BRILL

Seeing and Perceiving 25 (2012) 365–395

brill.nl/sp

A Bayesian Observer Replicates Convexity Context Effects in Figure–Ground Perception

Daniel Goldreich [1,*] **and Mary A. Peterson** [2]

[1] Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

[2] Department of Psychology and Cognitive Science Program, University of Arizona, Tucson, Arizona, USA

Received 14 September 2010; accepted 11 February 2012

Abstract

Peterson and Salvagio (2008) demonstrated convexity context effects in figure–ground perception. Subjects shown displays consisting of unfamiliar alternating convex and concave regions identified the convex regions as foreground objects progressively more frequently as the number of regions increased; this occurred only when the concave regions were homogeneously colored. The origins of these effects have been unclear. Here, we present a two-free-parameter Bayesian observer that replicates convexity context effects. The Bayesian observer incorporates two plausible expectations regarding three-dimensional scenes: (1) objects tend to be convex rather than concave, and (2) backgrounds tend (more than foreground objects) to be homogeneously colored. The Bayesian observer estimates the probability that a depicted scene is three-dimensional, and that the convex regions are figures. It responds stochastically by sampling from its posterior distributions. Like human observers, the Bayesian observer shows convexity context effects only for images with homogeneously colored concave regions. With optimal parameter settings, it performs similarly to the average human subject on the four display types tested. We propose that object convexity and background color homogeneity are environmental regularities exploited by human visual perception; vision achieves figure–ground perception by interpreting ambiguous images in light of these and other expected regularities in natural scenes.

Keywords

Figure–ground, scene segregation, object convexity, configural cue, Gestalt principles, natural scene statistics, Bayesian inference, computational model

[*] To whom correspondence should be addressed. E-mail: goldrd@mcmaster.ca

DOI:10.1163/187847612X634445

1. Introduction

Broadly speaking, the task of the visual system is to infer the physical scene from the image on the retina. This is a challenging task because multiple scenes could give rise to the same retinal image. One focus of scene segregation research has been *figure–ground perception*. Figure–ground perception is a possible outcome when two contiguous regions of the visual image share a border (Peterson, 2003; see Fig. 1A for illustration). When figure–ground perception occurs, only one of the contiguous regions (the *figure*) is perceived to have a definite shape; the figure is also perceived as the near surface at the shared border. The other region (the *ground*) is perceived to continue behind the figure, and appears to be unshaped near the shared border (although it can be shaped by other borders, as in Fig. 1A). Thus, figure–ground perception entails three-dimensionality, inasmuch as a figure is perceived in front of a background. We stress that figure–ground perception is not the only possible outcome when two contiguous regions share a border (cf., Kennedy, 1974). Shared borders could also delimit regions of a flat pattern (e.g., Fig. 1B) or the corners of a geometric object (e.g., Fig. 1C). These examples suggest that scene segregation perceived at a border cannot be predicted by local cues alone — the context matters.

Surprisingly, the role of context in figure–ground perception has received little attention. Instead, research has focused on identifying cues or shape properties that predict which regions of the visual field will be perceived as figures. The Gestalt psychologists identified a number of such cues, now known as the ‘classic’ Gestalt configural cues; other configural cues have been identified more recently (e.g., Hulleman and Humphreys, 2004; Palmer and Brooks, 2008; Palmer and Ghose, 2008; Peterson and Gibson, 1994; Peterson *et al.*, 1991; Vecera *et al.*, 2002). One of the classic cues, convexity (as opposed to concavity) has been considered by some theorists to be an extremely powerful configural cue. This conclusion followed from experiments such as those of Kanizsa and Gerbino (1976), who showed subjects 8-region black and white displays like the bottom sample in Fig. 2A in which 4 black regions with convex parts alternated with 4 white regions with concave parts (or *vice versa*). Their subjects reported perceiving the regions with convex parts as fig-

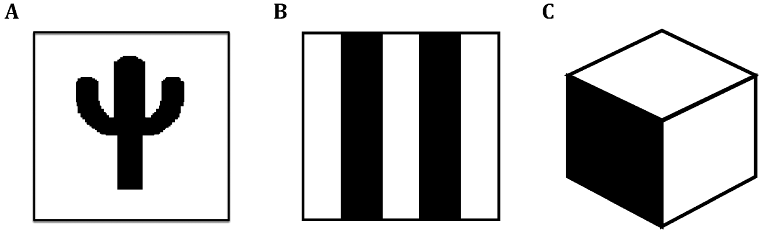

Figure 1. Shared borders, such as between the black and white regions in these displays, can evoke several types of percepts. (A) Figure–ground perception, with the black region perceived as the figure and the white region perceived as the ground. (B) A two-dimensional pattern. (C) The edges of a three-dimensional object.

