} = 15.88$), but there was no effect of adaptor size ($F(1,7) = 0.05$, p = 0.823, $\eta_g^2 = 0.001$, $log_{10}BF_{10} = -0.7$), nor any interaction ($F(6,42) = 0.59$, p = 0.738, $\eta_g^2 = 0.035$, $log_{10}BF_{10} = -0.88$). Note that spatial frequency-specific aftereffects of this kind were originally referred to as 'size adaptation' (Blakemore and Sutton, 1969) and understood in terms of a population code for spatial frequency. In principle, this scheme can explain the distorting effects (Altan and Boyaci, 2020; Kreutzer et al., 2015; Zeng et al., 2017) of perceived size for luminance-defined objects (see Supplementary Movie S4) by assuming that perception of size is mediated by spatial frequency selective channels. However, this approach does not explain the size adaptation aftereffects reported in Figure 3a where object size and spatial frequency are decoupled; the perceptual judgement here is specific to the size of the patch, not the spatial frequency of the grating texture it contains.

To reveal more about the level of processing at which these adaptation effects occur, we asked whether the aftereffects are tuned for orientation (see Supplementary Movie S3). For size adaptation, we found effects of similar magnitude when the adaptor orientation was orthogonal to that of the target (see Figure 3c). There was a main effect of target size ($F(6,42) = 27.51$, p <0.001, $\eta_g^2 = 0.588$, $log_{10}BF_{10} = 14.97$), but there was no effect of absolute adaptor size ($F(1,7) = 1.01$, p = 0.348, $\eta_g^2 = 0.03$, $log_{10}BF_{10} = -0.36$), or interaction ($F(6,42) = 1.27$, p = 0.292, $\eta_g^2 = 0.06$, $log_{10}BF_{10} = -0.64$). In fact, our informal explorations suggest that the size adaptation aftereffect is insensitive to the texture content of the stimuli and occurs even when this is substantially mismatched between the adaptor and target. For example, we get the same effects when the adaptor is a grating and the target is a face (see Supplementary Movie S5).

For spatial frequency adaptation (Figure 3d), aftereffects were also found when using an orthogonal adaptor, with a significant main effect of target spatial frequency ($F(6, 42) = 18.07$, p = <0.001, $\eta_g^2 = 0.509$, $log_{10}BF_{10}$ = 10.8), but no effect of adaptor size ($F(1, 7) = 0.34$, p = 0.576, $\eta_g^2 = 0.003$, $log_{10}BF_{10}$ = -0.67), or interaction ($F(6, 42) = 1.27$, p = 0.289, $\eta_g^2 = 0.082$, $log_{10}BF_{10}$ = -0.3). Early work on spatial frequency aftereffects showed strong orientation tuning, but did not test target-adaptor orientation differences beyond 40° (Blakemore et al., 1970; Blakemore and Nachmias, 1971). In the context of these classic early studies, our findings with an orthogonal adaptor might be surprising as they appear to show no orientation tuning for the aftereffect. However, subsequent work on the same phenomenon found strong effects on perceived spatial frequency from orthogonal adaptors (Heeley, 1979), consistent with our results.

Do the size and spatial frequency adaptation aftereffects derive from a single process?

Both of our adaptation aftereffects involve aspects of stimulus size – overall size in one case, the width of the stripes in the other. Might these two effects be different aspects of a single phenomenon? For example, when the perceived size of a target shrinks, do its stripes also appear closer together, causing the perceived spatial frequency to increase? To test this, we instructed participants to judge the perceived spatial frequency of the targets from the size adaptation experiment, and the perceived size of the targets from the spatial frequency experiment. Perceived size did not show any clear modulation following spatial frequency adaptation (see Figure 4a) suggesting that our size aftereffect was not a consequence of perceived bar width pulling in or pushing out the perceived diameter of the patch. There were no effects of target spatial frequency ($F(1.07, 7.46) = 0.79$, p = 0.582, $\eta_g^2 = 0.045$, $log_{10}BF_{10} = -1.01$), or adaptor size ($F(1, 7) = 0.44$, p = 0.529, $\eta_g^2 = 0.005$, $log_{10}BF_{10} = -0.58$), nor any interaction between them ($F(1.12, 7.84) = 0.79$, p = 0.585, $\eta_g^2 = 0.036$, $log_{10}BF_{10} = -0.82$). All Bayes Factor scores offered support for the null hypothesis (all $log_{10}BF_{10} < -0.5$).

Judgements of perceived spatial frequency for targets of different sizes did show a reduction in perceived spatial frequency for small target sizes (see Figure 4b). This was confirmed by ANOVA, showing a significant main effect of target size ($F(2.19, 15.33) = 4.99$, p <0.001, $\eta_g^2 = 0.2$, $log_{10}BF_{10} = 1.86$) but no effect of adaptor size ($F(1, 7) = 0.11$, p = 0.753, $\eta_g^2 = 0$, $log_{10}BF_{10} = -0.62$) and no interaction ($F(1.74, 12.17) = 0.5$, p = 0.806, $\eta_g^2 = 0.03$, $log_{10}BF_{10} = -1$). The effect of target size is surprising because the bar width

appears larger (lower spatial frequency) for grating patches that in fact look smaller due to the size adaptation aftereffect (Figure 3a). In other words, the adaptation aftereffect here (Figure 4b) is in the direction opposite to that expected if the size and spatial frequency aftereffects (Blakemore and Sutton, 1969) were caused by a single process.

Note that in our matching paradigm, the target grating would be reduced in perceived size (due to size adaptation), whereas the matching grating would not. Could this mismatch between the perceived sizes of target and adaptor (even when the physical sizes are identical) be somehow responsible for the low spatial frequency settings in Figure 4b?

To test this possibility, our participants completed a control experiment in which the size of the matching stimulus was constant in all conditions. Participants matched the perceived spatial frequency of targets of different sizes to the perceived spatial frequency of a match of fixed size. There was no adaptation. The results (Figure 4c) showed a significant main effect of target size ($F(2.19, 15.33) = 15.17$, p $=<0.001$, $\eta_g^2 = 0.61$, $log_{10}BF_{10} = 17.4$) with small and large targets underestimated and slightly overestimated in spatial frequency, respectively. There was no effect of match size ($F(1, 7) = 0.24$, p $=0.642$, $\eta_g^2 = 0.002$, $log_{10}BF_{10} = -0.69$). There was a significant interaction ($F(6, 42) = 5.35$, p $=<0.001$, $\eta_g^2 = 0.05$, $log_{10}BF_{10} = -1.04$) but the Bayes Factor score indicated greater evidence for the null hypothesis, and the result carries no particular theoretical importance in any case. This control experiment confirmed the perceived spatial frequency bias in the previous experiment (Figure 4b) when target and match stimuli were different sizes, but in this case, it cannot be attributed to size adaptation since there was no adaptor. This suggests that at least some of the spatial frequency effect seen in Figure 4b is a secondary effect deriving from a mismatch in the perceived sizes of the target and match stimuli. Thus, we are confident that the unexpected quirk in our data (Figure 4b) does not undermine our investigation, and that the effect might be related to other spatial frequency biases that have been reported before (Georgeson, 1980; Harris and Wink, 2000). We also conclude that the size and spatial frequency adaptation aftereffects derive from different processes.
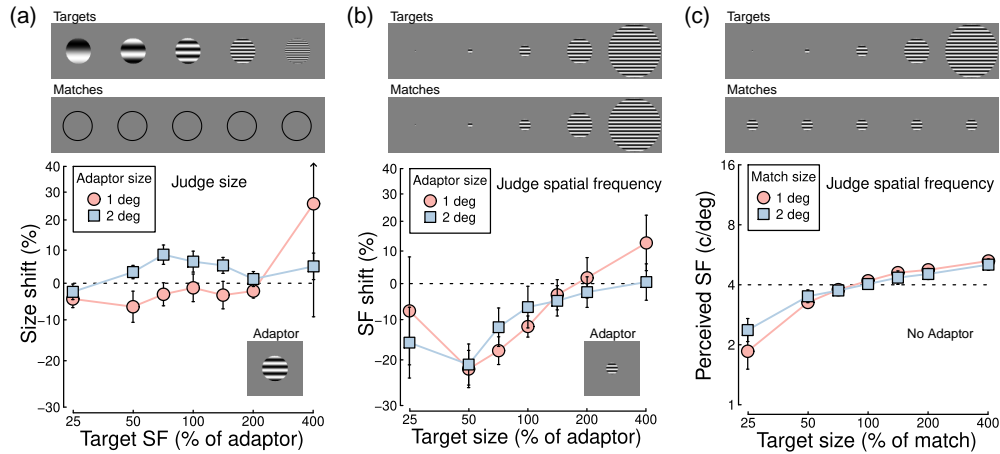

