

Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention

Marc S. Tibber

Institute of Ophthalmology, University College London,
Bath Street, London, UK



John A. Greenwood

Institute of Ophthalmology, University College London,
Bath Street, London, UK; Laboratoire Psychologie de la
Perception, Université Paris Descartes, Paris, France;
CNRS UMR 8158, Paris, France



Steven C. Dakin

Institute of Ophthalmology, University College London,
Bath Street, London, UK; NIHR Biomedical Research
Centre for Ophthalmology at Moorfields Eye Hospital
NHS Foundation Trust, London, UK



While observers are adept at judging the density of elements (e.g., in a random-dot image), it has recently been proposed that they also have an independent visual sense of number. To test the independence of number and density discrimination, we examined the effects of manipulating stimulus structure (patch size, element size, contrast, and contrast-polarity) and available attentional resources on both judgments. Five observers made a series of two-alternative, forced-choice discriminations based on the relative numerosity/density of two simultaneously presented patches containing 16–1,024 Gaussian blobs. Mismatches of patch size and element size (across reference and test) led to bias and reduced sensitivity in both tasks, whereas manipulations of contrast and contrast-polarity had varied effects on observers, implying differing strategies. Nonetheless, the effects reported were consistent across density and number judgments, the only exception being when luminance cues were made available. Finally, density and number judgment were similarly impaired by attentional load in a dual-task experiment. These results are consistent with a common underlying metric to density and number judgments, with the caveat that additional cues may be exploited when they are available.

Keywords: number, density, perception, psychophysics, spatial vision

Citation: Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision*, 12(6):8, 1–19, <http://www.journalofvision.org/content/12/6/8>, doi:10.1167/12.6.8

Introduction

Humans possess a formidable “number sense,” allowing them to make number estimates across widely varying conditions (for review, see Dehaene, 1992). For example, following habituation, newborn babies and young infants fixate longer—a measure of attention, and by inference, perceived novelty—when the numerosity of an array of objects is manipulated independently of their arrangement, shape, or identity (Antell & Keating, 1983; Jordan & Brannon, 2006; Starkey, Spelke, & Gelman, 1990; Xu & Arriaga, 2007). In the adult, this ability extends to judgments under conditions in which counting is not possible: i.e., large numerosities and short presentation times (Allik &

Tuulmets, 1991; Allik, Tuulmets, & Vos, 1991; Vos, van Oeffelen, Tibosch, & Allik, 1988). Human observers are also remarkably adept at judging the *density* of objects within a given area (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Durgin, 1995; Durgin & Huk, 1997). However, number and density are tightly linked, both conceptually (Density = Number/Area) and behaviorally. Weber fractions for number and density discrimination thresholds are frequently indistinguishable (Ross & Burr, 2010), and both dimensions are prone to adaptation following prolonged viewing (Durgin, 1995).

Given this close association between number and density, does the visual system need independent representations of both, or might they be derived from a common mechanism? This is a contentious issue

fuelled by a recent study in which Burr and colleagues (Burr & Ross, 2008b) showed that adaptation to a high numerosity random-dot array reduces the perceived numerosity of a subsequently presented patch of test dots. This led the authors to suggest that numerosity is a primary visual attribute (or “distinct qualia,” Burr & Ross, 2008a) that cannot be reduced to other continuous stimulus dimensions, e.g., density or spatial frequency (Ross & Burr, 2010). This claim has not gone uncontested. Durgin (2008) showed that, when number and density are uncoupled by changing the size of the adapter patch (the region over which the elements are distributed), adaptation follows the density of the adapter rather than its numerosity (Durgin, 2008), suggesting that density is in fact the adapted dimension (Durgin, 1995). In addition, although not uncontested themselves (Allik, Tuulmets, & Vos, 1991; Burr & Ross, 2008b; Ross & Burr, 2010), several studies have shown that number judgments are sensitive to manipulations of patch size (Tokita & Ishiguchi, 2010), element size (Ginsburg & Nicholls, 1988; Hurewitz, Gelman, & Schnitzer, 2006; Ross, 2003; Sophian, 2007; Tokita & Ishiguchi, 2010), element clustering (Frith & Frith, 1972; Ginsburg, 1978, 1991), and total element coverage (Hurewitz, Gelman, & Schnitzer, 2006; Tokita & Ishiguchi, 2010). Because a “pure” judgment of number should occur regardless of these spatial parameters, such findings are inconsistent with number being extracted as a primary visual attribute independent of other stimulus dimensions.

An alternative possibility is that number and density judgments tap into a common mechanism. We have recently uncovered a particularly close association between judgments of perceived number and perceived density (Dakin et al., 2011). Mismatches in overall patch size of a test and reference stimulus-pair were found to induce systematic biases (whereby larger patches appear both more numerous and more dense) and reduced sensitivity for both density and number discrimination judgments (see also Tokita & Ishiguchi, 2010). In addition, this study uncovered substantial intraindividual correlations between both bias and threshold measurements associated with number and density discriminations. On the basis of these findings, we developed a model of number and density discrimination that rests on the notion of a common underlying metric derived from the relative output of pairs of spatial frequency (SF)-tuned filters. This measure (referred to as the response-ratio) estimates a crude correlate of density using the ratio of activity between a high SF filter (which roughly captures element number) and a low SF filter (whose response varies in approximate proportion to stimulus area), and can be weighted by an estimate of relative patch size to derive a reliable estimate of numerosity¹. With just one

free parameter for density judgments, and one for number judgments (both noise terms), the model accurately captures observers’ discrimination performance under a range of experimental conditions (Dakin et al., 2011) as well as the effect of element-type and element-connectivity (He, Zhang, Zhou, & Chen, 2009). This model directly contradicts the notion of a dedicated mechanism for visual number extraction and is constrained by a necessary interdependence between visual number and density. Further, it predicts that experimental manipulation of stimulus structure should *similarly* affect number and density judgments, as both are derived from a common metric.

In order to determine the robustness of this association between perceived number and perceived density, and thereby test a basic assumption of any model based on a common metric or processing stage, we manipulated a range of stimulus attributes that might conceivably affect number and density performance—specifically patch size, element size, luminance contrast, contrast-polarity, and available attentional resources—whilst observers performed two-(spatial)-interval, forced-choice (2-IFC) density and number discriminations. Although there is a body of literature on the effects of such manipulations on perceived number (e.g., Ginsburg, 1978; Ginsburg & Nicholls, 1988; Ross, 2003), to the authors’ knowledge only a single study to date has made a direct comparison of their effects on number *and* density judgments using identical methodology and stimuli (Dakin et al., 2011). Further, this only involved manipulation of a single parameter: patch size. On the basis of the response-ratio model of density and number estimation, we predicted that density and number would be similarly affected by all manipulations tested. Contrary to these predictions, if density and number are processed independently by distinct mechanisms, the effects of stimulus manipulation should be uncorrelated. As we shall demonstrate, with the exception of one observers’ performance under conditions that rendered local luminance a useful cue, perceived density and number judgments were similarly affected by all experimental manipulations. Further, density and number thresholds were consistently correlated across all experiments. The most parsimonious interpretation of these data is that number and density are derived from a common metric.

General methods

All observers gave informed written consent in accordance with The Declaration of Helsinki. Each experiment was performed by five observers taken from a pool of eight (all experienced psychophysical observers, five naïve to the purpose of the study). Observers

made judgments about the relative density or numerosity of circular test and reference patches (Figure 1) presented for $250 \text{ ms} \pm 6.25^\circ$ left and right of central-fixation in a series of 2-IFC discriminations. Density and number judgments were performed in separate blocks. Patches were composed of a variable number of small 2D Gaussian blobs (elements) against a background grey display fixed at 70 cd/m^2 . To generate reference and test stimuli, Gaussian blobs (“elements”) were randomly dropped within a defined radius (the “patch”) with overlaps permitted. Where blobs overlapped, Gaussian profiles were added together and clipped to avoid exceeding the available luminance range. In order to decorrelate density and number, test and reference patches were independently varied in size

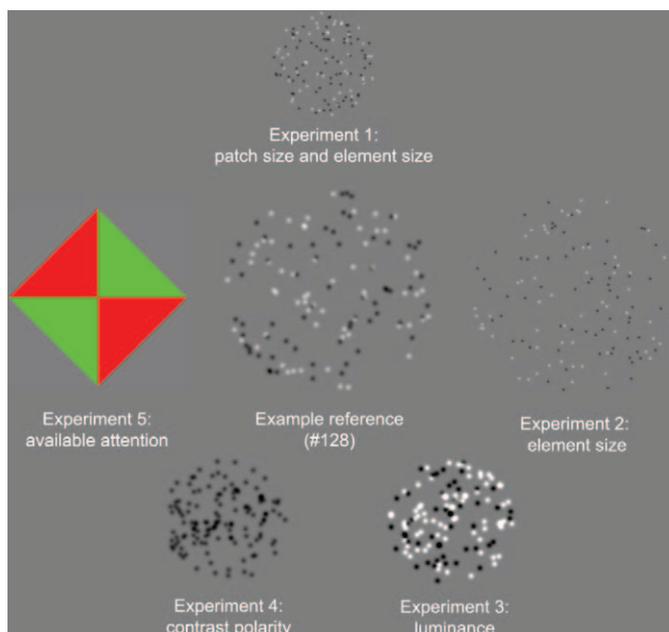


Figure 1. Schematic diagram representing the parameters manipulated and the stimuli used. The central patch shows an example reference (always containing 128 elements). In Experiment 1 (top), patch size and element size were both systematically manipulated. In Experiment 2 (right), element size was manipulated whilst patch size was fixed and matched across patches to examine the effects of element size under conditions of low uncertainty. In Experiments 3 and 4 (bottom), the effects of luminance and contrast-polarity respectively were examined; a smaller patch size mismatch was still introduced, however, in order to decouple number and density and encourage observers to make the appropriate judgment type. Finally, in Experiment 5 (left), available attentional resources were manipulated using a dual-task attentional load paradigm. Whilst performing density and number discriminations, observers had to simultaneously detect targets defined by a unique color (low attentional load) or a conjunction of color and the spatial arrangement of constituent segments (high attentional load). Note: the term “element” refers to an individual Gaussian blob, whilst the term “patch” refers to each collection of elements (i.e., the reference and test patch).

for all experiments except Experiment 2, where patch size was fixed to reduce stimulus uncertainty. The reference always contained 128 elements, whilst the test patch was set using a method of constant stimuli that varied stimulus level (density in the density judgments and number in the number judgments) over a 2-octave range that was split into seven steps. This was centered on 100% (i.e., ± 1 octave) relative to the reference patch, i.e., a physical match, so that for “number” runs, tests contained 64, 81, 102, 128, 162, 203, or 256 elements. For “density” runs, test patches were 50, 63, 79, 100, 126, 159, or 200% of the reference density (2.55 or 10.2 elements/deg² depending on reference size for Experiments 1 and 2, and 3.5 or 7.1 elements/deg² for Experiments 3 to 5). Each run consisted of 112 trials: 16 trials for each of seven stimulus levels, presented according to the method of constant stimuli. Other parameters, e.g., element size, varied from experiment to experiment (see following individual experimental methods sections). Experiments were programmed using Matlab (MathWorks, Cambridge, MA) running on a PC computer with PsychToolbox software (Brainard, 1997; Pelli, 1997). Stimuli were presented on a linearized LCD monitor at a spatial and temporal resolution of 1680×1072 pixels and 60 Hz respectively and viewed binocularly from a distance of 104 cm. Responses were given by keypress.