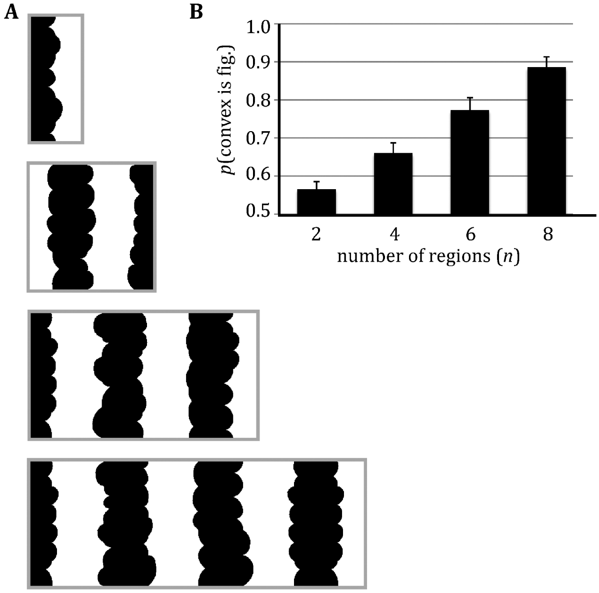

Figure 2. The convexity context effect in black and white displays. (A) Examples of black and white (homogeneous convex, homogeneous concave) displays used in Peterson and Salvagio (2008). In these examples, the convex regions are colored black and the concave regions white; in the experiment, an equal number of displays also had the opposite color scheme. The number of regions (n) was (top to bottom) 2, 4, 6 or 8. A display appeared for 100 ms on a medium gray backdrop, with the central border coinciding with the eyes' fixation point. A small rectangular red probe (not shown) lay within the region either directly to the left or right of the central border. Subjects were asked to indicate their first impression as to whether the probe lay on or off the region they perceived as the *figure*. Subjects were instructed that figures appear to have a definite shape and to lie in front of contiguous regions. (B) The mean frequency with which human subjects responded that a convex region was figure increased consistently with the number of regions in the display. Error bars: 1SE.

ures on an average of 90% of trials. Notably, theories have supposed implicitly that estimates of convexity's effectiveness obtained with displays like those in Fig. 2 would generalize to all displays, and even to single borders.

Recently, Peterson and Salvagio (2008) found that the effectiveness of convexity as a configural cue varied with the context, in particular with the number and color of alternating convex and concave regions (see Note 1). Using black and white displays like those used in previous experiments, they found that subjects were increasingly likely to perceive convex regions as figures as the total number of regions increased progressively from 2 to 8 (see Note 2). Using different coloration schemes, they found that displays in which both the convex and concave regions were heterogeneously colored failed to yield a convexity context effect (Fig. 3, display type 4). This result suggested that one or both region types must be homogeneously filled in order for the effect to occur. Investigating further, they found that homogeneous fill in the convex regions was not required; displays with either homogeneously or heterogeneously filled convex regions could yield convexity con-

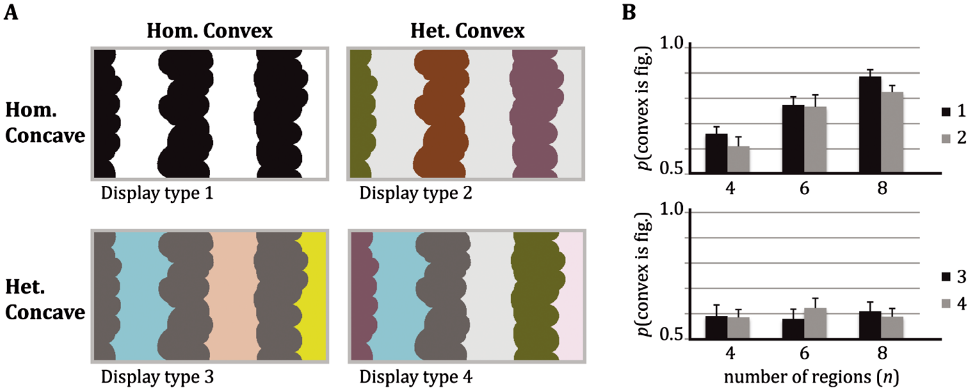

Figure 3. Four display types. (A) The four alternating convex-concave region display types on which we test the Bayesian observer in the present study. For simplicity, we show here only 6-region displays. Hom: homogeneously colored; Het: heterogeneously colored. (B) Human observers show the convexity context effect for display types 1 and 2, but not for display types 3 and 4. Data re-plotted from Peterson and Salvagio (2008).

text effects (Fig. 3A, display types 1 and 2; Fig. 3B, top). By contrast, homogeneous fill in the concave regions was a necessary condition, as displays without this feature failed to produce the effect (Fig. 3A, display types 3 and 4; Fig. 3B, bottom). These results, and others reported by Peterson and Salvagio (2008) using additional display types (e.g., displays with regions defined by straight edges, which did not yield context effects), suggested intriguingly that homogeneity in the color of every other region in fact operates as a *ground* cue, but only when a concomitant cue (i.e., convexity) favors the alternate regions as *figures* (however weakly).