Figure 4: Results from three control experiments ($N = 8$ participants). Panels (a) and (b) show effects of adaptation. Panel (c) shows the effect of size on perceived spatial frequency. The stimulus icons (top row) depict examples of the target stimuli relating to the x-axes. (All are for the 1 deg adaptor conditions, where the adaptors are shown by the lower right insets. Adaptors and targets had twice the diameters shown for the 2 deg adaptor condition.) (a) Perceived size for targets differing in spatial frequency from the 4c/deg adaptor but of the same physical size as the adaptor (see symbol shapes). The matching stimulus was a ring (second row of icons). (b) Perceived spatial frequency for targets differing in size from the adaptor but with the same physical spatial frequency (4 c/deg) as the adaptor. The matching stimulus was a patch of grating which was the same physical size as the target. Panels (a) and (b) show results relative to baseline measures where there was no adaptation. (c) The results of a control experiment in which the spatial frequencies of various sized targets were compared to matching stimuli of fixed size (see symbol shapes) and spatial frequency of 4c/deg (dashed line). There was no adaptor in this experiment. The difference between this experiment and the baseline condition for (b) is that in (c), the size of the matching stimulus was fixed, whereas in (b) it was the same as the target. Error bars show ± 1SE across participants ($N=8$).
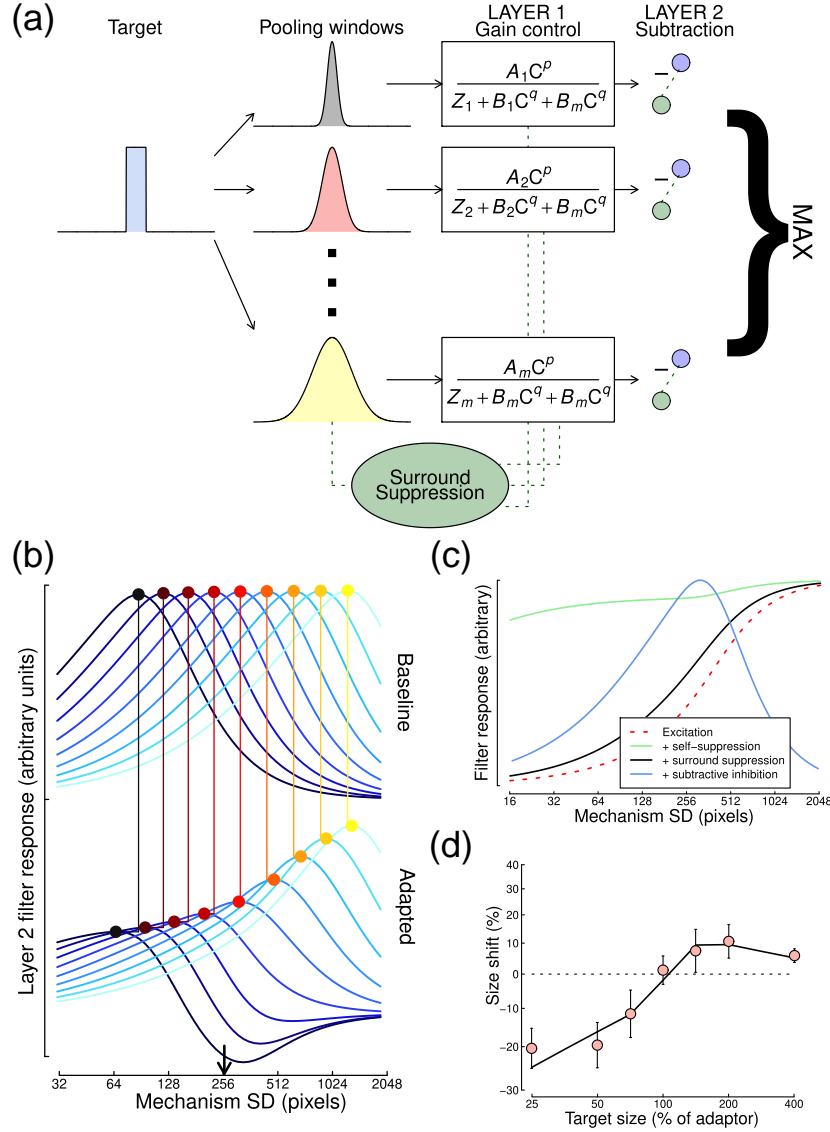
Figure 5: Overview of the computational model. Panel (a) is a schematic of the model. The contrast envelope (C) of a stimulus (blue shading) is multiplied by a bank of contrast pooling mechanisms of different spatial widths (only 3 are depicted here; the model had 91). Since we are not modelling the spatial frequency aftereffect, we have excluded the bandpass spatial filtering details for simplicity. Thus, the first explicit stage of our model (Layer 1) is what is traditionally described as a second-order pooling stage where outputs of spatially tuned linear filter-elements are summed over space, phase, orientation and spatial frequency. The results of pooling are passed through a nonlinear stage of gain control involving surround suppression from the largest window (B_m). Note that in general, the gain control coefficients, A & B, depend on the sizes of both the stimulus and the relevant pooling window as set out formally in the Methods section. The outputs of Layer 1 are then passed to a second layer involving subtractive inhibition between adjacent mechanisms. The size label of the mechanism with the greatest response in Layer 2 (implied by spline interpolation) delivers an estimate of image size. Adaptation increases the value of Z in Layer 1 in proportion to the mechanism's response to the adaptor. Panel (b) shows the Layer 2 responses for different sized excitatory pooling windows (x-axis) to stimuli of different sizes (curves), before (top) and after (bottom) adaptation to a 256 pixel-wide adaptor (indicated by the black arrow). The coloured dots indicate the peak response for each condition. Panel (c) shows the effects of the various inhibitory stages on the response distributions across Layers 1 (red, green, black) and 2 (blue) for a single 256 pixel-wide stimulus. Panel (d) shows the adaptation aftereffect, expressed as the percentage change in perceived size (derived from the lateral positions of the coloured dots in panel (c) ($N = 8$ participants)). The curve is the model behaviour, and the symbols are the 1 deg results replotted from Figure 3a.

10

Computational modelling

To understand the relationship between size adaptation and population coding involving mechanisms with superimposing selectivity for size, we developed a computational model (described more fully in Appendix A). This was guided by earlier work on contrast perception (Meese et al., 2017; Meese and Baker, 2011; e.g. Meese and Summers, 2007), and devised to overcome the difficulties involved with a simple application of the population model for size coding as described earlier (Figure 1).