Nomenclature

Where referenced in the text, patch size and element size are denoted by a “P” and “E” respectively. Upper-case letters indicate the larger sizes (“P” and “E”), and lower-case letters indicate the smaller (“p” and “e”). Details of the reference patch are presented first in any description. Hence, a condition described as P_e/p_E reflects a discrimination between a large reference patch with small elements and a small target patch with large elements.

Analyses

Raw psychophysical data were fit with a two-parameter model using the Palamedes fitting routine (Prins & Kingdom, 2009), which generates parameter estimates of the slope (a measure of sensitivity) and mean (a measure of bias) of the underlying cumulative Gaussian function. Data are presented in raw format (sensitivity and bias) expressed in octaves. For group statistical analyses, a series of repeated measures analyses of variance (ANOVAs) were performed on the slope and absolute (unsigned) biases using SPSS statistical analysis software (version 18.0; SPSS Inc., Chicago, IL). All other statistical tests used are described in the individual experimental sections.

Experiment 1: manipulations of both patch and element size

We first examined the effects of element size and patch size on judgments of number and density. We have previously shown that mismatching relative patch size impairs performance—in terms of increasing threshold and bias—on both of these judgments (Dakin et al., 2011). As previously outlined, a common mechanism for number and density discrimination predicts a convergence of effects of all experimental manipulations. The influence of element size mismatch is of particular interest, however, as the response-ratio model of density and number perception would predict impaired performance when test and reference elements differ in size, if—and only if—the spatial frequency of filters is fixed for the two stimulus intervals, i.e., if no compensation can be made for differences in element size. This could be avoided, however, if the visual system is capable of weighting a response-ratio estimate to compensate for differences in element size, as has been shown for mismatches in patch size (Dakin et al., 2011). The aims of [Experiment 1](#) therefore are two-fold: (1) to test the prediction that density and number judgments are similarly affected by manipulations of patch and element size, and (2) to determine whether the visual system is capable of compensating for relative element size in its estimates of relative number and density.

Methods

Patch size and element size were systematically varied at two levels for both the reference and test patches, creating a 4×4 design. Test and reference patches could have radii of 2° or 4° . The standard deviation of the Gaussian envelope describing test and reference elements could have a value of 2.5 or 5 arcmins. Hence, patches could be matched on both dimensions (PE/PE, Pe/Pe, pE/pE, pe/pe), mismatched on a single dimension (patch size [PE/pE, Pe/pe, pE/PE, pe/Pe] or element size [PE/Pe, Pe/PE, pE/pe, pe/pE]), or mismatched on both (PE/pe, Pe/pE, pE/Pe, pe/PE), generating 16 unique conditions. Elements were of random contrast-polarity, and patches had a peak Michelson contrast of 50%. All observers performed a minimum of two runs per discrimination type (number and density).

Results

For each of the 16 combinations of patch and element size, and each of the two judgments (number or density), psychometric functions were derived for each

observer. These functions plot performance as the frequency that the reference stimulus was seen as either more numerous or more dense than the test stimulus. From this, we derived both estimates of bias (point of subjective equality) and sensitivity (the slope of the function). In [Figure 2](#), biases and sensitivities are shown for density and number judgments when both parameters were matched (red and black symbols) as well as when mismatched for a single parameter (blue and green symbols). In [Figure 3](#), data are plotted for the double mismatched conditions (PE/pe, Pe/pE, pE/Pe, pe/PE). These were presented in a separate figure for clarity. First, considering the effects of mismatching patch size when element size is held constant ([Figure 2a and b](#)), performance is clearly most sensitive and least biased when reference and test patches are matched in size; red and black data points are centered about zero on the abscissa and ordinate values are maximal. When a mismatch in patch size is introduced, however (green and blue data points), sensitivity drops and large biases are introduced; these biases are systematic such that when the reference patch is larger than the test patch (blue data points), more elements must be added to the test for it to appear equally dense or numerous (approximately 240% relative to the reference). That is, large patches look more dense and more numerous than they truly are. Conversely, when the reference patch is smaller than the test patch (green data points), there must be fewer elements in the test for it to be perceived as equally dense or numerous (approximately 50% relative to the reference). This effect, also shown previously (Dakin et al., 2011), captures the fact that the perceived numerosity and density of large patches are overestimated. Note also, that as in the original study, the effect is greater for density than it is for number.

Next, consider the effects of mismatching element size on number and density performance when patch size is held constant ([Figure 2c and d](#)). Once again, judgments are minimally biased and most sensitive when patches are matched (red and black data points). Similar to manipulations of patch size, mismatches of element size induce large biases; green and blue data points fall away from the midline. However, these differ in sign between individuals, with some observers' data being consistent with small elements making a patch appear more dense/numerous, whereas others' are consistent with the reverse. Note, however, that for most observers, the sign of the bias is maintained across judgment types, such that if their green data points (test > reference) are to the left of center in the density plot (negative biases; [Figure 2c](#)), they will also fall to the left in the number plot ([Figure 2d](#)). This was true of four of our five observers on all conditions involving a mismatch of element size.

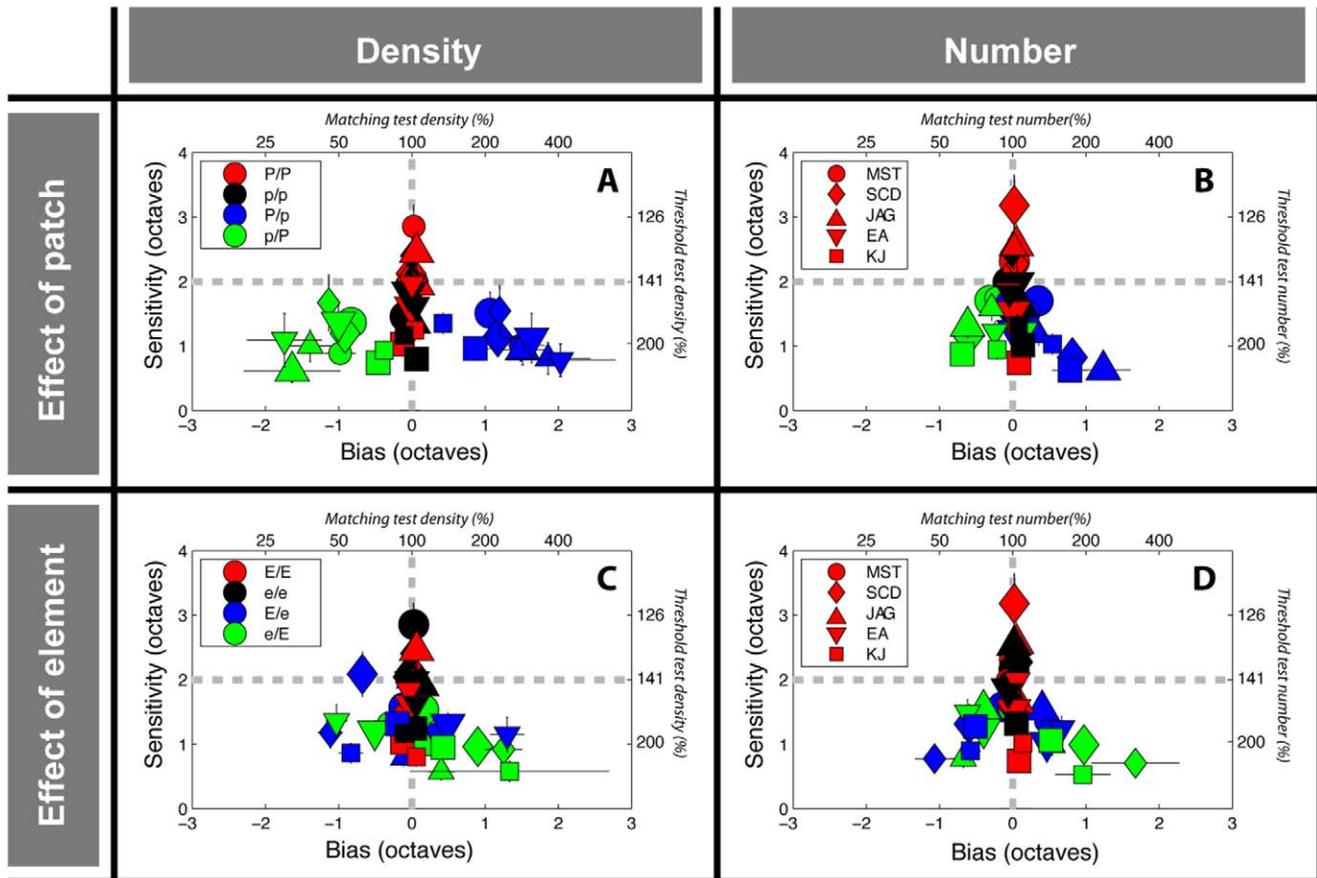


Figure 2. Sensitivity and bias are shown for density judgments (a and c) and number judgments (b and d) under conditions of double parameter matched and single parameter mismatched conditions. A and b plot the effects of patch size when element size is held constant (large symbols = large elements; small symbols = small elements). C and d plot the effects of element size when patch size is held constant (large symbols = large patches; small symbols = small patches). Red and black symbols represent conditions in which the parameter of interest is matched across target and reference; blue and green symbols represent conditions in which the parameter of interest is mismatched. The legend of a and c shows a key to the conditions (“P” = large patch; “p” = small patch; “E” = large elements; “e” = small elements). Details of the reference are always given first. For example, in a, the green symbols (p/P) show data from conditions in which the reference is small and the target is large (i.e., patch size is mismatched). The abscissa shows bias in octaves (lower axis labels) as well as matching test density/number (%) (upper axis labels) for cross-reference. Along the ordinate axes, sensitivity data are plotted in octaves (left axis labels) as well as threshold test number/density (%) (right axis labels) for cross-reference. For example, in a, a bias of +1 octave implies that, for a perceptual match to be made, the test patch must have a density twice that (200%) of the reference. Vertical dotted grey lines denote zero bias. Horizontal dotted grey lines denote average sensitivity as recorded in the original study (Dakin et al., 2011). Error bars plot the standard deviation of fit parameters derived from bootstrapping.