Why does perception gives rise to convexity context effects? A growing body of research suggests that perception relies upon probabilistic (Bayesian) inference, in which observers incorporate their understanding of natural scene statistics, acquired perhaps over a lifetime, into the interpretation of the present image. Thus, an image that is open to multiple interpretations may be disambiguated with the aid of expectation based in experience. Here, we show that the convexity context effects reported by Peterson and Salvagio (2008) are replicated by a Bayesian observer that incorporates two plausible expectations regarding natural scenes: (1) objects tend to be convex, and (2) the background tends to be homogeneously colored. Object convexity is directly supported by natural scene data (Burge *et al.*, 2010). Ground color homogeneity is indirectly supported by the observation that in natural scenes color varies less across the surface of a single object than it does between objects (Fine *et al.*, 2003; Ing *et al.*, 2010). Because the two ground regions that flank an object may belong to a single occluded surface, it follows that these flanking ground regions should tend, more than separate objects do, to share the same color. The results of the present study are consistent with the hypothesis that human visual perception exploits these regularities of the environment.

2. Methods

Here we develop a Bayesian observer for figure–ground perception in alternating convex-concave region displays such as those used by Peterson and Salvagio (2008). In that study, subjects reported in a forced-choice experiment which one of two adjacent regions (a convex region and a concave region) they perceived as *figure* (i.e., a foreground object). Our interest lies in the proportion of times that subjects identified the convex region as figure, as a function of the total number of regions and the type of display (see Fig. 3).

2.1. Definitions

It is useful at this point to define some important terms. We define a *figure* as a foreground object in front of a partially visible background (*ground*). Examples of figures are dogs playing on the grass, people walking in front of a wall, and trees outlined against the morning sky. The words 'on' 'in front of', and 'against' all convey the idea that figures are closer to the observer than is the ground, which continues behind the figures, occluded from view. We will use the term *three-dimensional* (3D) to refer to scenes that contain figure–ground structure (i.e., depth separation between objects and background).

Importantly, not all scenes contain figure–ground structure. A scene may instead occupy a single (not necessarily flat) surface. Like scenes with figure–ground structure, many single-surface scenes project a multi-region visual image. Examples include the body surfaces of many animals, such as leopards (spotted), tigers (striped), zebras (striped) and dogs (various); patterned fabrics such as shirts (plaid, striped, etc.), wallpaper, rugs, curtains and flags; surfaces formed by adjoining objects, such as brick walls, hardwood floors, tiled ceilings and the books pressed tightly against one another in a bookshelf; and patterns of sunlight and shadows cast on a surface (a wall, etc.). These examples lack figures and ground, as we have defined those terms. We will somewhat loosely use the term *two-dimensional* (2D) to refer to such scenes.

We define a *scene* (S) as an arrangement of elements in the external world that casts a visual *image* (I) onto the retina. We are concerned here with scenes whose retinal images consist of n alternating convex and concave *regions*. Each region has two attributes: (1) a *shape* (i.e., a particular convex or concave form), and (2) a *color* (Fig. 3).

2.2. Overview of the Bayesian Observer

The task of the visual system is to infer the scene — or specific scene properties — from the retinal image. In particular, our Bayesian observer will infer whether a scene is 2D or 3D and, if it is 3D, which of the regions in the image represent foreground objects. This is a formidable task, because there is not a one-to-one mapping between scenes and images: different scenes can produce the same retinal image. The visual system, in attempting to reason backwards from effects (images) to their causes (scenes), thus faces a challenging *inverse problem* (Pizlo, 2001), whose solution is inherently probabilistic, that is, uncertain.

The observer can perform optimally in such situations by joining its observation (the visual image) with its knowledge of scene statistics. For instance, the observer may know the approximate value of $p(S)$, the *prior probability* of a particular scene (e.g., that scene's prevalence in nature). The observer may additionally be able to calculate $p(I|S)$, the probability of the visual image, given the scene; this is the *likelihood* of the scene. If the observer is able to specify the priors and likelihoods for different scenes, it can combine these using Bayes' theorem to calculate $p(S|I)$, the *posterior probability* of a scene, given the visual image. The distribution of posterior probabilities over scenes provides the basis for a perceptual inference.

Our Bayesian observer incorporates the expectation that, in naturally occurring 3D scenes, convexity is a figure cue, and homogeneity of color is a ground cue; that is, the Bayesian observer has the expectation that real-world objects tend to be convex, and that real-world backgrounds tend to be at least locally homogeneous in color. These two biases are represented by the observer's q and a parameters, respectively (see below). These are the free parameters of the model, and our efforts here are focused primarily on estimating them.