For simplicity, our spatially one-dimensional model (see Figure 5a) takes stimulus contrast as input over the spatial extent of the stimulus[1] (i.e. it treats the envelope of the stimulus as a local measure of contrast)[2], regardless of spatial frequency and orientation (Meese and Baker, 2013). (Our model was not intended to explain the spatial frequency aftereffect which has been modelled elsewhere (e.g. Klein et al., 1974), but could be extended to do this with little or no impact on our conclusions here). These spatially distributed contrast responses (Moutsiana et al., 2016) are multiplied by each element in a one-dimensional 'size' array of Gaussian pooling mechanisms of various spatial extents (windows) (dashed red curve in Figure 5c), each member of the array being subject to nonlinear gain control. This includes (i) self-suppression to protect the system's image contrast code from the influence of stimulus size above threshold (Meese and Baker, 2011) (green curve in Figure 5c), (ii) a divisive surround suppression term (Cavanaugh et al., 2002; Sengpiel et al., 1998; Webb et al., 2003; Xing and Heeger, 2001), where each mechanism is suppressed by the largest mechanism in the array (black curve in Figure 5c) to provide the initial basis for size coding (Meese and Baker, 2011) and (iii) a saturation constant (Z) that allows for the benefit of stimulus size at detection threshold (Meese and Summers, 2012). This has the same initial value for each mechanism (Meese, 2004) but is influenced by the mechanism's response to the adaptor (R) and the gain of adaptation (α) (Foley and Chen, 1997; Meese and Holmes, 2002) such that $Z_j = 1 + \alpha R_j$, where j indexes the array (i.e. population) of size mechanisms. In general, with no adaptation this model arrangement produces a distribution of activity across the size array (model Layer 1) with the form of a blurred step edge (black curve in Figure 5c). (See Supplementary Figures S3-S5 and their captions for further discussion of the model development with and without adaptation.)

First-order image contrast (Meese and Baker, 2011) (not our focus here) and stimulus size (Meese and Baker, 2011) are encoded by the height and location of the distribution of activity across Layer 1, respectively. Note that mechanisms with excitatory pooling regions greater than or equal to the size of the stimulus have comparable responses because they receive the same excitatory drives and suppression. This sets our Layer 1 population code apart from many others in visual perception, where the usual bandpass properties of the mechanisms involved (Figure 1a) result in peaked (bell-shaped) distributions (Figure 1b). Instead, the computational task here is to find where the response transition in the Layer 1 array is located (see black curve in Figure 5c). Following the logic of edge-emphasis by lateral inhibition in the retina (Barlow, 1953), Layer 2 of our model takes a copy of the Layer 1 responses and subjects each to subtractive inhibition from its immediate neighbour, approximating first-order differentiation. This results in size-tuning of the population (blue curve in Figure 5c), with perceived size being given by the peak of the spline-interpolated distribution (Marr and Hildreth, 1980) (see points in the upper row of Figure 5b).

Adaptation is implemented by setting the saturation constant (Z, see Foley and Chen, 1997; Meese and Holmes, 2002) in the gain control of Layer 1 such that $Z_j - 1$ is proportional to the jth mechanism's response to the adaptor (see Appendix A for details.) Adaptation changes the population size-tuning curves, such that the peaks of the tuning curves in Layer 2 shift away from the adapting stimulus (i.e. smaller sized stimuli produce peaks at smaller mechanism widths, and vice versa), as shown in the lower row of Figure 5b:

---

[1] Our model here is devised to deal with psychophysical stimuli—patches of luminance grating in a nominal world of otherwise uniform luminance. In this simple case, no stopping rule is needed for the contrast integration process at the stimulus boundary since the absence of stimulus contrast does the job. In the more general case, an additional process of image segmentation would be needed to identify the boundary of the integration process. We pick up on this point in the discussion.

[2] To a first approximation this simplification is equivalent to a measure of local (receptive field size) RMS contrast at each point across the image which can be derived from the square root of the sum of the squares of a quadrature pair of filter outputs (i.e. a standard model complex cell). All of our stimuli were uniform patches of grating (albeit of various sizes). Thus, across most of the stimulus, the local RMS contrast is simply the contrast of the stimulus which is uniform across the stimulus region and zero beyond the stimulus region (see footnote 1). By using the stimulus envelope as a proxy for this measure, we are ignoring minor edge effects (contrast blurring) that would be produced by first-order spatial filters had we included a convolution stage in the model.

the peaks (marked by the coloured dots) are repulsed laterally from the location of the adaptor following adaptation. The ratio of adapted to unadapted perceived size is shown in Figure 5d (curve) and produces the same bi-directional size adaptation aftereffect that we observed empirically (symbols). Note that the decreases in size are around twice as large as the increases, and that this asymmetry is present in both model and human behaviour. Further investigation on the second author confirmed that the size reduction effects remained strong for even smaller targets, and that size increase effects returned to baseline for larger targets. This is also predicted by the model (see Supplementary Figure S2). These unusual effects derive from natural asymmetries in our model's architecture. For example, within the mechanisms of Layer 1, adaptation is more influential for smaller integration windows because with less potential for the summing of signal (in the inhibitory pathway as well as the excitatory pathway) these mechanisms are more labile. This is to say that the impact of an increase in Z (see Figure 5a) is greater when the values of A and B ($A_1 \leq A_2 \leq A_m$; and $B_1 \leq B_2 \leq B_m$) are small than when they are large, and since Z is the parameter that carries desensitisation by adaptation, aftereffects are larger for smaller mechanisms. Similarly, for any given mechanism, as target size becomes smaller than the integration window, the impact of adaptation increases for that mechanism. (See Supplementary Figures S4 and S5 and their captions for further discussion.)

In general, the behaviour of our model derives primarily from its architecture; the black curve in Figure 5d was derived by setting a single adaptation parameter (α) by eye (see Methods), yet the model describes the key features of our results (i.e. their bi-directional asymmetries) very well.

Note that the first layer of our model is consistent with our motivating work on area summation of luminance contrast at threshold and above (Meese, 2004; Meese and Baker, 2011; Meese and Summers, 2007), extending it here by pooling over first-order spatial frequency and orientation, the details of which no doubt deserve further investigation. The second, subtractive, layer is a novel extension designed to deliver the retinal image-size code.

Discussion

We have used an adaptation paradigm to provide the first psychophysical evidence for size mechanisms in human vision, and supported this with a computational model. Other studies have also shown distortions of perceived object shape or size, including various size adaptation aftereffects (Altan and Boyaci, 2020; Köhler and Wallach, 1944; Kreutzer et al., 2015; Laycock et al., 2017; Pooresmaeili et al., 2013; Zeng et al., 2017; Zimmermann et al., 2016). However, none of these studies have demonstrated that vision involves neural mechanisms for size coding. For example, where static luminance-defined shapes or contours have been used, perceptual distortions can be attributed to spatial repulsion effects (Ganz, 1964) or to adaptation of low spatial frequency mechanisms (Blakemore and Sutton, 1969). In some cases, the involvement of retinal afterimages might also have been important (see Köhler and Wallach, 1944). In contrast, our own experiments are the first ones designed to investigate the size-mechanism hypothesis directly, ruling out the possibilities above by using (i) a rapid spatially jittering adaptor and (ii) stimuli that are narrowband in spatial frequency. This approach decouples object size from (i) retinal contour location and (ii) carrier (or dominant object) spatial frequency. We have also placed our results in a computational context developed here by building in a natural way on known cortical physiology and a large body of previous psychophysical work on early vision. We know of no other image-driven model of size perception that has done this. Our model accommodated our hypothesised bi-directional adaptation aftereffects for size, and in this respect the work here delivers a successful test of our proposal about size-coding that emerged from work on image contrast (Meese and Baker, 2011). But more than that, two asymmetrical features of our results that we had not anticipated are emergent properties of our model and provide further support for the scheme we have been advocating.

Our work is complemented by several other studies and observations. For example, the perceived aspect ratio of a shape can be distorted by adaptation, appearing narrower or wider when observers were adapted to other shapes (Storrs and Arnold, 2017) or to large grating patterns (Frome et al., 1979). Haptic size adaptation has also been demonstrated using paradigms where participants grasped objects of different sizes (Walker and Shea, 1974). And more generally, strong simultaneous effects on perceived size can also be induced by surrounding elements that are either smaller or larger than a central target, such as in the well-known Ebbinghaus illusion (Ebbinghaus, 1902).