Finally, consider the conditions in which the test and reference are mismatched on both dimensions (element size and patch size; Figure 3). The first thing to note is that relative to the double *matched* conditions (white symbols, presented for comparison), data are once again heavily biased and sensitivity is reduced. Thus, all other (colored) data points fall away from the midline on the abscissa and ordinate values are minimal. Further, fits to the data are comparatively poor with several data points falling out of range and error bars dominating the plot. This implies that, for both density and number tasks, several observers had difficulty making judgments on trials in which relative patch size

and element size were simultaneously mismatched. For data points that do show a reasonable fit, however (i.e., those with small error bars), it is clear that the effects of size-mismatching are dominated by relative patch size rather than relative element size. This is true of both density and number judgments. Thus, there is a general trend for purple and yellow data points (which are the conditions in which the reference patch is smaller than the test patch, irrespective of element size) to fall to the left of center on the abscissa: a negative bias, implying that at the perceived match point there were insufficient elements in the test. Similarly, there is a trend for black and orange symbols (which are conditions in which the

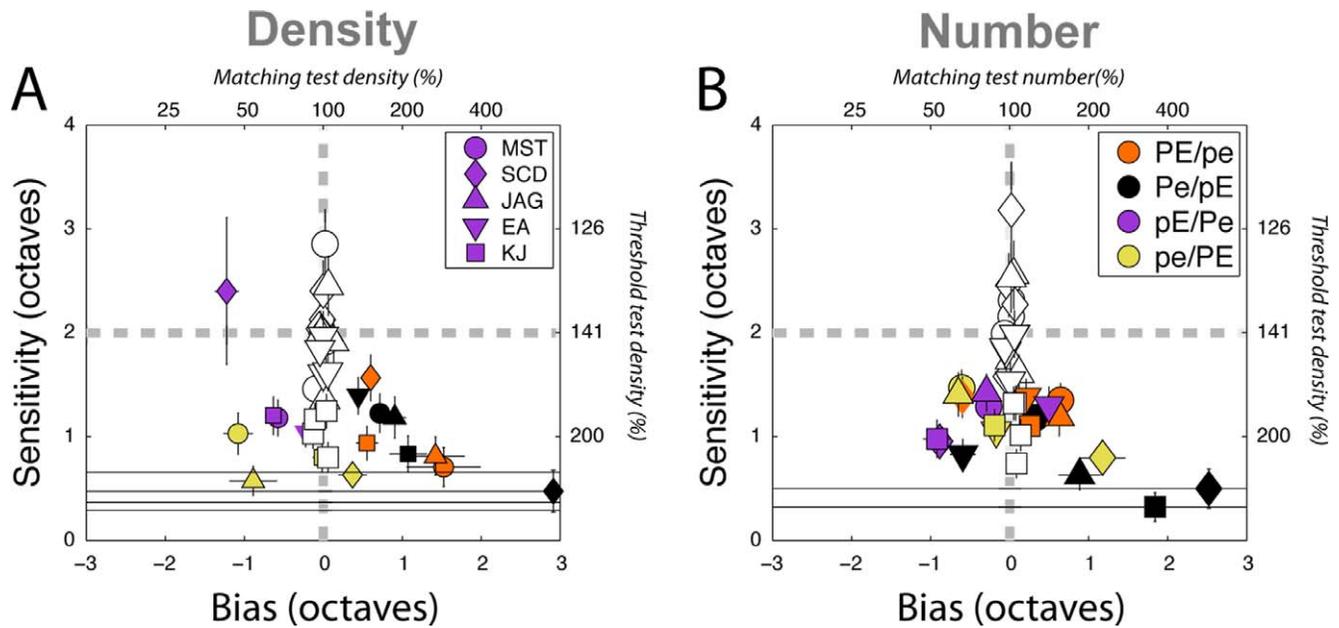


Figure 3. Sensitivity and bias are shown for density judgments (a) and number judgments (b) when both parameters (patch size and element size) are mismatched between the test and reference. Data from the double mismatched conditions are presented with colored symbols (orange, black, purple, and yellow). Note that there are four double mismatched conditions as patch and element sizes could be mismatched in the same direction (PE/pe and pe/PE) or in opposite directions (Pe/pE and pE/Pe). For comparison, data are also shown for double matched conditions, i.e., trials in which the test and reference are matched with respect to element size and patch size (white symbols: PE/PE, for example). See legend to Figure 2 for further details.

reference patch is larger than the test patch irrespective of element size) to fall to the right of center: a positive bias, implying that at the perceived match point there were too many elements in the test. This pattern once again reflects the tendency for larger patches to be perceived as more dense and more numerous. The double mismatched conditions therefore reinforce the findings of the single mismatched conditions and suggest that, for both number and density judgments, mismatching patch size and element size introduces bias and reduces observer sensitivity. This close correspondence between number and density judgments is also captured by a highly significant correlation between density sensitivity and number sensitivity, as well as density bias and number bias ($r_s > 0.6$, $p_s < 0.0001$; Figure 4), and is consistent with a shared underlying metric driving both judgments.

To quantitatively test these findings at the group level, data from Experiment 1 were subjected to two repeated-measures ANOVAs (one for absolute biases and one for sensitivities), each with four factors: task (density or number judgment), relative patch size (matched or mismatched), relative element size (matched or mismatched), and replicate number (1, 2, 3, or 4). Replicate number was included as a factor in the model because, for each condition type, e.g., a density judgment matched for patch and element size, four different patch size/element size combinations were possible (PE/PE, Pe/Pe, pE/pE, pe/pe; see previous discussion). Analyses

show that, with respect to absolute biases, there was no effect of task type ($F_{(1,4)} = 4.47$, $p = 0.1$) or replicate number ($F_{(3,12)} = 0.07$, $p = 0.98$), but a highly significant effect of relative patch size ($F_{(1,4)} = 22.29$, $p = 0.009$) and a weaker (though significant) effect of relative element size ($F_{(1,4)} = 11.48$, $p = 0.03$). Similar results were found with respect to sensitivity: no effect of task type ($F_{(1,4)} = 1.23$, $p = 0.33$) or replicate number ($F_{(3,12)} = 1.59$, $p = 0.24$) with significant effects of relative patch size ($F_{(1,4)} = 24.02$, $p = 0.008$) and relative element size ($F_{(1,4)} = 17.71$, $p = 0.01$).

Discussion

These data demonstrate that mismatching patch size reduces the sensitivity of relative density and number judgments and introduces systematic biases that are consistent with larger patches being perceived as more numerous and more dense, replicating our previous study (Dakin et al., 2011). A similar bias, studied using a comparable 2-IFC paradigm, has been reported for number judgments (Tokita & Ishiguchi, 2010), although no effect of patch size manipulation was found when the reference consisted of an internal standard (Allik, Tuulmets, & Vos, 1991; Ross & Burr, 2010), highlighting the importance of observers making explicit judgments on a pair of stimuli. That there is a substantial effect of patch size on number and density

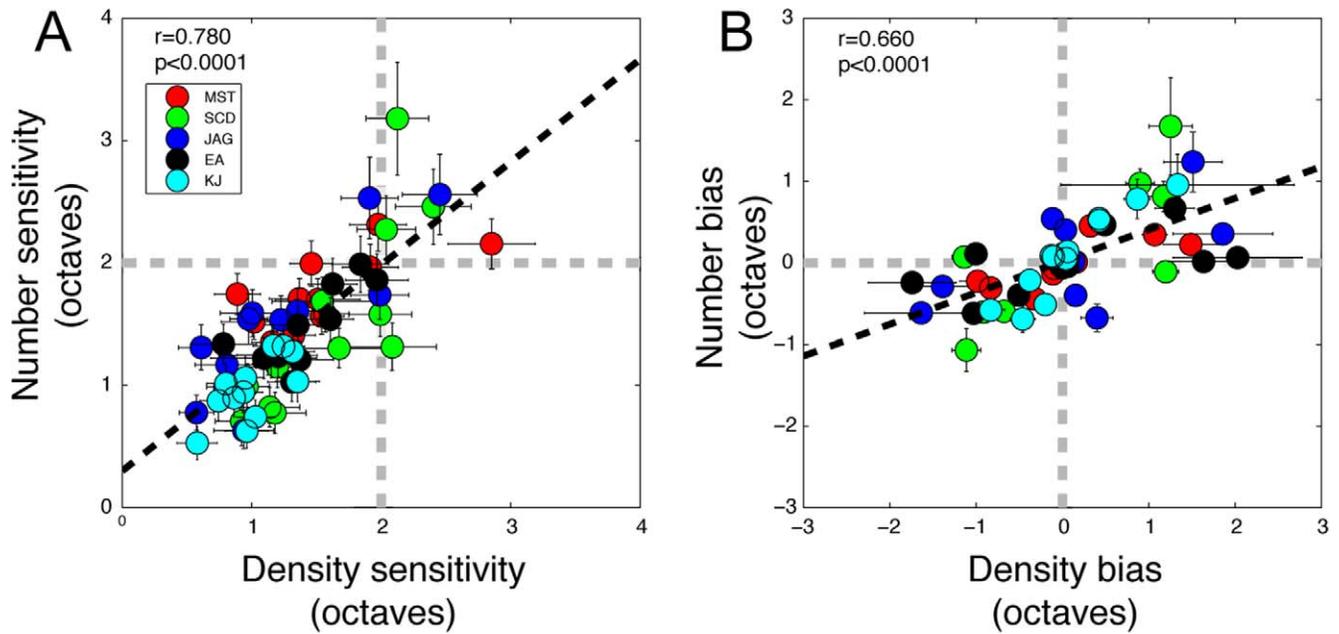


Figure 4. Correlations between density/number sensitivity (a) and density/number bias (b). each data point represents a distinct condition/observer (observers coded by color). Note that, while both sensitivity and bias are highly correlated, biases are greater for density judgments than they are for number judgments (b); hence the slope is < 1 (0.39). In contrast, the slope for sensitivity is closer to 1 (0.84). Vertical dotted grey lines denote zero bias. Horizontal dotted grey lines denote average sensitivity as recorded in the original study (Dakin et al., 2011). Error bars plot the standard deviation of fit parameters derived from bootstrapping.

judgments is inconsistent with either density or number being independent visual attributes. Rather, we suggest that each is derived from a common perceptual metric and that the response-ratio model (Dakin et al., 2011) can account for the effect of patch size mismatch on the bias and sensitivity of both number and density discriminations (see modeling section). To reiterate, this model takes the relative output of high and low spatial frequency (SF) filters (each of which roughly captures element number and patch size respectively) as a common first stage in the estimation of both density and number. However, because the low SF output does not rise fast enough as patch size increases, the response ratio is biased toward overestimation of density for larger patches (mirroring the observers' own biases). In order to estimate *number* under conditions of patch size mismatch, the authors propose that the visual system attempts to essentially recover the raw high SF response by reweighting the response-ratio by an estimate of relative patch size (actually derived using the relative low SF responses). This manipulation gives a workable numerosity estimate with relatively low bias. (See Dakin et al., 2011 for a more in-depth discussion.)