In 2D scenes, the terms *figure* and *ground* do not carry clear meaning. Before responding whether a convex or concave region is figure, then, the observer first estimates the probability that the scene itself is 3D or 2D. The observer then answers that the convex region is figure with probability:

$$\begin{aligned} &p(\mathit{convex\ region\ is\ figure}|I) \\ &\quad = p(\mathit{convex\ region\ is\ figure}|3D, I)p(3D|I) + (1/2)p(2D|I). \end{aligned} \tag{1}$$

As a reminder, here I refers to the *image* (visual data) and *3D* and *2D* refer to the three-dimensional and two-dimensional scene interpretations.

Note that we assume the observer perceives by sampling from its posterior probability distributions (*Bayesian sampling*; see, e.g., Mamassian *et al.*, 2002; Moreno-Bote *et al.*, 2011; Wozny *et al.*, 2010). For instance, if the observer computes $p(3D|I) = 0.9$, then it will perceive the scene to be 3D on 90% of repeated trials. When the observer perceives the scene as 3D, it answers that the convex region is figure with a frequency equal to the posterior probability, under the 3D scene interpretation, that the convex region is figure. When the observer perceives the scene as 2D, it simply answers randomly (probability $1/2$) that the convex region is figure, because it is forced to give one answer or the other (see Note 3). This simple decision rule leads the observer, like human subjects, to respond stochastically rather than deterministically; thus, the observer will show binomial response variability upon viewing the same display in repeated trials, even in the absence of image or neural noise. Posterior sampling does not maximize expected utility under standard cost functions; in this sense, it is a suboptimal strategy (but see Discussion).

In the following sections, we proceed through the steps needed to solve equation (1) for each of the displays (Fig. 3). In Section 2.3, we state the Bayesian observer's assumptions regarding the visual world, and define its two free parameters. In Section 2.4, we derive the observer's posterior probability that the convex region is figure, under the 3D scene interpretation: $p(\textit{convex region is figure}|3D, I)$. In Section 2.5, we derive the posterior probabilities of the 3D and 2D scene interpretations: $p(3D|I)$ and $p(2D|I)$. Finally, in Section 2.6 we put these results together to solve equation (1) for each display.

2.3. The Bayesian Observer's Assumptions Regarding the Visual World

Here we enumerate the Bayesian observer's assumptions regarding the statistics of the visual world. In the process, we define the two free parameters of the model.

Assumption 1: The observer assumes that 3D objects (a.k.a. *figures*) are more likely to be convex than concave in shape. In particular, with respect to the types of stimuli used by Peterson and Salvagio (2008) (P & S), the observer considers that an object has probability p_1 of looking like one of the convex regions, and probability p_2 of looking like one of the concave regions used in that study, where $p_1 > p_2$. We will see that only the ratio, p_1/p_2, matters to the observer's perception. This ratio we call q. Note that $1 < q \leqslant \infty$:

$$q = \frac{p(\textit{convex}_{P\&S}|\textit{object})}{p(\textit{concave}_{P\&S}|\textit{object})} = \frac{p_1}{p_2}. \tag{2}$$

Assumption 2: The observer assumes that there are N_c possible colors, and that an object can have any color, with equal probability, unrelated to the colors of other objects in the scene. Thus, the probability of a particular color for any object is $1/N_\mathrm{c}$. By contrast, the observer assumes that consecutive ground regions have a tendency to share the same color (because consecutive ground regions often reflect a single structure occluded by an object). Specifically (indexing the regions in the scene from $k = 1$ to n, from left to right), the observer assumes that the probability that region $(k + 2)$ has the same color as region (k), given that these are ground regions (i.e., the probability that the ground does not change color when 'out of sight' behind a figure) is a/N_c, where $1 < a < N_\mathrm{c}$. Thus, the observer's generative model for ground color is a Markov chain, the probability assigned by the observer to the color of a ground region depending only on the color of the preceding ground region. Note that the a-parameter is the likelihood ratio:

$$a = \frac{p(\textit{color of region}_{k+2} = \textit{color of region}_k|\textit{the regions are grounds})}{p(\textit{color of region}_{k+2} = \textit{color of region}_k|\textit{the regions are figures})}. \tag{3}$$

Assumptions 3–5: In addition to the above fundamental assumptions embodied by its q and a parameters, the observer makes three simplifying equiprobability assumptions:

(3) it is equally probable to encounter an alternating figure–ground scene that begins (first region on left) with a figure, as one that begins with ground;

(4) it is equally probable to encounter a 3D scene whose visual image has alternating convex and concave shapes like those in Fig. 2A, as it is to encounter a 2D scene whose visual image has those shapes; and

(5) a region in a 2D scene can take on any color, with probability $(1/N_c)$.

Note that we have purposefully endowed our observer with what we consider to be the minimal set of plausible expectations regarding the statistics of the visual world that might enable it to interpret the displays shown (Figs 2 and 3). The observer expects only that objects tend to be convex (Assumption 1, q-parameter) and that the ground tends to be homogeneously colored (Assumption 2, a-parameter). In all other respects, it is agnostic with respect to the visual world. For instance, it does not have any expectation for dependencies among the colors of figures in a 3D scene (Assumption 2) or among the colors of regions in a 2D scene (Assumption 5), and it is unable to discern whether a scene is 3D or 2D from the shapes presented (Assumption 4). Our goal is to determine whether this intentionally minimally endowed observer can replicate the convexity context effect in figure ground perception, and if so, what parameter settings (q and a values) it requires. We embark on this project with the goal of developing a minimal observer that, if successful here, will provide a scaffold upon which more sophisticated models may be built.