Our focus here has not been on temporal dynamics, but in additional pilot experiments conducted on the second author, we found that the size effect increased monotonically during the first 16 seconds of adaptation but was constant thereafter. Size perception returned to baseline by around 60 seconds after the offset of the adaptor. This rapid attack and long persistence mean that judgements of object size in natural environments might be affected in repetitive tasks such as fruit picking or production line work, for example. There might also be important clinical implications relating to judgements of body size in patients with eating disorders (Challinor et al., 2017).

Our motivation for this work was the prediction that populations of neurons which pool image texture across various regions of the visual field might code for object size (Meese and Baker, 2011). Receptive fields in V1 are too small for this, but neurons with large receptive fields that also exhibit suppression effects are found in extra-striate visual areas such as V4 (Desimone and Schein, 1987; Pollen et al., 2002) and have many of the required properties. Results from fMRI have shown that adaptation can reduce or increase the area of V1 activated by a stimulus (Pooresmaeili et al., 2013) and this might suggest that our own adaptation effects occur at an earlier stage than we suppose. However, the slow time-course of fMRI means that feedback from later stages might also be involved in these imaging results. A recent study applying TMS to lateral occipital cortex (Zeng et al., 2020) found that size judgements were disrupted at an earlier time point than when TMS was applied to early visual cortex, consistent with this feedback hypothesis.

Our modelling shows how mutually inhibitory mechanisms that pool over different regions of the visual field can produce an adaptable population code for size, but several developments are needed before a more comprehensive understanding of perceived object size can emerge. For simplicity, our current model is one dimensional and (quite straightforwardly) would need to be extended to the two spatial dimensions of the retina to have general applicability. More challenging is that our current model has no method for segmenting the background to deliver a stopping rule for contrast integration, this shortfall owing to the simplicity of our psychophysical constraints. However, our work is not undermined by this, and remains valid so long as the segmentation task can be achieved. Everyday observations serve as an existence proof for this, and our finding that size adaptation aftereffects extend across the orientation and spatial frequency of luminance modulations points to the sort of general second-order process that we should expect (see also Richard et al., 2019): one that operates on the envelope of pooled local contrasts (e.g. the boundary of an object). Furthermore, visual neurons in V2 and V4 (Zhou et al., 2000) are known to have the border ownership properties that might be an important part of this process, the details of which continue to be investigated (Heydt and Zhang, 2018). Finally, our experiments did not distinguish between image size (which depends on viewing distance) and physical object size (which does not), so whether the process identified here comes before or after the processes of depth perception and size constancy remains to be elucidated. We hope that future work in neuroimaging and neurophysiology as well as psychophysics, will help to illuminate these issues.

Appendix A: A computational model of size adaptation

We constructed a computational model to simulate the size adaptation experiments. For simplicity, the model operates along only one spatial dimension with $i = 1$ to n pixels ($n = 8,192$). This gives a notional resolution of 256 pixels per degree. This is higher than in the experiments (64 pixels per degree) which were limited by the pixel resolution of the display. The model has a population of $j = 1$ to m pooling mechanisms ($m = 91$) defined by Gaussian spatial profiles (G_{ij}) with standard deviations ranging from 4 to 2048 pixels (or ~1 arcmin to 8 degrees) in logarithmic steps of the index j. Model inputs were rectangular functions (see Figure 5a) with unity height and widths between 16 and 4,096 pixels (3.75 arcmin to 16 degrees; 17 sizes in log steps). Half a cycle of a raised cosine function (16 pixels, or 3.75 arcmin wide, consistent with the stimuli in our experiments) was added to each end of these functions to produce profiles with smooth edges. These inputs describe the contrast envelope of our stimuli across space which, to a first approximation, is equivalent to the output of an array of standard model complex cells. (Recall that stimulus and adaptor contrasts were fixed and identical in our experiments.) Stimulus response was calculated by multiplying model inputs by each pooling mechanism, and their products were passed through a nonlinear gain control equation (Layer 1), given by:

$$L1_j = \frac{\Sigma_{i=1:n}(G_{ij} \times C_i)^p}{Z_j + \Sigma_{i=1:n}(G_{ij} \times C_i)^q + \Sigma_{i=1:n}(G_{im} \times C_i)^q} \tag{1}$$

where C_i is the model input (the contrast envelope) at each pixel location (i). Note that the contrast terms are exponentiated before summation, as required by previous work (Baker et al., 2013). The exponents p and q were fixed at 2.4 and 2 respectively, also based on previous work (Legge and Foley, 1980; Meese and Summers, 2012; Meese and Summers, 2007). Z_j is a saturation constant with a value of 1 in the absence of adaptation (described further below). The term that includes G_{im} represents surround suppression from the largest mechanism in the population (indexed by m). The Σ symbol denotes summation across space ($i = 1{:}n$). All stimuli and pooling mechanisms were centred in the middle of the spatial array (of size n).

There is a special case of our gain control model worthy of consideration. This is when the target stimulus is a uniform contrast signal, as in our experiments here. In this case, and by ignoring the minor complication of the small skirts of our stimuli, Equation 1 simplifies to the following:

$$\widehat{L1_j} \approx \frac{A_j C^p}{Z_j + B_j C^q + B_m C^q} \tag{2}$$

where,

$$A_j = S_j \Sigma_{i=1:n} G_{ij}^p \tag{3}$$

and

$$B_j = S_j \Sigma_{i=1:n} G_{ij}^q \tag{4}$$

and where S_j is either the size of the j-th mechanism ($j = 1{:}m$) or the size of the stimulus (both in units of i), whichever is smaller. Equation 2 is used in Figure 5a to facilitate interpretation of our Layer 1 architecture, but our formal modelling used the more general Equation 1.

Layer 2 of the model involves subtractive inhibition between neighbouring mechanisms:

$$L2_j = L1_j - L1_{j-1}. \tag{5}$$

The model was presented with stimuli of various sizes, and the size of each stimulus was estimated by fitting a spline to the population response and finding the maximum (Figure 5b). We obtained equivalent results using a maximum likelihood readout rule (Jazayeri and Movshon, 2006), but this did not improve the analysis and was omitted for simplicity. To model the influence of adaptation, we set the saturation constant such that $Z_j = 1 + \alpha R_j$, where α is a free parameter that determines the weight of adaptation, consistent with previous work (Foley and Chen, 1997; Meese and Holmes, 2002), and R_j is the normalized mechanism response of the j-th mechanism to the adaptor. We scaled the values of vector R between 0 and 1 by dividing by the most responsive mechanism. (Note that $R = 0$ in the absence of adaptation). We set $\alpha = 360$ by eye (for all mechanisms) to give an adaptation effect consistent with our empirical results. (Note that α is the only free parameter in the model.) Model estimates of stimulus size were then recalculated using the adapted values of Z_j. We define the adaptation aftereffect as the ratio of perceived sizes with and without adaptation, expressed as percentage change (see Figure 5d). We found that key model behaviours derived from the model architecture and were not critically dependent on precise parameter values. We also anticipate that a more elaborate two-dimensional version of the model with traditional bandpass front-end filters would behave similarly for the experiments here.

Acknowledgments

References

Altan E, Boyaci H. 2020. Size aftereffect is non-local. *Vision Research* **176**:40–47. doi:10.1016/j.visres.2020.07.006

Arnold KD, Winer GA, Wickens DD. 1982. Veridical and nonveridical interpretations to perceived temperature differences by children and adults. *Bulletin of the Psychonomic Society* **20**:237–238. doi:10.3758/BF03334826

Baker DH, Meese TS. 2012. Size adaptation effects are independent of spatial frequency aftereffects. *Perception* **41(S)**:33.