The data from Experiment 1 demonstrate that mismatches in element size also severely disrupt performance, resulting in a significant increase in bias and decrease in sensitivity. Though the sign of this bias varies between observers, for any given observer it is

largely consistent across judgment types (number and density), implying that biases are nonrandom. It is interesting to note therefore that previously reported effects of element size on perceived numerosity, where found, are varied, with two studies being consistent with a small element size/high numerosity bias (Sophian, 2007; Tokita & Ishiguchi, 2010), one with a large element/high numerosity bias (Hurewitz, Gelman, & Schnitzer, 2006), and a third reporting changes in sensitivity only (Ross, 2003). Discrepancies in the literature and our own intersubject differences may therefore reflect multiple cognitive strategies, a notion that is at least consistent with several of our observers reporting a tendency to compensate for intrinsic bias by systematically reversing the sign of their responses. Regardless of the origin of these effects, however, both density and number judgments were biased by mismatches of element size, and sensitivity was reduced (Figure 2 and 3). Interindividual variation in these values was also tightly correlated across judgment types (Figure 4), findings that are entirely consistent with a common metric underlying density and number judgments. Could this common metric again reflect the use of the response-ratio? To do so, the response-ratio estimate could be weighted by relative element size (as well as patch size). However, it would seem that, at least within this experimental paradigm, observers often scale inappropriately, albeit in a similar manner, for both density and number estimates.

Experiment 2: manipulation of element size alone

Whilst the effects of mismatching patch size in [Experiment 1](#) were consistent with the pattern of results previously reported (Dakin et al., 2011), the magnitude of the biases reported here were approximately double those of our earlier study. The only major difference between this and the earlier study was that here observers made judgments under conditions of high stimulus uncertainty as element size was systematically manipulated along with patch size, whereas in the original experiment only patch size was manipulated. Accordingly, judgments under conditions of dual uncertainty (i.e., when both patch and element size simultaneously differed within a trial) were highly unreliable ([Figure 3](#)). Consequently, we wondered whether the effect of element size mismatch observed when patch size and element size varied within the same block (high stimulus uncertainty; [Experiment 1](#)) would be diminished with reduced stimulus uncertainty. To test this hypothesis, element size was manipulated whilst observers performed density and number discriminations on test and reference patches that were fixed and matched with respect to patch size.

Methods

Five observers (four from [Experiment 1](#); all experienced psychophysics observers, three naïve to the purpose of the study) performed number and density discriminations under conditions of fixed patch size and varying element size. Patch radius was fixed at 4° , whilst test and reference elements could have Gaussian envelopes with a standard deviation of 2.5 or 5 arcmins (PE/PE, Pe/Pe, PE/Pe, Pe/PE). All observers completed a minimum of one run per judgment type. Note that in this experiment, because test and reference patches were always matched in size, density and number covaried; as a result, both cues were simultaneously available on both “density” and “number” runs.

Results

To determine whether observers’ biases and sensitivities differed between density and number runs, a series of paired-samples *t*-tests were performed for each condition comparing parameter estimates between judgment types. As expected (given that density and number covaried), there were no significant differences between parameter estimates for the two judgment types (all p s > 0.05) even without correction for multiple comparisons. In addition, density/number

sensitivities and biases were both highly correlated (p s < 0.01). Consequently, data from density and number runs were pooled (before refitting the data afresh) to increase the confidence of the fits. The parameter estimates for this pooled data are presented in [Figure 5](#). As is clearly evident, whilst sensitivity is still greatest when element size is matched between test and reference patches (red and black data points typically fall above the blue and green), with the exception of one observer (SCD) all biases reported in [Experiment 1](#) collapse, so that the data are tightly clustered about zero on the abscissa. This is reflected in the outcome of two repeated-measures ANOVAs, one for sensitivity and one for bias (two factors: relative element size [matched or mismatched] and replicate [1 and 2]). Replicate was included as a factor in the analysis as, for each condition type, two different element size combinations were possible: matched (E/E and e/e) and mismatched (E/e and e/E). Analyses indicate a significant effect of element size on sensitivity ($F_{(1,4)} = 10.03$, $p = 0.03$), but not on absolute bias ($F_{(1,4)} = 1.68$, $p = 0.27$), and no effect of replicate number on either ($F_{(1,4)} = 1.51$, $p = 0.29$; $F_{(1,4)} = 1.48$, $p = 0.29$).

Discussion

The data reported in [Experiments 1](#) and [2](#), in conjunction with our previous results (Dakin et al., 2011), are consistent with observers being able to at

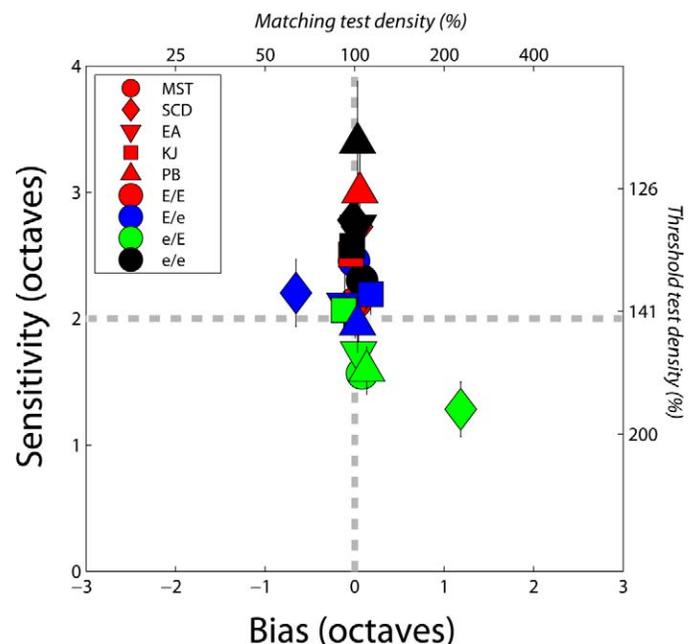


Figure 5. Sensitivity and bias are shown for density and number judgments (combined prior to fitting) under conditions of matched patch size and varying element size. Error bars plot the standard deviation of fit parameters derived from bootstrapping.

least partially compensate for mismatches in patch size or element size when making judgments of relative number or density. Under conditions of low uncertainty, biases are greatly reduced (effect of patch size mismatch; compare [Experiment 1](#) with the original study, Dakin et al., 2011) or largely non-existent (effect of element size mismatch; compare [Experiment 2](#) with [Experiment 1](#)). In contrast, under conditions of high uncertainty (i.e., in blocks in which patch size and element size are simultaneously manipulated/interleaved; [Experiment 1](#)), observers are typically poor at correcting for patch/element size mismatches, and resulting biases are large. This provides further evidence against the use of the method of single stimuli in experiments examining numerosity and/or density perception; the resulting reduction in stimulus uncertainty through the development of an internal standard (Morgan, Watamaniuk, & McKee, 2000) may hide underlying bias and sensitivity changes that would otherwise be present.

Experiment 3: manipulation of luminance contrast

Having shown that mismatches in patch size and element size have similar effects on density and number sensitivity/bias, we tested whether this close association would survive mismatches in luminance contrast between test and reference patches. An independent estimate of number might be highly susceptible to these manipulations—an estimator based on raw stimulus energy should see higher contrast displays as more numerous, for instance. In contrast, because density necessarily requires a comparison between the elements present (e.g., high SFs in the response-ratio model) and the total area covered by the stimulus (e.g., low SFs in the response-ratio model), luminance contrast would be divided out by these computations. Of course, as above, if both numerosity and density discriminations are subserved by a common mechanism, we would expect to find similar effects on each.

Methods

The relative mismatch in patch size was reduced from an octave to half an octave (test and reference patches could have radii of 2.4° or 3.4°) to reduce observer bias and hence the likelihood that the tails of fit psychometric functions would fall out of the stimulus range examined. The relative mismatch in patch size was reduced rather than increasing the stimulus range because of the space restrictions involved in generating small patches with high density

and numerosity. In addition, only mismatched patch size conditions were included in this experiment (P/p and p/P), as patch size was not the parameter of interest, only a tool to decouple number and density. Element size was fixed at 5 minutes of arc. Luminance contrast was systematically manipulated so that, on half of all trials, the elements in one of the patches had a contrast that was twice that of the other patch; resulting patches had Michelson contrasts of 100% and 50% respectively. The increased contrast was randomly assigned to either the test or the reference patch on a trial-by-trial basis. This resulted in a $2 \times 2 \times 2$ design with three factors: task (density and number), relative patch size (reference > test [P/p] and test > reference [p/P]) and luminance contrast (matched and mismatched). All observers performed a minimum of two runs per judgment type (number and density), except for observer RS who performed one of each.

Results

In [Figure 6](#), density (a) and number (b) sensitivities under conditions of matched and mismatched contrast are plotted for all five observers. For the purposes of presentation, the relative patch size factor was collapsed; sensitivity parameter estimates from the reference > test and test > reference conditions were averaged following fitting (for each observer). For three out of the five observers, there seemed to be very little effect of mismatching contrast on density or number sensitivity (observers MST, RS, and KJ). For the two remaining observers (EA and JAG), sensitivity was clearly elevated in the contrast-matched conditions (relative to the unmatched). However, when sensitivity was analyzed at the group level in a repeated measures ANOVA, no effect of task type ($F_{(1,4)} = 1.04$, $p = 0.37$), relative patch size ($F_{(1,4)} = 1.04$, $p = 0.37$), or relative contrast ($F_{(1,4)} = 4.35$, $p = 0.11$) was found. Despite this, individuals' density and number sensitivities were significantly correlated ($r = 0.45$, $p = 0.04$). Thus, although there is little-to-no effect of variations in luminance contrast, the effects that do occur are consistently applied to both number and density judgments.