2.4. Derivation of $p(\textit{convex region is figure}|3D, I)$

Here we derive the observer's posterior probability that the convex region is figure, under the 3D scene interpretation. We assume that, confronted with one of the displays used by Peterson and Salvagio (2008), the observer considers two plausible 3D scene interpretations, which we call hypotheses H_1 (FGFG...) and H_2 (GFGF...). Here, F signifies a region that is a *figure*, G signifies a region that is a *ground*, and the ellipsis indicates that the alternating pattern (FG or GF) continues for the full length of the display (until the total number of regions in the display is reached). Each hypothesis represents the idea that the shapes of ground regions are formed (by occlusion) from the borders of figures (foreground objects). Thus, if the figures are convex, then the visible portions of the ground must be concave, and *vice versa*. We assume that the observer considers any other 3D scene interpretation to have negligible probability.

In the derivations that follow, we consider without loss of generality only displays in which the first (leftmost) of the alternating regions is convex, the second concave, and so on (as in Fig. 3A) (see Note 4). As a consequence, H_1 is equivalent to the hypothesis that the convex regions are figures (and the concave regions are grounds) and H_2 is equivalent to the hypothesis that the concave regions are figures (and the convex regions are grounds). Our goal, then, is to derive, for the four display types (Fig. 3A), $p(H_1|3D, I)$.

In calculating the likelihoods for the two hypotheses, the Bayesian observer considers the entire visual image, specifically the shape (convex or concave) of each region, the color of each region, and number, n, of regions. So, we have:

$$\begin{aligned} p(I|3D, H_1) &= p(\textit{shapes}, \textit{colors}|3D, H_1) \\ &= p(\textit{shapes}|3D, H_1)\, p(\textit{colors}|\textit{shapes}, 3D, H_1), \end{aligned}$$

$$\begin{aligned} p(I|3D, H_2) &= p(\mathit{shapes}, \mathit{colors}|3D, H_2) \\ &= p(\mathit{shapes}|3D, H_2)p(\mathit{colors}|\mathit{shapes}, 3D, H_2). \end{aligned} \tag{4}$$

The observer then uses Bayes' theorem to calculate the posterior probability of H_1:

$$p(H_1|3D, I) = \frac{p(I|3D, H_1)p(H_1|3D)}{p(I|3D, H_1)p(H_1|3D) + p(I|3D, H_2)p(H_2|3D)}. \tag{5}$$

The observer considers H_1 and H_2 to be equally probable, *a priori* (see Section 2.3, Assumption 3); thus, $p(H_1|3D) = p(H_2|3D)$, and these priors cancel out in Bayes' formula. Note also that we can express Bayes' formula in terms of the H_2-to-H_1 likelihood ratio:

$$p(H_1|3D, I) = \frac{p(I|3D, H_1)}{p(I|3D, H_1) + p(I|3D, H_2)} = \frac{1}{1 + \frac{p(I|3D,H_2)}{p(I|3D,H_1)}}. \tag{6}$$

We now determine the posterior probability of H_1, for each of the four display types (Fig. 3).

For *display type 1* (homogeneous convex, homogeneous concave), the Hypothesis 1 and 2 likelihoods are:

$$\begin{aligned} p(I_1|3D, H_1) &= p_1^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})(a/N_\mathrm{c})^{(n/2)-1} \\ &= p_1^{n/2}(1/N_\mathrm{c})^{n}a^{(n/2)-1}, \end{aligned} \tag{7}$$

$$\begin{aligned} p(I_1|3D, H_2) &= p_2^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})(a/N_\mathrm{c})^{(n/2)-1} \\ &= p_2^{n/2}(1/N_\mathrm{c})^{n}a^{(n/2)-1}. \end{aligned} \tag{8}$$

The first factor following the first equal sign in equation (7) is the probability of $n/2$ convex objects (this is equivalent to the probability of the shapes of all regions in the image, because the object shapes determine the ground shapes). The second factor is the probability of the colors of those objects. The third factor is the probability of the color of the leftmost ground region. The fourth factor is the probability of each subsequent ground region having the same color as the ground region that precedes it. Thus, the second, third, and fourth factors together represent the probability of the colors of all regions, given the shapes of those regions. The explanation for equation (8) is identical, except that the first factor is the probability of $n/2$ concave objects, as according to H_2 the objects are concave.