Baker DH, Meese TS, Georgeson MA. 2013. Paradoxical psychometric functions ("swan functions") are explained by dilution masking in four stimulus dimensions. *i-Perception* **4**:17–35. doi:10.1068/i0552

Baldwin AS, Meese TS. 2015. Fourth-root summation of contrast over area: No end in sight when spatially inhomogeneous sensitivity is compensated by a witch's hat. *Journal of Vision* **15**:4. doi:10.1167/15.15.4

Barlow HB. 1953. Summation and inhibition in the frog's retina. *The Journal of Physiology* **119**:69–88. doi:10.1113/jphysiol.1953.sp004829

Blakemore C, Julesz B. 1971. Stereoscopic depth aftereffect produced without monocular cues. *Science* **171**:286–288. doi:10.1126/science.171.3968.286

Blakemore C, Nachmias J. 1971. The orientation specificity of two visual after-effects. *The Journal of Physiology* **213**:157–174. doi:10.1113/jphysiol.1971.sp009374

Blakemore C, Nachmias J, Sutton P. 1970. The perceived spatial frequency shift: Evidence for frequency-selective neurones in the human brain. *The Journal of Physiology* **210**:727–750. doi:10.1113/jphysiol.1970.sp009238

Blakemore C, Sutton P. 1969. Size adaptation: A new aftereffect. *Science* **166**:245–247. doi:10.1126/science.166.3902.245

Cannon MW, Fullenkamp SC. 1988. Perceived contrast and stimulus size: Experiment and simulation. *Vision Research* **28**:695–709. doi:10.1016/0042-6989(88)90049-1

Cavanaugh JR, Bair W, Movshon JA. 2002. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology* **88**:2530–2546. doi:10.1152/jn.00692.2001

Challinor KL, Mond J, Stephen ID, Mitchison D, Stevenson RJ, Hay P, Brooks KR. 2017. Body size and shape misperception and visual adaptation: An overview of an emerging research paradigm. *Journal of International Medical Research* **45**:2001–2008. doi:10.1177/0300060517726440

Christman RJ. 1954. Shifts in pitch as a function of prolonged stimulation with pure tones. *The American Journal of Psychology* **67**:484. doi:10.2307/1417939

Coltheart M. 1971. Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review* **78**:114–121. doi:10.1037/h0030639

Desimone R, Schein SJ. 1987. Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology* **57**:835–868. doi:10.1152/jn.1987.57.3.835

Ebbinghaus H. 1902. Grundzuge der psychologie. Leipzig: Verlag von Viet & Co.

Foley JM. 1994. Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, Optics, Image Science, and Vision* **11**:1710–1719. doi:10.1364/josaa.11.001710

Foley JM, Chen CC. 1997. Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model. *Vision Research* **37**:2779–2788. doi:10.1016/s0042-6989(97)00081-3

Frome F, Levinson J, Danielson J, Clavadetscher J. 1979. Shifts in perception of size after adaptation to gratings. *Science* **206**:1327–1329. doi:10.1126/science.515736

Ganz L. 1964. Lateral inhibition and the location of visual contours. An analysis of figural after-effects. *Vision Research* **4**:465–481. doi:10.1016/0042-6989(64)90053-7

Georgeson MA. 1980. Spatial frequency analysis in early visual processing. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **290**:11–22. doi:10.1098/rstb.1980.0079

Gibson JJ, Radner M. 1937. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology* **20**:453–467. doi:10.1037/h0059826

Harris JP, Wink B. 2000. Invariance of the perceived spatial frequency shift of peripherally viewed gratings with manipulations of contrast, duration, and luminance. *Vision Research* **40**:931–941. doi:10.1016/S0042-6989(99)00236-9

Heeley DW. 1979. A perceived spatial frequency shift at orientations orthogonal to adapting gratings. *Vision Research* **19**:1229–1236. doi:10.1016/0042-6989(79)90188-3

Heron J, Aaen-Stockdale C, Hotchkiss J, Roach NW, McGraw PV, Whitaker D. 2012. Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences* **279**:690–698. doi:10.1098/rspb.2011.1131

Heydt R von der, Zhang NR. 2018. Figure and ground: How the visual cortex integrates local cues for global organization. *Journal of Neurophysiology* **120**:3085–3098. doi:10.1152/jn.00125.2018

Jazayeri M, Movshon JA. 2006. Optimal representation of sensory information by neural populations. *Nature Neuroscience* **9**:690–696. doi:10.1038/nn1691

Klein S, Stromeyer CF, Ganz L. 1974. The simultaneous spatial frequency shift: A dissociation between the detection and perception of gratings. *Vision Research* **14**:1421–1432. doi:10.1016/0042-6989(74)90017-0

Köhler W, Wallach H. 1944. Figural after-effects. An investigation of visual processes. *Proceedings of the American Philosophical Society* **88**:269–357.

Kreutzer S, Fink GR, Weidner R. 2015. Attention modulates visual size adaptation. *Journal of Vision* **15**:10. doi:10.1167/15.15.10

Lawless HT. 1991. A sequential contrast effect in odor perception. *Bulletin of the Psychonomic Society* **29**:317–319. doi:10.3758/BF03333930

Laycock R, Sherman JA, Sperandio I, Chouinard PA. 2017. Size aftereffects are eliminated when adaptor stimuli are prevented from reaching awareness by continuous flash suppression. *Frontiers in Human Neuroscience* **11**:479. doi:10.3389/fnhum.2017.00479

Legge GE, Foley JM. 1980. Contrast masking in human vision. *Journal of the Optical Society of America* **70**:1458–1471.

Marks LE. 1994. "Recalibrating" the auditory system: The perception of loudness. *J Exp Psychol Hum Percept Perform* **20**:382–96. doi:10.1037//0096-1523.20.2.382

Marr D, Hildreth E. 1980. Theory of edge detection. *Proceedings of the Royal Society of London Series B, Biological Sciences* **207**:187–217. doi:10.1098/rspb.1980.0020

Mather G, Pavan A, Campana G, Casco C. 2008. The motion aftereffect reloaded. *Trends in Cognitive Sciences* **12**:481–487. doi:10.1016/j.tics.2008.09.002

Meese TS. 2010. Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision* **10**:14. doi:10.1167/10.8.14

Meese TS. 2004. Area summation and masking. *Journal of Vision* **4**:8. doi:10.1167/4.10.8

Meese TS, Baker DH. 2013. A common rule for integration and suppression of luminance contrast across eyes, space, time, and pattern. *i-Perception* **4**:1–16. doi:10.1068/i0556

Meese TS, Baker DH. 2011. Contrast summation across eyes and space is revealed along the entire dipper function by a "swiss cheese" stimulus. *Journal of Vision* **11**:23–23. doi:10.1167/11.1.23

Meese TS, Baker DH, Summers RJ. 2017. Perception of global image contrast involves transparent spatial filtering and the integration and suppression of local contrasts (not RMS contrast). *Royal Society Open Science* **4**:170285. doi:10.1098/rsos.170285

Meese TS, Hess RF, Williams CB. 2005. Size matters, but not for everyone: Individual differences for contrast discrimination. *Journal of Vision* **5**:928–947. doi:10.1167/5.11.2

Meese TS, Holmes DJ. 2002. Adaptation and gain pool summation: Alternative models and masking data. *Vision Research* **42**:1113–1125. doi:10.1016/S0042-6989(01)00291-7

Meese TS, Summers RJ. 2012. Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. *Journal of Vision* **12**:9–9. doi:10.1167/12.11.9

Meese TS, Summers RJ. 2009. Neuronal convergence in early contrast vision: Binocular summation is followed by response nonlinearity and area summation. *Journal of Vision* **9**:7–7. doi:10.1167/9.4.7

Meese TS, Summers RJ. 2007. Area summation in human vision at and above detection threshold. *Proceedings of the Royal Society B: Biological Sciences* **274**:2891–2900. doi:10.1098/rspb.2007.0957

Morgan MJ, Ward RM, Hole GJ. 1990. Evidence for positional coding in hyperacuity. *Journal of the Optical Society of America A* **7**:297. doi:10.1364/JOSAA.7.000297

Moutsiana C, Haas B de, Papageorgiou A, Dijk JA van, Balraj A, Greenwood JA, Schwarzkopf DS.