Discussion

Our results demonstrate that there is little-to-no effect of variations in luminance contrast on judgments of number and density. This is consistent with the finding that number/density aftereffects using random dot stimuli are relatively insensitive to changes in contrast (Durgin, 2001), which may reflect an early monocular contrast-normalization stage upstream from the locus of density/number adaptation (Durgin,

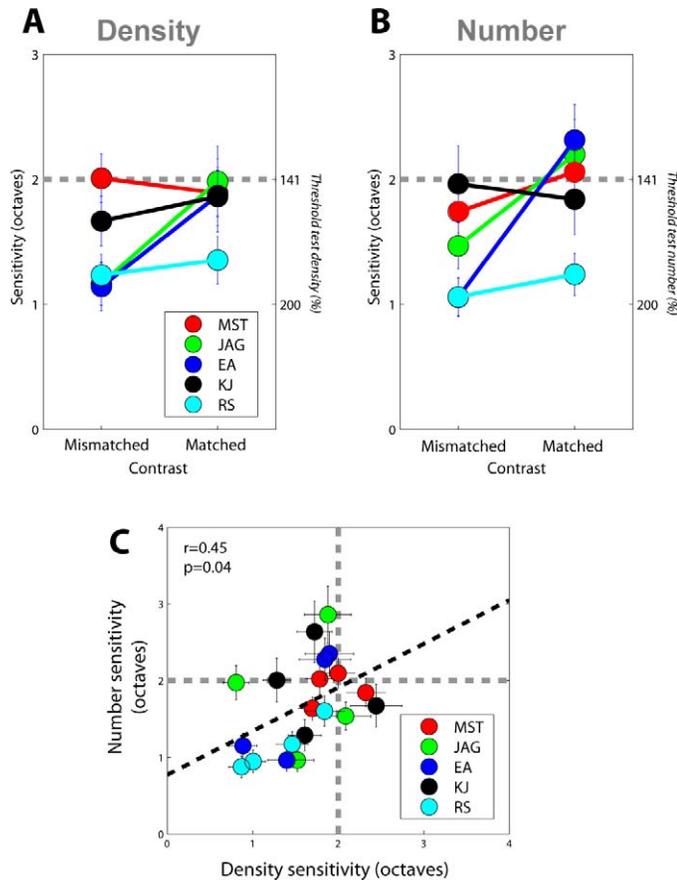


Figure 6. Sensitivity for density (a) and number (b) judgments under conditions of matched luminance contrast (contrast of test and reference elements fixed at 50%) and mismatched contrast (contrast of reference or test elements randomly doubled). (c) Correlation between density/number sensitivity; each data point represents a specific condition/observer (observers coded by color). Vertical dotted grey line denotes zero bias. Horizontal dotted grey line denotes average sensitivity as recorded in the original study (Dakin et al., 2011). Error bars plot the standard deviation of fit parameters derived from bootstrapping.

2001; Durgin & Hammer, 2001). There was some variation between observers, however, suggesting either interobserver variation in strategy or differential sensitivity to the manipulation. It is possible that, if a wider range of contrasts had been tested, sensitivity might have fallen for all observers. For the observers who were more strongly affected by these manipulations, reduced sensitivity in the contrast mismatched condition could imply either that observers' decisions were biased toward selecting the higher contrast (more salient) patch as a result of attentional capture (a nonperceptual bias), or alternatively, that luminance contrast genuinely affects the perceived numerosity/density of a stimulus (a perceptual bias). (See Carrasco, Fuller, & Ling, 2008 and Prinzmetal, Long, & Leonhardt, 2008 for a related discussion on the relative role or perceptual versus nonperceptual biases in the

field of attention.) Notwithstanding, the critical findings of this experiment as they relate to our first prediction in the Introduction are unambiguous: (1) within each individual, the effect of contrast on number and density performance was virtually identical (compare Figure 6a and 6b); (2) there was no main effect of task type at the group level; and (3) density/number sensitivities were significantly correlated (Figure 6c). This is again consistent with the use of a common metric for both number and density judgments.

Experiment 4: manipulation of contrast-polarity

Experiments 1 to 3 have shown that density and number judgments are affected almost identically by manipulations of patch size, element size, and element contrast. In Experiment 4, we examined the effects of manipulating element contrast-polarity. In our earlier study (Dakin et al., 2011), contrast-polarity was shown to have little effect on predicted number discrimination thresholds, a property that is achieved as a result of an early rectification stage to the model (Figure 3). Once again, a model based on a common metric and shared mechanisms would predict similar effects on number and density judgments.

Methods

Unless stated otherwise, stimulus parameters were identical to those outlined in Experiment 3. Instead of contrast being manipulated (fixed here at 50%), the contrast-polarity of individual elements in test and reference patches was systematically varied. On one-third of trials, both patches contained elements of random contrast-polarity (as in Experiments 1 and 2; random condition); on another third of trials, both patches contained elements of a single uniform (and matched) polarity (all elements were either white or black; matched condition); and on the remaining trials, one patch was comprised of random contrast-polarity elements, while the other was comprised of a single uniform polarity (black or white; mismatched condition). Hence, there were three factors to the manipulation: task type (density and number), relative patch size (reference > target [P/p] and target > reference [p/P]), and relative polarity (random, matched, mismatched).

Results

In Figure 7, density (a) and number (b) sensitivities are plotted for each of five observers under conditions

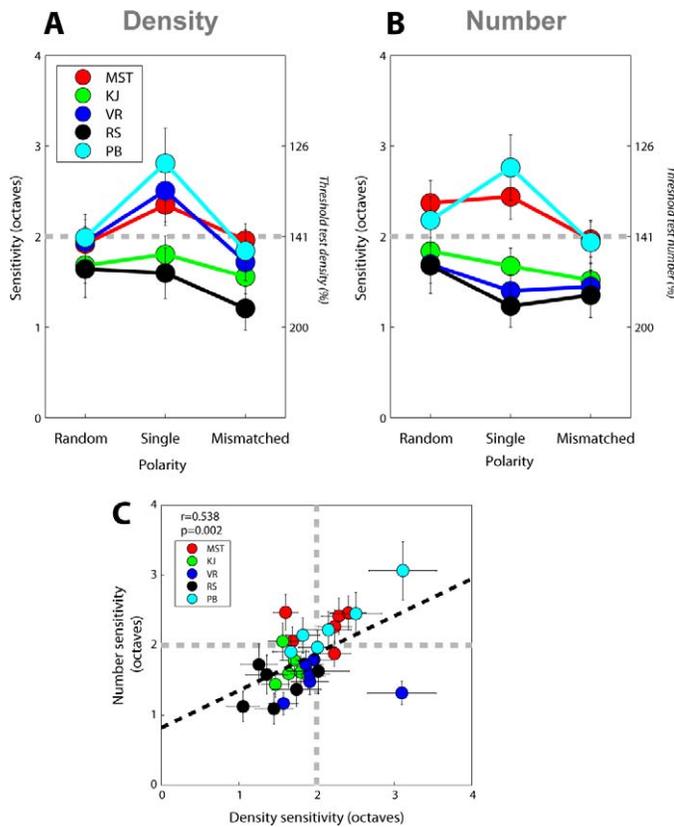


Figure 7. Sensitivity for density (a) and number (b) judgments with either random contrast-polarity (all elements in the test and reference are of random contrast-polarity), single contrast-polarity (both test and reference of a single contrast-polarity), and mismatched contrast-polarity (all elements in one of the patches of random contrast-polarity, all elements in the other patch of a single uniform polarity). In (c), the correlation between density and number sensitivity is shown for all five observers tested. Vertical dotted grey line denotes zero bias. Horizontal dotted grey line denotes average sensitivity as recorded in the original study (Dakin et al., 2011). Error bars plot the standard deviation of fit parameters derived from bootstrapping.

of random, single, and mismatched contrast-polarity. For the purpose of presentation, the patch size factor was collapsed; sensitivity parameter estimates from the reference > test and test > reference conditions were averaged following fitting. Once again, there appears to be considerable variation in strategy amongst observers. Two observers (MST and PB) showed a clear tendency toward increased sensitivity in the single polarity condition for both judgment types, whilst VR showed a similar effect for density, but not number. At the group level, this is reflected in a small but significant effect of relative contrast-polarity ($F_{(2,8)} = 5.7$, $p = 0.03$), a significant interaction between task type and relative contrast-polarity ($F_{(2,8)} = 6.08$, $p = 0.03$), but no main effects of task type ($F_{(1,4)} = 0.3$, $p = 0.61$) or relative patch size ($F_{(1,4)} = 0.51$, $p = 0.52$). Finally, there was a

tight correlation between density and number sensitivities ($r = 0.54$, $p = 0.002$).

Discussion

As in previous experiments, we observe broadly similar patterns of dependence of number and density thresholds on the independent variable. That said, whilst there was a general trend for elevated sensitivity in the single polarity condition, this pattern was more pronounced/consistent in the density task. Notice, however, that it is only the performance of observer VR that differs markedly between density and number tasks: VR exhibits a clear single polarity advantage for density, but none for number. (In Figure 7c, the outlying nature of this single condition performance is highly evident.) This pattern of results can be explained by assuming that observer VR uses local luminance as a cue to density in the single polarity condition, a strategy that would not be reliable for number judgments, or any judgments in Experiments 1 to 3 for that matter. In contrast, to explain PB's and MST's increased sensitivity for both tasks in the single polarity condition, one would have to assume that these observers are using local energy and total energy as cues to the density and number tasks respectively. Thus, it would seem that by fixing the contrast, contrast-polarity, and size of individual elements, additional cues were made available to the observers (e.g., local/global luminance). Irrespective of the interindividual variability reported, it is worth noting that, once again at the group level, there was no main effect on task type, and further, density/number sensitivities were tightly correlated. These findings are still consistent with a common underlying metric for relative number and density judgments, whilst highlighting a potential for exploitation of additional cues when available.

Experiment 5: manipulation of available attentional resources

If density and number estimation rely on a common underlying metric and engage the same computations, they should present comparable demands on available attentional resources. In contrast, if either were “primary”, then it should be less influenced by divided attention. If density estimates were derived from number estimates, for instance, the higher-order calculation of density would likely require higher cognitive resources than the more “direct” estimation of number. To test this prediction, we used a dual-task attentional load paradigm to manipulate available attentional resources whilst observers performed basic number and density discriminations.