Bayes' formula (equation (6)) then yields the posterior probability for Hypothesis 1:

$$p(H_1|3D, I_1) = \frac{p_1^{n/2}}{p_1^{n/2} + p_2^{n/2}} = \frac{1}{1 + (\frac{1}{q})^{n/2}}. \tag{9}$$

Note that this is the same formula that would result were the observer not to consider color at all, just shapes. The color information has effectively canceled out in Bayes' theorem, since the convex and concave regions are both homogeneous in color. Note that, if the observer were to have no object convexity bias ($q = 1$), then its posterior probability for Hypothesis 1 would equal 0.5, but as q increases towards infinity, the posterior probability approaches one. Finally, note that, for $q > 1$, the posterior probability asymptotically approaches one as the number of regions increases.

For *display type 2* (heterogeneous convex, homogeneous concave), the Hypothesis 1 and 2 likelihoods are:

$$\begin{aligned} p(I_2|3D, H_1) &= p_1^{n/2}(1/N_c)^{n/2}(1/N_c)(a/N_c)^{(n/2)-1} \\ &= p_1^{n/2}(1/N_c)^n a^{(n/2)-1} \end{aligned} \tag{10}$$

$$\begin{aligned} p(I_2|3D, H_2) &= p_2^{n/2}(1/N_c)^{n/2}(1/N_c)\left(\frac{1-(a/N_c)}{N_c-1}\right)^{(n/2)-1} \\ &= p_2^{n/2}\left(\frac{1}{N_c}\right)^n\left(\frac{1-(a/N_c)}{1-(1/N_c)}\right)^{(n/2)-1}. \end{aligned} \tag{11}$$

Equation (10) is identical to equation (7): display types 1 and 2 have identical probability under Hypothesis 1, because in each case the hypothesized ground is homogeneous in color (and the Bayesian observer assigns a single probability — $1/N_c$ — to the color of any object, independently of the colors of the other objects). Note that the fourth factor following the first equal sign in equation (11) arises because each ground region has probability $(1 - (a/N_c))$ of not matching the color of the preceding ground region, divided among $(N_c - 1)$ possible non-matching colors.

The posterior probability for Hypothesis 1 is therefore:

$$\begin{aligned} p(H_1|3D, I_2) &= \frac{p_1^{n/2}a^{(n/2)-1}}{p_1^{n/2}a^{(n/2)-1} + p_2^{n/2}(\frac{1-(a/N_c)}{1-(1/N_c)})^{(n/2)-1}} \\ &= \frac{1}{1+(\frac{1}{q})^{n/2}(\frac{1}{b})^{(n/2)-1}}, \end{aligned} \tag{12}$$

where for convenience of expression we have defined:

$$b = \left(\frac{1-\frac{1}{N_c}}{\frac{1}{N_c}}\right)\Big/\left(\frac{1-\frac{a}{N_c}}{\frac{a}{N_c}}\right) = \frac{N_c - 1}{(N_c - a)/a}. \tag{13}$$

The numerator of b is the probability that a foreground object will not match the color of its (foreground) neighbor, divided by the probability that it will match. The denominator is the probability that a ground region will not match the color of its (ground) neighbor, divided by the probability that it will match. Note that b is not an independent parameter; it is specified by a and N_c. Because $1 < a < N_c$, it follows that $1 < b < \infty$.

Note that, as b approaches one, the observer loses any expectation for a uniform ground color, and so the posterior probability of Hypothesis 1 (equation (12)) approaches the value it would have for a color-free image (or for display type 1). As b approaches infinity, by contrast, the observer increasingly expects the ground to maintain a single color, and so, since only the concave regions are unchanging in color, these are confidently perceived as ground, and the posterior probability of Hypothesis 1 approaches one.

For *display type 3* (homogeneous convex, heterogeneous concave), the Hypothesis 1 and 2 likelihoods (following the same reasoning, and expressing the factors in the same order as in the likelihood formulae above) are:

$$\begin{aligned} p(I_3|3D,H_1) &= p_1^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})\left(\frac{1-(a/N_\mathrm{c})}{N_\mathrm{c}-1}\right)^{(n/2)-1} \\ &= p_1^{n/2}\left(\frac{1}{N_\mathrm{c}}\right)^n\left(\frac{1-(a/N_\mathrm{c})}{1-(1/N_\mathrm{c})}\right)^{(n/2)-1}, \end{aligned} \tag{14}$$

$$\begin{aligned} p(I_3|3D,H_2) &= p_2^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})(a/N_\mathrm{c})^{(n/2)-1} \\ &= p_2^{n/2}(1/N_\mathrm{c})^n a^{(n/2)-1}. \end{aligned} \tag{15}$$

The posterior probability for Hypothesis 1 is therefore:

$$\begin{aligned} p(H_1|3D,I_3) &= \frac{p_1^{n/2}(\frac{1-(a/N_\mathrm{c})}{1-(1/N_\mathrm{c})})^{(n/2)-1}}{p_1^{n/2}(\frac{1-(a/N_\mathrm{c})}{1-(1/N_\mathrm{c})})^{(n/2)-1}+p_2^{n/2}a^{(n/2)-1}} \\ &= \frac{1}{1+(\frac{1}{q})^{n/2}b^{(n/2)-1}}. \end{aligned} \tag{16}$$

Note that, for this display type, the observer's q and b parameters are at odds. Larger q-values indicate higher expectations for convex objects (favoring Hypothesis 1), and larger b-values indicate higher expectations for uniform ground color (favoring Hypothesis 2). As b approaches infinity the observer increasingly expects the ground to maintain a single color, and so, since only the convex regions are unchanging in color, these are more confidently perceived as ground, and the posterior probability of Hypothesis 1 approaches zero.