2016. Cortical idiosyncrasies predict the perception of object size. *Nature Communications* **7**:12110. doi:10.1038/ncomms12110

Olejnik S, Algina J. 2003. Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* **8**:434–447. doi:10.1037/1082-989X.8.4.434

Pollen DA, Przybyszewski AW, Rubin MA, Foote W. 2002. Spatial receptive field organization of macaque V4 neurons. *Cerebral Cortex* **12**:601–616. doi:10.1093/cercor/12.6.601

Pooresmaeili A, Arrighi R, Biagi L, Morrone MC. 2013. Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *Journal of Neuroscience* **33**:15999–16008. doi:10.1523/JNEUROSCI.1770-13.2013

Richard B, Hansen BC, Johnson AP, Shafto P. 2019. Spatial summation of broadband contrast. *Journal of Vision* **19**:16. doi:10.1167/19.5.16

Roach NW, Heron J, Whitaker D, McGraw PV. 2011. Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B: Biological Sciences* **278**:1314–1322. doi:10.1098/rspb.2010.1737

Rouder JN, Morey RD, Speckman PL, Province JM. 2012. Default bayes factors for ANOVA designs. *Journal of Mathematical Psychology* **56**:356–374. doi:10.1016/j.jmp.2012.08.001

Sengpiel F, Baddeley RJ, Freeman TC, Harrad R, Blakemore C. 1998. Different mechanisms underlie three inhibitory phenomena in cat area 17. *Vision Research* **38**:2067–2080. doi:10.1016/s0042-6989(97)00413-6

Snowden RJ, Hammett ST. 1998. The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Research* **38**:1935–1945. doi:10.1016/s0042-6989(97)00379-9

Storrs KR, Arnold DH. 2017. Shape adaptation exaggerates shape differences. *Journal of Experimental Psychology: Human Perception and Performance* **43**:181–191. doi:10.1037/xhp0000292

Sutherland NS. 1961. Figural after-effects and apparent size. *Quarterly Journal of Experimental Psychology* **13**:222–228. doi:10.1080/17470216108416498

Tsai JJ, Victor JD. 2003. Reading a population code: A multi-scale neural model for representing binocular disparity. *Vision Research* **43**:445–466. doi:10.1016/s0042-6989(02)00510-2

Walker JT, Irion AL, Gordon DG. 1981. Simple and contingent aftereffects of perceived duration in vision and audition. *Perception & Psychophysics* **29**:475–486. doi:10.3758/BF03207361

Walker JT, Shea KS. 1974. A tactual size aftereffect contingent on hand position. *Journal of Experimental Psychology* **103**:668–674. doi:10.1037/h0037136

Webb BS, Tinsley CJ, Barraclough NE, Parker A, Derrington AM. 2003. Gain control from beyond the classical receptive field in primate primary visual cortex. *Visual Neuroscience* **20**:221–230. doi:10.1017/s0952523803203011

Webster MA, Leonard D. 2008. Adaptation and perceptual norms in color vision. *Journal of the Optical Society of America A* **25**:2817. doi:10.1364/JOSAA.25.002817

Xing J, Heeger DJ. 2001. Measurement and modeling of center-surround suppression and enhancement. *Vision Research* **41**:571–583. doi:10.1016/s0042-6989(00)00270-4

Zeng H, Fink GR, Weidner R. 2020. Visual size processing in early visual cortex follows lateral occipital cortex involvement. *The Journal of Neuroscience* **40**:4410–4417. doi:10.1523/JNEUROSCI.2437-19.2020

Zeng H, Kreutzer S, Fink GR, Weidner R. 2017. The source of visual size adaptation. *Journal of Vision* **17**:8. doi:10.1167/17.14.8

Zhou H, Friedman HS, Heydt R von der. 2000. Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* **20**:6594–6611. doi:10.1523/JNEUROSCI.20-17-06594.2000

Zimmermann E, Morrone MC, Burr D. 2016. Adaptation to size affects saccades with long but not short latencies. *Journal of Vision* **16**:2. doi:10.1167/16.7.2
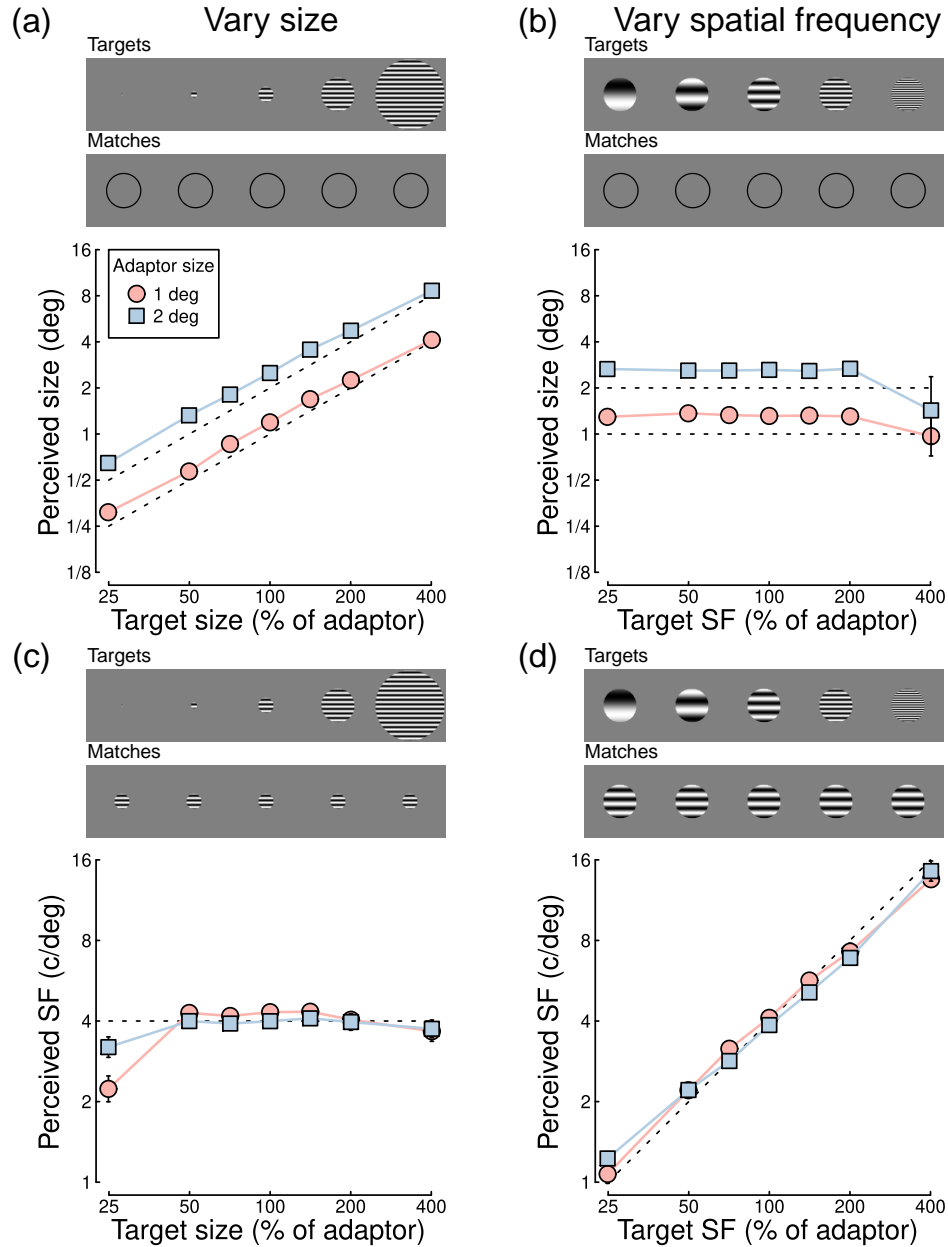
Supplementary information