Methods

To manipulate available attentional resources, a dual-task attentional load paradigm adapted from that described by Vetter, Butterworth, and Bahrami (2008) was used. In addition to number and density discriminations, observers had to perform a speeded target detection task based on a simultaneously presented diamond (4 DVA wide; presented at fixation) that was divided into composite colored segments. Under low load conditions, observers had to detect targets based solely on the presence of the color red, whereas in the high load conditions, targets were determined by the conjunction of color and the spatial arrangement of constituent segments (Figure 1, as well as Figure 1 in Vetter, Butterworth, & Bahrami, 2008). On each trial, observers first responded to the central task, indicating by button press whether a target was “present” or “absent,” before responding to the secondary discrimination task, indicating whether the left patch or the right patch was more dense/numerous. For the secondary task, all stimulus parameters matched those described in Experiment 3 with the exception that all elements were of random contrast-polarity. Hence, there were three factors to the main body of the experiment: secondary task type (density or number judgment), relative patch size (reference > target and target > reference), and primary task attentional load (high and low). In addition, all tasks were performed individually to get an estimate of baseline performance (single task density discrimination, single task number discrimination, primary task low attentional load, primary task high attentional load). Observers were instructed to prioritize accuracy over speed. All observers performed a minimum of two runs for the dual-task conditions. Response time and accuracy (% correct) were measured for the primary target detection task. Response times greater than 3 seconds were excluded from the analysis on the basis that they involved a pause in the experiment rather than a slow response. Whilst this cutoff is arbitrary, changing it had negligible effects on results. For the secondary task, estimates of sensitivity and bias were derived as outlined in previous experiments.

Results

In Figure 8, group level data are presented. For the purpose of presentation, the patch size factor was collapsed; sensitivity estimates from the reference > test and test > reference conditions were averaged at the observer level following fitting. In Figure 8a, group mean response times are shown for the low and high attentional load detection tasks. Data are presented from single task and both double task performances

(with a concurrent density judgment [red data points] and with a concurrent number judgment [green data points]). Attentional load clearly elevates response times (main effect of attentional load: $F_{(1,4)} = 498$, $p < 0.0001$), as does the concurrent performance of a secondary task ($F_{(2,8)} = 36.04$, $p < 0.0001$). This was driven by a cost of performing a concurrent secondary task (i.e., either secondary task), rather than an exceptionally high cost of a density or number judgment specifically. Thus, there was no significant difference between response times for double task number and double task density conditions at either low or high attentional load levels (*post hoc* paired samples *t*-tests: $ps > 0.05$).

Next, consider task accuracy on the primary color detection task as measured by percent correct responses (Figure 8b); whilst a similar pattern emerges (there is a trend for attentional load and the concurrent performance of a secondary task to impair performance), there were no significant effects of secondary task ($F_{(2,8)} = 3.26$, $p = 0.09$) or attentional load ($F_{(1,4)} = 6$, $p = 0.07$). This may be partially due to a ceiling effect (accuracy rarely dropped below 95%, even in the high load double task attentional conditions). Irrespective, both primary task measures indicate that density and number tasks have no differential effects on concurrent target detection performance.

In Figure 8c and d, group level sensitivities and absolute biases are shown for the secondary (density and number) tasks at three different primary task attentional load levels: none (no concurrent detection task), low (concurrent color detection task), and high (concurrent color/spatial arrangement detection task). It is clearly evident from both these plots that performance for density and number judgments are identically affected by the primary task manipulation; attentional load reduces both density and number sensitivity with very little effect on biases. Hence there was a main effect of attention ($F_{(2,8)} = 21.39$, $p = 0.001$) with no effect of task ($F_{(1,4)} = 5.61$, $p = 0.08$) or relative patch size ($F_{(1,4)} = 0.12$, $p = 0.75$) on sensitivity. Nor was there any interaction between attention and task ($F_{(2,8)} = 0.21$, $p = 0.82$). With respect to biases, there was no main effect of task ($F_{(1,4)} = 0.17$, $p = 0.71$), attention ($F_{(2,8)} = 3.48$, $p = 0.08$), or relative patch size ($F_{(1,4)} = 1.46$, $p = 0.29$), nor an interaction between task and attention ($F_{(2,8)} = 1.14$, $p = 0.37$).

Discussion

The results indicate that density and number discriminations incur a comparable load on attention. This contrasts with an earlier study in which number estimation (above the subitizing range) was found to be unaffected by attentional load (Burr, Turi, & Anobile,

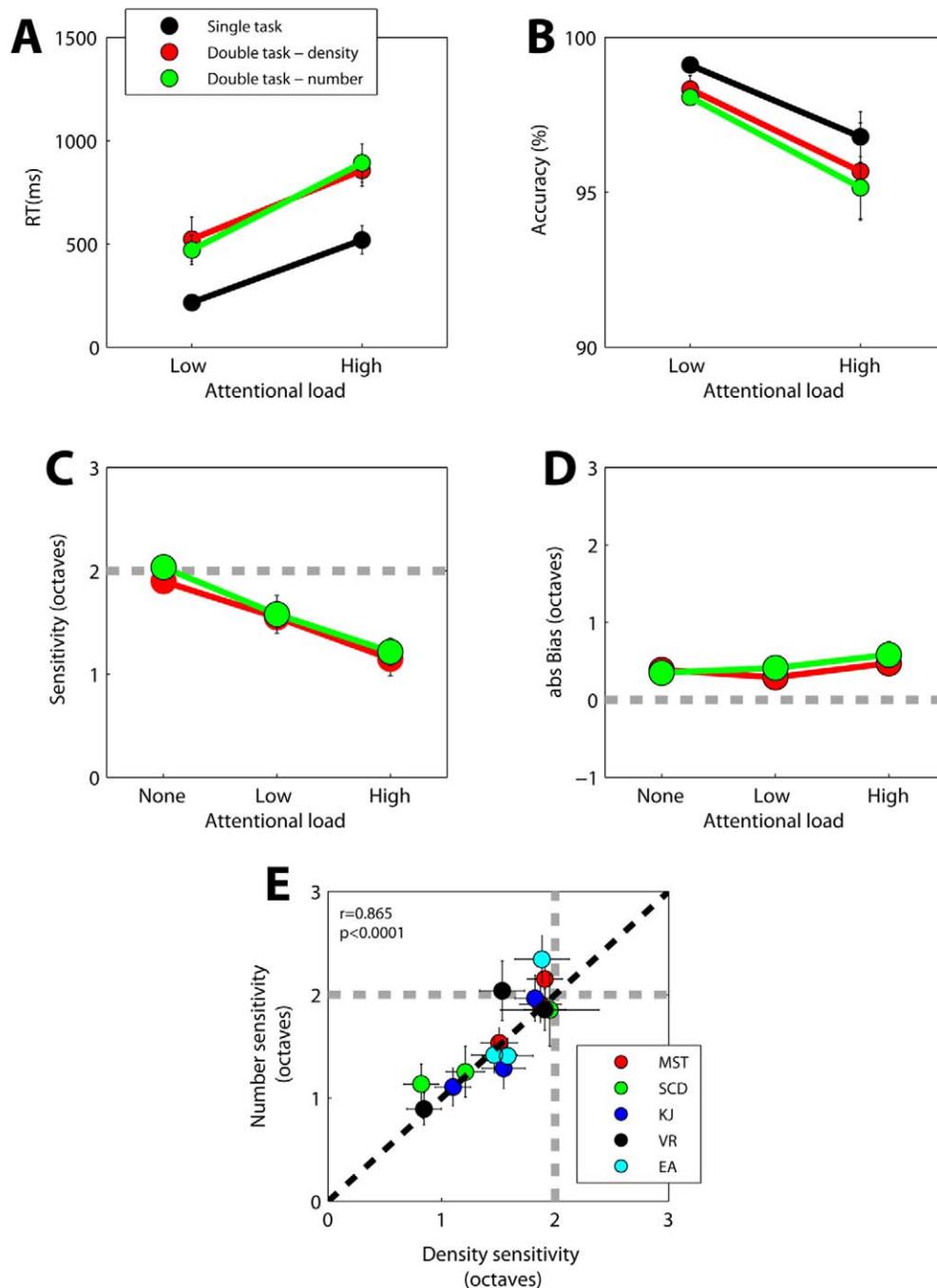


Figure 8. Data from the dual-task attentional manipulation paradigm. (a) Group mean response times, and (b) group mean accuracies (% correct) plotted for the primary (target detection) task as a function of attentional load. Data are presented for single task and both dual-task conditions (i.e., when performed alone [black data points], in conjunction with a density judgment [red data points], or in conjunction with a number judgment [green data points]). (c) and (d) plot group mean sensitivities and biases for the secondary tasks (density judgments [red data points] and number judgments [green data points]) as a function of attentional load (none, low, or high). In (e), individual density sensitivities are plotted against number sensitivities for all observers. Error bars plot the standard error of the mean (a to d) or the standard deviation of fit parameters (e) derived from bootstrapping.

2010). However, there are a number of critical differences between their study and ours, so it is possible that the primary source of attentional interference in our experiment is not derived from density/

number estimation per se but the comparison process itself. That is, the estimation process may be preattentive, but the process of comparison between patches to perform the 2-IFC discrimination might require atten-

tion. Enumeration processes may also be more efficient for low element numbers, making Burr and colleagues' use of eight elements or fewer another significant difference from our own study, which used a test range centered on 128 elements. Alternatively, the dominant source of interference reported here may reflect the need to simultaneously distribute attention across multiple spatial locations (Duncan, 1980; Tibber, Grant, & Morgan, 2009). This seems unlikely, however, because in the paradigm used by Burr, Turi, and Anobile (2010), observers were required to attend to a large region of the visual field (16° of visual angle). Irrespective, the data reported suggest that density and number judgments incur either negligible attentional costs or else equivalent attentional costs over and above those derived from other aspects of the paradigm. Once again, the data are perfectly consistent with a common underlying metric for number and density discrimination, and provide no reason to posit the existence of distinct mechanisms.

Experiment 6: modeling of the data

Finally, we tested the ability of the response-ratio model (Dakin et al., 2011), which is based on the notion of a common metric for number and density discrimination, to predict human psychophysical data gathered in Experiments 1 to 4. Results from Experiment 5 were not modeled as the response-ratio captures front-end processes involved in the initial stages of number/density estimation, and as such, has little to say about cognitive effects or the role of attention.

Methods

Predictions of the response-ratio model to Experiments 1 to 4 were derived from Monte Carlo simulations of the data. Reference and test stimuli were identical to those described in the Methods sections for Experiments 1 to 4, except that a broader stimulus range and finer sampling of cue levels were used (64 trials per each of 17 cue levels spanning 4 octaves [reference level ± 2 octaves]). For each condition and task, the model's responses were plotted as a function of stimulus cue; these were then fit with a cumulative Gaussian function (as described in the General Methods section), so that estimates of sensitivity and bias could be obtained.