For *display type 4* (heterogeneous convex, heterogeneous concave), the Hypothesis 1 and 2 likelihoods are:

$$\begin{aligned} p(I_4|3D,H_1) &= p_1^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})\left(\frac{1-(a/N_\mathrm{c})}{N_\mathrm{c}-1}\right)^{(n/2)-1} \\ &= p_1^{n/2}\left(\frac{1}{N_\mathrm{c}}\right)^n\left(\frac{1-(a/N_\mathrm{c})}{1-(1/N_\mathrm{c})}\right)^{(n/2)-1}, \end{aligned} \tag{17}$$

$$\begin{aligned} p(I_4|3D,H_2) &= p_2^{n/2}(1/N_\mathrm{c})^{n/2}(1/N_\mathrm{c})\left(\frac{1-(a/N_\mathrm{c})}{N_\mathrm{c}-1}\right)^{(n/2)-1} \\ &= p_2^{n/2}\left(\frac{1}{N_\mathrm{c}}\right)^n\left(\frac{1-(a/N_\mathrm{c})}{1-(1/N_\mathrm{c})}\right)^{(n/2)-1}. \end{aligned} \tag{18}$$

The posterior probability for Hypothesis 1 is therefore:

$$p(H_1|3D, I_4) = \frac{p_1^{n/2}}{p_1^{n/2} + p_2^{n/2}} = \frac{1}{1 + (\frac{1}{q})^{n/2}}. \tag{19}$$

Note that this is the same posterior that resulted from display type 1 (equation (9)). Under the 3D scene interpretation, the likelihoods for both hypotheses are smaller for display type 4 (equations (17) and (18)) than for display type 1 (equations (7) and (8)), but the likelihood ratios, and therefore the posterior probabilities, are the same. This does not mean, however, that the observer will perform equivalently on the two display types. Indeed, display type 4 favors a 2D scene interpretation, whereas display type 1 favors a 3D scene interpretation, as the following section shows.

2.5. *Derivation of $p(3D|I)$ and $p(2D|I)$*

We now derive the observer's posterior probabilities for the 3D and 2D scene interpretations. We begin by writing Bayes' formula for each posterior probability:

$$\begin{aligned} p(3D|I) &= \frac{p(I|3D)p(3D)}{p(I|3D)p(3D) + p(I|2D)p(2D)} = \frac{1}{1 + \frac{p(I|2D)p(2D)}{p(I|3D)P(3D)}}, \\ p(2D|I) &= \frac{p(I|2D)p(2D)}{p(I|3D)p(3D) + p(I|2D)p(2D)} = \frac{1}{\frac{p(I|3D)p(3D)}{p(I|2D)p(2D)} + 1}. \end{aligned} \tag{20}$$

Since the visual image, I, consists of regions characterized by both shapes and colors, we can write the likelihoods as:

$$\begin{aligned} p(I|3D) &= p(colors|shapes, 3D)p(shapes|3D), \\ p(I|2D) &= p(colors|shapes, 2D)p(shapes|2D). \end{aligned} \tag{21}$$

Substituting equations (21) into equations (20), we obtain:

$$\begin{aligned} p(3D|I) &= \frac{1}{1 + (1/clr)(1/z)}, \\ p(2D|I) &= \frac{1}{(clr)(z) + 1}. \end{aligned} \tag{22}$$

Here *clr* is the *color likelihood ratio* that compares the probability of the colors of the regions in the visual image, given the shapes of the regions, under the 3D and 2D scene interpretations,

$$clr = \frac{p(colors|shapes, 3D)}{p(colors|shapes, 2D)} \tag{23}$$

and z is the probability of encountering a 3D scene, as compared to a 2D scene, whose visual image has those shapes:

$$z = \frac{p(shapes|3D)p(3D)}{p(shapes|2D)p(2D)}. \tag{24}$$

As explained previously (Section 2.3, Assumption 4), our Bayesian observer assumes that $z = 1$. We later explore the effect of varying z, however, so for future reference we leave it in equations (22).