Figure S1: Baseline settings for perceived size (top) and perceived spatial frequency (bottom) (N = 8 participants). The stimuli were targets that varied in size (left) or in spatial frequency (right). Dashed lines in each panel show the expectation for veridical matches. In general, perceived size was slightly overestimated, most likely a consequence of the raised cosine envelope used to window the grating stimuli which added to the overall diameter. The high variance in the perceived sizes of the highest spatial frequency stimuli (far right points in panel b) is probably because these stimuli were close to detection threshold. Similarly, the spatial frequency was difficult to judge for the smallest targets (far left points in panel c) and was underestimated overall. However, the matches for the baseline conditions from the main experiments were very close to veridical (panels a, d). Error bars indicate ±1SE across participants (N=8), and in most cases are smaller than the symbols.
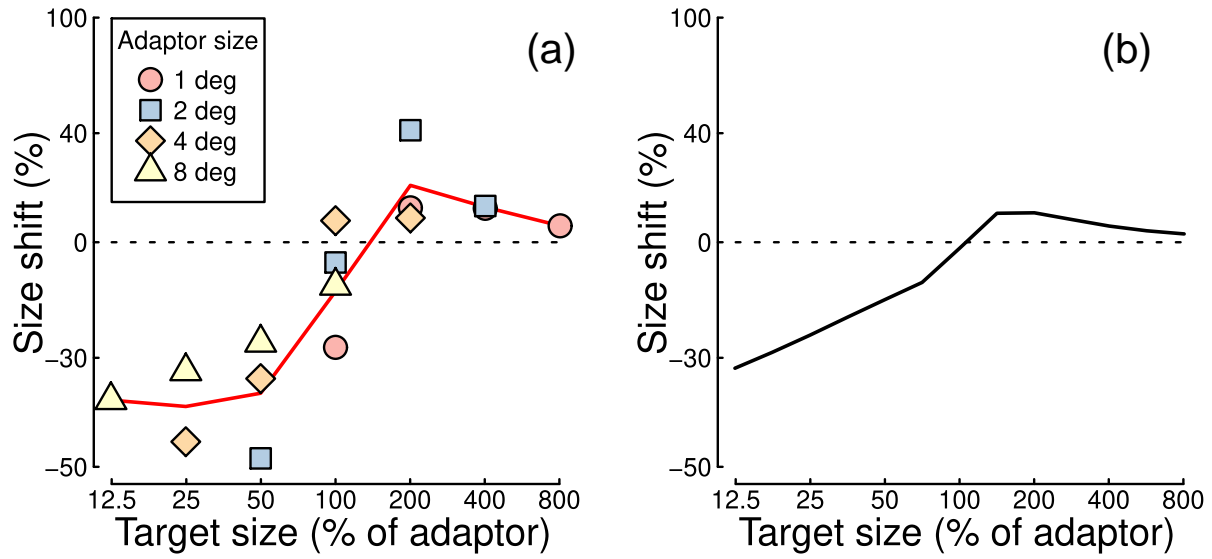
Figure S2: Results for the second author (a) and model predictions (b) across a wider range of relative target/adaptor sizes than was used in the main experiment (N = 1 participant). The conditions involved factorial combination of four target and adaptor sizes (see legend in panel (a)), and the red curve shows their average. Across this extended range the results confirm our earlier observations: the size reduction aftereffect (left of the plot) does not return to baseline, whereas the size increase aftereffect (right of the plot) returns to baseline for large targets. As in the main experiment, the size reduction aftereffects are much larger than the size increase aftereffects. The model behaviour in panel (b) predicts these features of the results. The adaptation parameter, α, is unchanged from before.
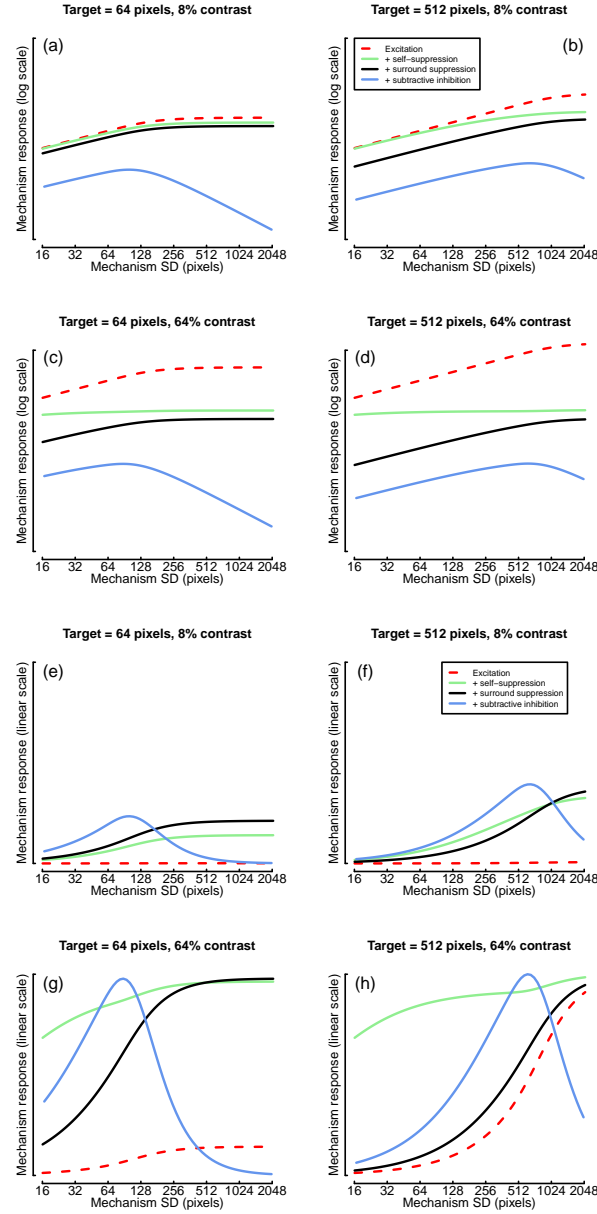
Figure S3: Mechanism responses as functions of the size (SD) of the integration window. This illustrates population responses with the progressive introduction of various model features (see legend). Panels (a-d) are for double-log axes. The same mechanism responses are shown on linear axes in panels (e-h). The responses are shown for a small (64 pixel wide) stimulus in the left column (a, c, e, g) and a larger (512 pixel wide) stimulus in the right column (b, d, f, h). The first (a, b) and third (e, f) rows are for low (8%) contrast stimuli, and the second (c, d) and fourth (g, h) rows are for high (64%) contrast stimuli. For the purpose of illustration in panels (e-h), each function is normalised by the greatest response across panels. The dashed red curves (Layer 1) are for a model with no suppression (no gain control), only excitation and exponentiation ($p = 2.4$) of the contrast responses for a population of different sized spatial integration windows. The right-hand side of the red response function rises with an increase in stimulus contrast and an increase in stimulus size, resulting in response ambiguity. This ambiguity can be removed by implementing gain control by self-suppression (green curves) in Layer 1. This leaves a contrast effect in the model (first row versus second row), but no clear effect of size (e.g. panel c versus d). This problem is fixed by introducing surround suppression (black curves) in Layer 1. With this arrangement, the height of the population response depends on stimulus contrast (consistent with published contrast discrimination results) and the lateral position of the transition of the population response depends on stimulus size. The blue curves derive from the introduction of lateral inhibition in Layer 2. This converts the maximum gradients in the black functions of Layer 1 to peaks in Layer 2 by a discrete approximation to differentiation. The peaks of these curves are located by a max operator.
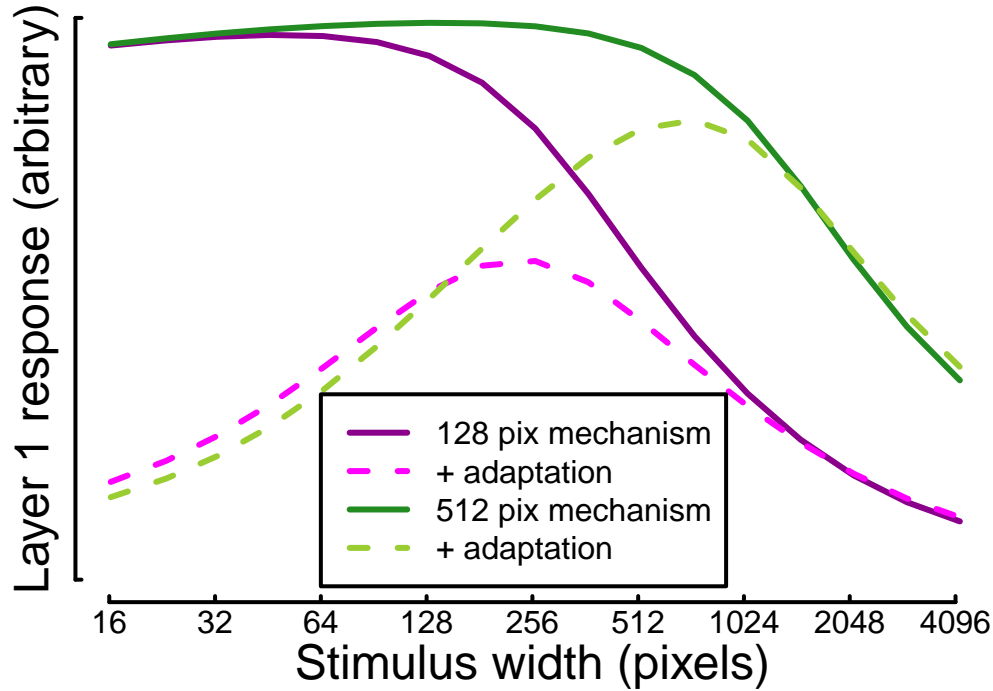