For full details of the model, please see Dakin et al. (2011). In brief, when presented with a stimulus (a grey-scale image), the image is rectified and convolved with a pair of Laplacian-of-Gaussian (center-surround) filters:

one characterized by a high SF (R_{hi}) and one characterized by a relatively low SF (R_{lo}), the output of which broadly correlate with element number and patch size respectively (see Figure 5 in Dakin et al., 2011). A ratio of these outputs (C), corrupted by multiplicative Gaussian random noise ($2^{\gamma\sigma}$) with a standard deviation σ is then taken as a correlate of a patch's density ($C = 2^{\gamma\sigma} R_{hi}/R_{lo}$). To simulate a discrimination judgment, this process is carried out independently for both a test (C_t) and a reference (C_r) patch, and a ratio of the two is taken, generating a density response-ratio ($d_{tr} = C_t/C_r$). The denser stimulus is then selected on the basis of whether d_{tr} is less than, or greater than, 1. To model number discriminations, the same process is performed, except that the density response ratio is weighted by an estimate of the size-mismatch between patches (W), thereby generating a number response ratio ($n_{rt} = Wd_{tr}$). This size-mismatch weighting is derived from the relative output of low SF filters to the reference and test image, and includes a second noise term, S ($W = [2^{\gamma\sigma} [tR_{lo}/rR_{lo}]] \wedge 2^{\gamma S}$).

The basic model therefore has two free parameters: early multiplicative noise (σ) and late noise (S). To fit data in the original study, these were set to 0.1 and 1.9 respectively (Dakin et al., 2011). To obtain the fits described here, the local noise term (σ) was kept at 0.1 for tasks 2–4, but was increased to 0.2 for Experiment 1. This may reflect increased stimulus uncertainty in Experiment 1, which putatively arises from the simultaneous manipulation of patch size *and* element size in interleaved conditions (see Experiment 2 discussion). The second noise term (S) was fixed at 0.6. The only modification made to the original model is that, here, the SF of the fine-scale filter was dependent on the element sizes presented on any given trial. The standard deviation of the fine-scale filter was thus set to 1 or 5 arcmins for trials containing exclusively small or large elements respectively, and to an intermediate value (3 arcmins) for trials containing a mixture of small and large elements. The standard deviation of the coarse filter was always fixed at 13 arcmins.

Results

In Figure 9, human group mean psychophysical performance for density and number discriminations (green and red bars respectively) are presented for Experiments 1 to 4, along with the predictions of the response-ratio model (blue dots). The small black crosses denote individual data. Considering bias first (right-hand columns in each panel), whilst there is considerable noise in the individual data, the predictions of the model capture many of the general trends

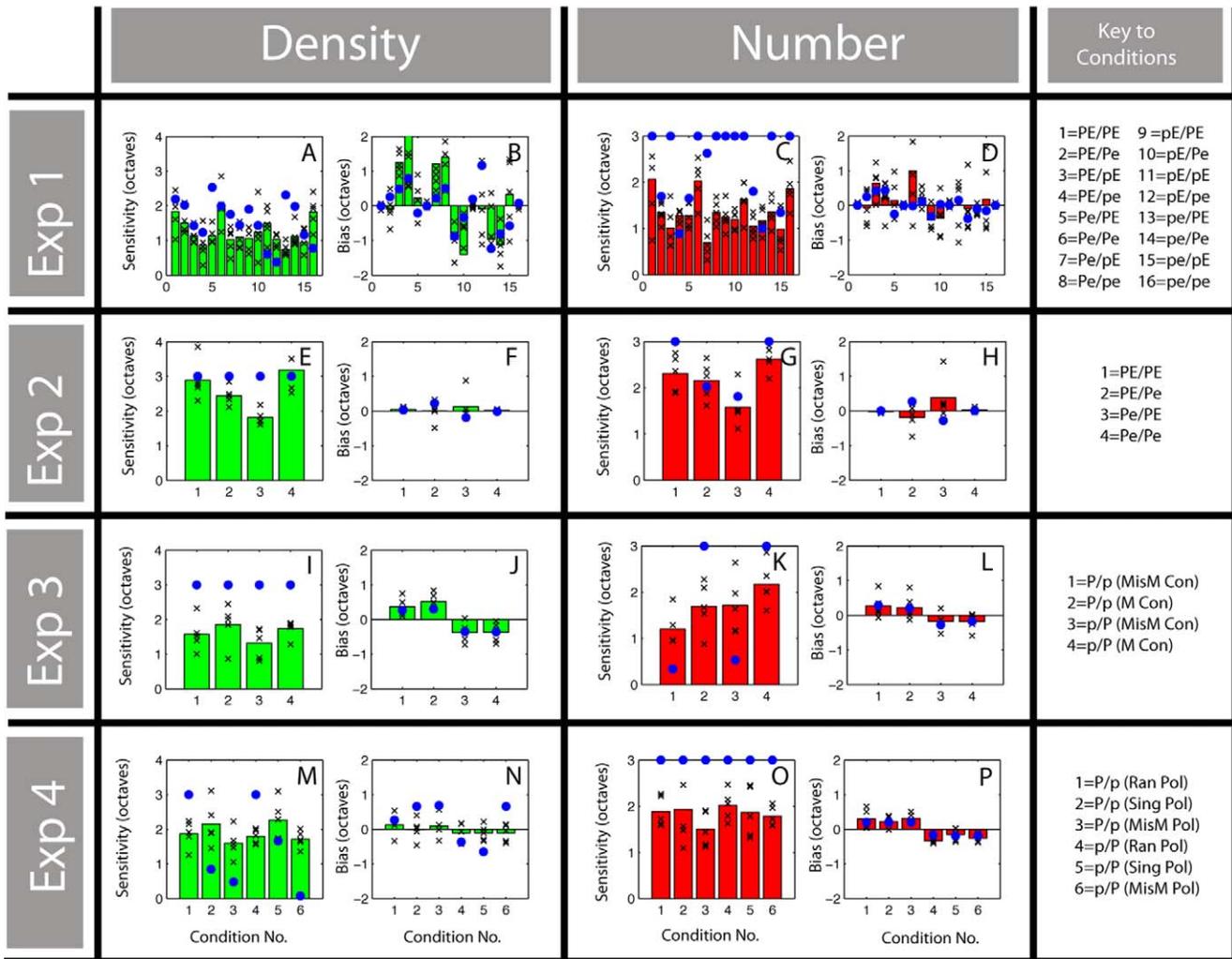


Figure 9. Group mean psychophysical performances for density (green bars) and number (red bars) discriminations are presented for [Experiments 1](#) to [4](#) along with the predictions of the response-ratio model (blue dots). Individual data are presented as small black crosses. Condition numbers are listed on the abscissa. The key to each condition is presented in the right-hand panel. (“P” = large patch; “p” = small patch; “E” = large elements; “e” = small elements; MisM Con = mismatched contrast; M Con = matched contrast; Ran Pol = random polarity; Sing Pol = single polarity; MisM Pol = mismatched polarity). Remember that the mismatch in patch sizes was reduced in [Experiments 3](#) and [4](#) (see Methods for further details).

at the group level. In [Experiment 1](#), the predictions capture the flip in the sign of bias between conditions 1–8 and conditions 9–16 ([Figure 9b](#) and [d](#)), reflecting the tendency for larger patches to be perceived as more dense or more numerous (in conditions 1–8, the reference is large; in conditions 9–16, the reference is small; see key to conditions in [Figure 9](#), right-hand panel). Similarly, the model predicts that a similar pattern of bias is induced—though reduced in magnitude—for smaller mismatches in patch sizes ([Experiments 3](#) and [4](#); [Figure 9j](#) and [l](#), [n](#), and [p](#)), and that biases for number tasks exhibit the same sign, but are typically attenuated further ([Figure 9d](#) relative to [Figure 9b](#)). These findings imply that bias is largely dominated by (and scales with) the relative mismatch in patch sizes. While the model predicts that mismatches

in *element size* should also induce modest biases (conditions 2 and 3 in [Experiment 2](#); [Figure 9f](#) and [H](#)), with the exception of a single observer (SCD), human performance is largely unbiased. As described in the discussion to [Experiment 2](#), we suggest therefore that under conditions of low stimulus uncertainty, observers are able to at least partially correct for the small effects of mismatches in element size.

The model also does reasonably well at predicting density thresholds for most conditions from [Experiments 1](#) to [4](#) ([Figure 9a](#), [e](#), [i](#), and [m](#)), as well as for number thresholds from [Experiments 2](#) and [3](#) ([Figure 9g](#) and [k](#)). However, it clearly fails to predict behavior successfully in several conditions. Where the model outperforms humans (e.g., [Experiment 1](#), number task; [Figure 9c](#)) it is not overly informative; the model merely

defines a theoretical ceiling in performance. More interesting, however, are the conditions in which the human observers outperform the model. These are indicative of limitations in the model and may reflect the existence of additional cues exploited by the human observer. For example, the model underperforms on a subset of conditions in the number task of [Experiment 3](#) ([Figure 9k](#)). Whilst the relative trends in the human data are captured—sensitivity is highest when contrast is matched (conditions 2 and 4)—the model performs worse than every human observer in the contrast-mismatched conditions (conditions 1 and 3). Critically, this is not the case for the density task ([Figure 9i](#)). One possibility therefore is that when the density response-ratio (d_{tr}) is weighted by an estimate of relative patch size to derive an estimate of relative number ($n_{tr} = Wd_{tr}$), contrast dependence (originally removed by the taking of filter output ratios to *each* patch) is reintroduced into the metric. This is because the weighting is based on the relative output of the coarse filter to the reference and test stimuli and is therefore not contrast normalized. Because the higher contrast signal was randomly assigned to the reference or test on each trial, this will have manifested itself as a trial-by-trial fluctuation in bias and, specifically, elevated thresholds in the contrast mismatched number conditions. As the human behavior did not show such a pronounced collapse in performance in the contrast-mismatched conditions, a role for additional (non-contrast-based) cues in the estimation of relative patch size are implicated.

The second place where the model underperforms is in the density task from [Experiment 4](#) ([Figure 9m](#)). The model's sensitivity essentially collapses in conditions 2, 3, 5, and 6. These are all conditions that involve at least one patch comprised of single-polarity (i.e., all black or all white) elements. We believe that this may arise because when two matched-polarity elements overlap, they sum, resulting in clipping and a disproportionate boost in the low SFs. Hence, the high SF filter's response saturates and sensitivity falls drastically, an effect that would only occur in the density tasks because of the extremely high densities reached at the highest cue levels. As human performance does not collapse under these conditions, the implication is that additional cues, e.g., luminance cues, may begin to be informative in such high-density stimuli.

Discussion

Whilst there is considerable noise in the human psychophysical data, the response-ratio model does a surprisingly good job of predicting general trends in observers' behavior across a range of stimulus manipulations. The strength of this finding is increased by the

fact that [Experiments 1 to 4](#) were not originally designed to constrain the model, and that predictions have been fit to some 120 data-points ([Figure 9](#)) with few free parameters. Further, conditions in which the model's predictions and human psychophysical behavior diverged proved to be highly informative and will inform further developments of the response-ratio model. Nonetheless, the broad correspondence between observed and predicted behavior (particularly with respect to biases) support the notion of a shared common metric to density and number estimation.

General discussion

The data reported here are unambiguously consistent with density and number estimates being derived from a common underlying metric. Thus, we have shown that:

- Density and number estimates are not independent. Manipulations of either dimension (by mismatching patch size) result in reduced sensitivity and elevated biases in the other (replicating previous findings; [Experiment 1](#)).
- The effects of manipulating patch size, element size, and luminance contrast are near identical on relative density and number judgments ([Experiments 1 to 4](#)).
- Density sensitivity and number sensitivity are consistently correlated across all experimental manipulations undertaken ([Experiments 1 to 5](#)).
- When the range is not restricted, density biases and number biases are also significantly correlated ([Experiment 1](#)).
- Number and density judgments make indistinguishable demands on available attentional resources ([Experiment 5](#)).
- General trends in the data from [Experiments 1 to 4](#) can be predicted with a relatively simple filter-energy-based model that relies on a common first-stage metric for number and density estimation.

There were only two findings that distinguished density and number estimation behaviorally. First, biases under conditions of mismatched patch size were considerably greater for density judgments. However, this pattern of results falls out of the response-ratio model by assuming that number estimates are weighted by an approximation of relative patch sizes, a necessary step to the judgment if estimates of numerosity are to be anything but wildly inaccurate ([Figure 9](#)). Another important point to note is that, whilst the magnitude of these biases differed, they were highly correlated, which is to be expected if both are derived from a common metric. Secondly, when luminance contrast, contrast-polarity, and the size of individual elements were held

constant (Experiment 4; matched polarity condition), one observer's behavior for density and number estimation diverged. Thus, VR showed elevated sensitivity for density judgments (compared with when contrast-polarity was mismatched across patches or randomized), but no improvement for number estimation. However, when contrast, contrast-polarity, and element size are held constant, absolute local luminance provides an additional cue to the density task, but not to the number task. Consequently, it is likely that this divergence of behavior merely reflects the potential to use additional cues when they are available. This possibility is supported by the fact that a similar benefit of uniform polarity elements was not found when luminance-balanced difference of Gaussian (DoG) elements—which would not provide absolute luminance cues—were used (Ross, 2003).

How can these data be reconciled with the notion that number is extracted independently of other stimulus dimensions? Evidence for this position comes from two sets of studies by Burr and Ross: the first involving adaptation (Burr & Ross, 2008b) and the second is a set of discrimination judgments similar to those reported here (Ross & Burr, 2010). We will not discuss the former here; firstly, because we did not use an adaptation paradigm so that our results do not speak directly to their data, and secondly, because we feel that Durgin convincingly demonstrated that the data are consistent with adaptation following the density of the adapter (or some correlate of density) rather than numerosity per se (Durgin, 2008), i.e., patch area was critical. With respect to their second study, Ross and Burr (2010) suggest that “perhaps the strongest evidence for independent mechanisms for sensing numerosity and texture (at least dense texture) was that while numerosity estimates show strong dependency on luminance [...], texture density was completely independent of luminance over this range.” However, in the study performed therein, very different experimental stimuli were used to probe the effects of luminance on density and number estimation. This was to some extent necessary, as the authors wanted to test for any association between number estimation and *texture density* specifically (as proposed by Durgin, 2008). However, it is therefore unclear whether the differential effects they report reflect distinct mechanisms of density and number estimation per se, or differences in the stimuli used and cues that they carry. Indeed, when we assessed the effects of luminance contrast on perceived number and density discrimination using matched stimuli and paradigms, we found near-identical effects on density and number performance (Experiment 3). Note, however, that unlike Ross and Burr (2010) who manipulated luminance using single-polarity elements, we used mixed-polarity elements, and thus, at the level of the stimulus patch,

manipulated luminance *contrast*, not *absolute* luminance. Nonetheless, it would be of interest to compare the effects of luminance on *texture density* and *random dot density* estimation; as taken together, our data and that of Ross and Burr (2010) raise the possibility that they are not subserved by identical processes.

In the same study (Ross & Burr, 2010), the authors provide additional results they suggest imply number is estimated independent of density; number thresholds were no greater than density thresholds, and further, holding density constant (by yoking patch size and number) did not impair number performance. Presumably, the assumption here is that if density (D) and number (N) are connected, it is because the visual system calculates density first, and then derives number by multiplying through by area (A) ($D = N/A$, therefore $N = D * A$) in a noisy process that elevates thresholds. Thus, if density cues are available, this noisy transformation may be bypassed, resulting in reduced thresholds. However, in the response-ratio model, density is not calculated first. Instead, a common metric based on the relative output of high and low SF filters is estimated, which may then be corrected for mismatches in patch size (or element size) to generate a correlate of density and/or number. The model, therefore, does not necessarily predict reduced thresholds when density cues are present. In fact, a reverse prediction could be made: when density cues are not present, and patch size is yoked to element number (constant density condition in Ross and Burr's experiment, 2010), patch size provides an additional cue to the number task, and thresholds could fall. Indeed, this is precisely what they report (figure 3a, Ross & Burr, 2010). Thus, the finding that number thresholds were no greater than density thresholds (a result that we replicate here), and further, that performance was no worse when density was held constant does not provide evidence of independent processing. If anything, the close and robust correspondence between density and number thresholds may simply be indicative of a common underlying mechanism.

In conclusion, whilst the data reported are entirely consistent with density and number estimation being based on a common metric, we have also shown that the effect sizes are extremely sensitive to manipulations of stimulus uncertainty (Experiments 1 and 2), that observers may exploit additional cues when they are available (Experiment 4), and also that there may be considerable interindividual variation in strategy (Experiments 3 and 4). When taken together with previous reports that number biases are lost with practice (Tokita & Ishiguchi, 2010) and that the range, i.e., number of elements in the stimulus (Durgin, 1995), as well as the experimental paradigm used (Allik, Tuulmets, & Vos, 1991; Dakin et al., 2011; Ross & Burr, 2010; Tokita & Ishiguchi, 2010) are highly critical to

many of the reported effects, it is no longer surprising that there are discrepancies in the existing number/density literature. Nonetheless, by testing parallel effects on density and number judgments using identical stimuli and matched experimental paradigms, we believe that our findings are robust, and further, that they support a response-ratio model of density/number discrimination.

Acknowledgments

This research was funded by the Wellcome Trust. SCD is also supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. JAG is currently supported by a Marie Curie Intra-European Fellowship.

Commercial relationships: None.

Corresponding author: Marc S. Tibber.

Email: mtibber@yahoo.com.

Address: Institute of Ophthalmology, University College London, Bath Street, London, UK.

Footnote

¹Note that while number estimates could in principle be derived from the sole output of high spatial frequency detectors, this would fail to capture the biases that we observe with mismatches in patch size (Dakin et al., 2011).

References

- Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, 49(4), 303–314. [PubMed].
- Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, 53(4), 290–295. [PubMed].
- Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, 54(3), 695–701. [PubMed].
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. [PubMed].
- Burr, D., & Ross, J. (2008a). A visual sense of number. *Journal of Vision*, 8(6):691, <http://www.journalofvision.org/content/8/6/691.abstract>, doi:10.1167/8.6.691.
- Burr, D., & Ross, J. (2008b). A visual sense of number. *Current Biology*, 18(6):425–428. [PubMed].
- Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*, 10(6):20, <http://www.journalofvision.org/content/10/6/20>, doi:10.1167/10.6.20.
- Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: a reply to Prinzmetal, Long, and Leonhardt (2008). *Perception & Psychophysics*, 70(7), 1151–1164. [Article].
- Dakin, S., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences USA*, 108(49), 19552–19557. [Article].
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1–2), 1–42. [PubMed].
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychology Review*, 87(3), 272–300. [PubMed].
- Durgin, D. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology*, 21(1), 149–169. [PubMed].
- Durgin, F. H. (2001). Texture contrast aftereffects are monocular; texture density aftereffects are binocular. *Vision Research*, 41, 2619–2630. [Article].
- Durgin, F. H., & Hammer, J. T. (2001). Visual aftereffects of sequential perception: dynamic adaptation to changes in texture density and contrast. *Vision Research*, 41, 2607–2617. [Article].
- Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology*, 18(18), R855–856. ; author reply R857–858.
- Durgin, F. H., & Huk, A. C. (1997). Texture density aftereffects in the perception of artificial and natural textures. *Vision Research*, 37(23), 3273–3282. [Article].
- Frith, C. D., & Frith, U. T. (1972). The solitaire illusion: An illusion of numerosity. *Perception & Psychophysics*, 11(6), 409–410.
- Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *American Journal of Psychology*, 91(2), 267–273. [PubMed].
- Ginsburg, N. (1991). Numerosity estimation as a function of stimulus organization. *Perception*, 20(5), 681–686. [PubMed].

- Ginsburg, N., & Nicholls, A. (1988). Perceived numerosity as a function of item size. *Perceptual & Motor Skills*, 67(2), 656–658. [PubMed].
- He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: implications for configural processing. *Psychonomic Bulletin & Review*, 16(3), 509–517. [PubMed].
- Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences USA*, 103(51), 19 599–19 604. [Article].
- Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences USA*, 103(9), 3486–3489. [Article].
- Morgan, M. J., Watamaniuk, S. N. J., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, 40(17), 2341–2349. [Article].
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. [PubMed].
- Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. <http://www.palamedestoolbox.org/>.
- Prinzmetal, W., Long, V., & Leonhardt, J. (2008). Involuntary attention and brightness contrast. *Perception & Psychophysics*, 70(7), 1139–1150. [PubMed].
- Ross, J. (2003). Visual discrimination of number without counting. *Perception*, 32(7), 867–870. [PubMed].
- Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, 10(2):10, 1–8, <http://www.journalofvision.org/content/10/2/10>, doi:10.1167/10.2.10.
- Sophian, C. (2007). Measuring spatial factors in comparative judgments about large numerosities. Paper presented at the Proceedings of the 3rd international conference on Foundations of augmented cognition.
- Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, 36(2), 97–127. [PubMed].
- Tibber, M. S., Grant, S., & Morgan, M. J. (2009). Oculomotor responses and visuospatial perceptual judgments compete for common limited resources. *Journal of Vision*, 9(12):21, 1–31, <http://www.journalofvision.org/content/9/12/21>, doi:10.1167/9.12.21.
- Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception, & Psychophysics*, 72(7), 1839–1853. [PubMed].
- Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: evidence against a pre-attentive subitizing mechanism. *PLoS One*, 3(9), e3269. [Article].
- Vos, P. G., van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research*, 50(3), 148–154. [PubMed].
- Xu, F., & Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology*, 25, 103–108.