Figure S4: Responses of two Layer 1 mechanisms in the model as a function of the width/size of a high contrast stimulus. The standard deviations of the integration windows were 128 pixels (pink/purple) and 512 pixels (green). Solid and dashed curves are for the unadapted and adapted states, respectively (for a 256-pixel adaptor). Note that when the stimulus is much smaller than the integration window, increasing stimulus size causes little or no change in response because the benefits of excitatory integration are offset by the effects of suppression in the contrast gain control equation (Equations 1 and 2, and Figure 5a of the main report). However, for stimulus sizes greater than the integration window, responses decrease with stimulus size because of the increasing contribution of surround suppression (the second contrast term on the denominator of Equations 1 and 2, and the Bm term in Figure 5 of the main report) without a concomitant increase in excitation. This arrangement means that the Layer 1 mechanisms can carry a stimulus contrast code unconfound by stimulus size (Meese, 2004; Meese and Summers, 2007; Meese and Baker, 2011, 2013). The situation is different in the adapted state. Here, adaptation is most severe when the test stimulus is smaller than the mechanism's integration window. This is because as the stimulus gets smaller, the A and B terms in Figure 5 of the main report get smaller, and the mechanism is more readily influenced by adaptation which increases the value of Z (see also Equation 1 in the main report). This asymmetry in adaptation contributes to the asymmetries seen in the model's account of our results.
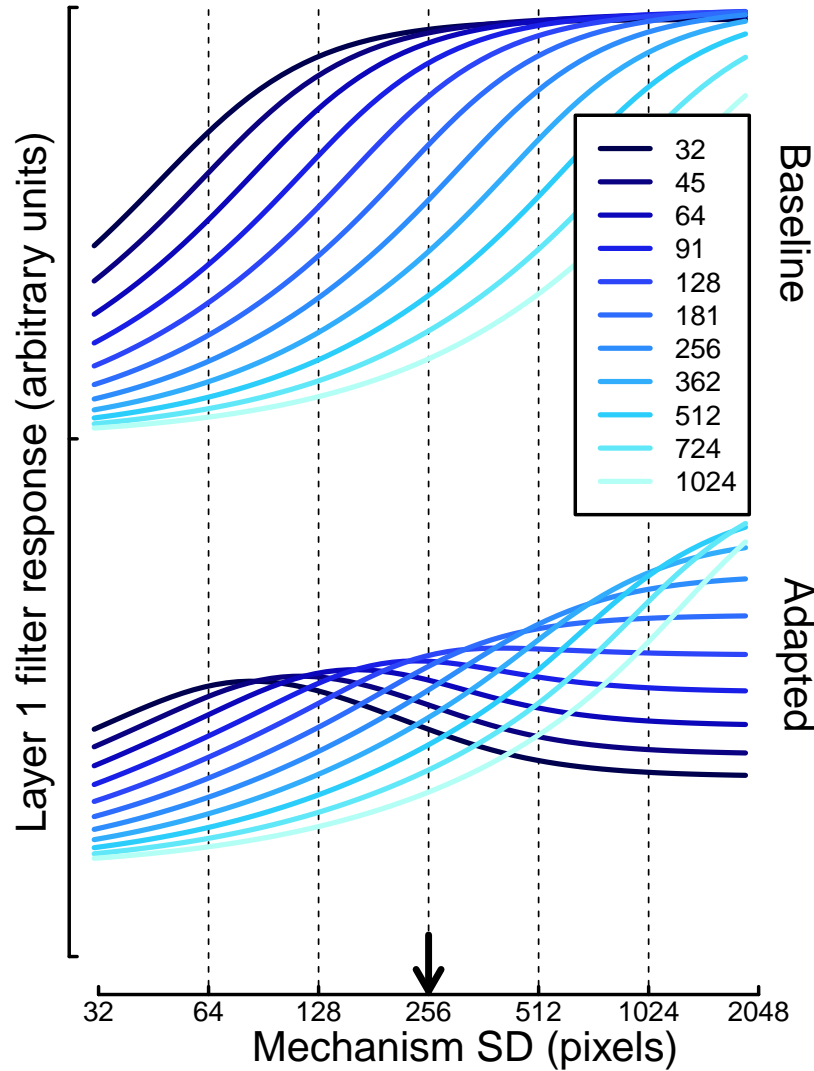
Figure S5: Responses of Layer 1 mechanisms for different high contrast stimulus width/sizes (shown in the legend in pixel units) as a function of the standard deviation of each mechanism's integration window. The top panel is for the unadapted/baseline state and the bottom panel is for adaptation to a stimulus with a width/size of 256 pixels (black arrow). This figure is the Layer 1 equivalent of the Layer 2 responses shown in Figure 5b of the main report but extended to smaller stimulus sizes. (The top panel is also the family of curves for which a single instance is shown in Figure 5c of the main report.) At baseline, the population of Layer 1 responses provides a code for image contrast (plateau height) and stimulus size (given by the mechanism's size label at the transition point in the response distribution and picked off by the differentiation process in Layer 2). The aftereffects of adaptation on the population of Layer 1 responses is asymmetric. This is because the impact of adaptation depends on (i) the mechanism's sensitivity to the adaptor (top panel), which is asymmetric to the right, but also (ii) the effect of the adaptation term (Z in Figure 5a and Equation 1 of the main report) on the overall mechanism response. This effect increases with a decrease in the integration of signal contrast and is asymmetric to the left. These two asymmetries are not balanced in their influences on the adapted Layer 1 network which is shown by the asymmetry in the bottom panel of the figure. This imbalance contributes to the asymmetries seen in the model's account of our results.

Supplementary Movies

Movie S1. Demonstration of size adaptation with horizontal adaptors and targets. To facilitate a large effect, the viewer should fixate the central cross throughout. The adaptors will jitter on either size of the display. After a brief blank interval, two identical gratings are shown, but they will look different in size following adaptation. Note that the demonstration does not correspond to the stimulus arrangement in the experiment, where a single adaptor was presented to one side only of the display.

Movie S2. As for Movie S1, except that the adaptors have the same size and different spatial frequencies. The aftereffect is a change in the perceived spatial frequency of the carrier but with no change in perceived size.

Movie S3. As for Movie S1, except that the adaptors and targets have orthogonal orientations. The size-aftereffect is seen nonetheless.

Movie S4. As for Movie S1, except that the grating textures are replaced by uniform luminance.

Movie S5. As for Movie S1, except that the target grating is replaced by a face image.