To determine the posterior probabilities (equations (22)), then, we must derive the color likelihood ratio (equation (23)) for each of the four display types. In order to proceed with the derivation, we recall the simplifying assumption (Section 2.3, Assumption 5) that, under the observer's 2D interpretation, all colors are equally probable for any region, regardless of the region's shape:

$$p(colors|shapes, 2D) = (1/N_\mathrm{c})^n. \tag{25}$$

Under the observer's 3D interpretation, by contrast, the probability of the colors depends on the hypothesized identity (figure or ground) of the regions (Section 2.3, Assumption 2). Therefore:

$$\begin{aligned} &p(colors|shapes, 3D) \\ &\quad = p(colors|H_1, shapes, 3D)p(H_1|shapes, 3D) \\ &\qquad + p(colors|H_2, shapes, 3D)p(H_2|shapes, 3D), \end{aligned} \tag{26}$$

where, for all four display types (see, e.g., equation (9)),

$$\begin{aligned} p(H_1|shapes, 3D) &= \frac{1}{1 + (1/q)^{n/2}}, \\ p(H_2|shapes, 3D) &= 1 - p(H_1|shapes, 3D) = \frac{1}{1 + q^{n/2}} \end{aligned} \tag{27}$$

and $p(colors|H_1, shapes, 3D)$ and $p(colors|H_2, shapes, 3D)$ are as shown within the likelihood formulae (Section 2.4), for each display type. For instance, for display type 1 (see equations (7) and (8)):

$$p(colors|H_1, shapes, 3D) = p(colors|H_2, shapes, 3D) = (1/N_\mathrm{c})^n (a)^{(n/2)-1}. \tag{28}$$

It is now a simple matter to evaluate equation (26) for each display type, and substitute it with equation (25) into equation (23). The resulting color likelihood ratios for the four display types are:

$$clr_1 = a^{(n/2)-1}, \tag{29}$$

$$clr_2 = \frac{a^{(n/2)-1}}{1 + (\frac{1}{q})^{n/2}} \left(1 + \left(\frac{1}{q}\right)^{n/2} \left(\frac{1}{b}\right)^{(n/2)-1}\right), \tag{30}$$

$$clr_3 = \frac{a^{(n/2)-1}}{1 + (\frac{1}{q})^{n/2}} \left(\left(\frac{1}{q}\right)^{n/2} + \left(\frac{1}{b}\right)^{(n/2)-1}\right), \tag{31}$$

$$clr_4 = \left(\frac{a}{b}\right)^{(n/2)-1}, \tag{32}$$

where the subscript denotes the display type. Note that for $n > 2$, clr_4 is less than one (since $b > a$), whereas clr_1 is greater than one (since $a > 1$). Thus, display type 4 is more probable under a 2D than under a 3D scene interpretation, whereas the opposite is true for display type 1. Note that, for all four display types, the color likelihood ratio approaches 1 in the limit that a (and therefore b) approaches one. As a approaches one, the observer loses the expectation for ground color homogeneity; for the observer without this expectation, the visual image is uninformative with respect to the dimensionality of the scene.

2.6. Derivation of $p(\textit{convex region is figure}|I)$

Using the results of Sections 2.4 and 2.5, we can now derive, for each display type, the probability that the Bayesian observer will answer that the convex region is figure. We begin by substituting equations (22) into equation (1) and rearranging to obtain:

$$p(\textit{convex region is figure}|I) = 0.5 + \frac{p(H_1|3D, I) - 0.5}{1 + (1/z)(1/clr)}. \tag{33}$$

We then substitute into equation (33) the relevant expressions for each display type, derived in Sections 2.4 (equations (9), (12), (16), (19)) and 2.5 (equations (29)–(32)), to obtain the final formulae we have sought:

$$p(\textit{convex region is figure}|I_1) = 0.5 + \frac{\frac{1}{1+(1/q)^{n/2}} - 0.5}{1 + (1/z)(1/a)^{(n/2)-1}}, \tag{34}$$

$$p(\textit{convex region is figure}|I_2) = 0.5 + \frac{\frac{1}{1+(1/q)^{n/2}(1/b)^{(n/2)-1}} - 0.5}{1 + \frac{1+(1/q)^{n/2}}{za^{(n/2)-1}(1+(1/q)^{n/2}(1/b)^{(n/2)-1})}}, \tag{35}$$

$$p(\textit{convex region is figure}|I_3) = 0.5 + \frac{\frac{1}{1+(1/q)^{n/2}b^{(n/2)-1}} - 0.5}{1 + \frac{1+(1/q)^{n/2}}{za^{(n/2)-1}((1/q)^{n/2}+(1/b)^{(n/2)-1})}}, \tag{36}$$

$$p(\textit{convex region is figure}|I_4) = 0.5 + \frac{\frac{1}{1+(1/q)^{n/2}} - 0.5}{1 + (1/z)(b/a)^{(n/2)-1}}. \tag{37}$$

3. Results

We developed a Bayesian observer that makes figure–ground judgments for alternating convex-concave region displays. The observer's convex-region-is-figure response probability depends upon its knowledge of natural scene statistics, specifically, its expectation that objects tend to be convex (represented by the q parameter) and that backgrounds tend to be homogeneously colored (represented by the a parameter). The observer responds that the convex regions in the display are foreground objects (figures) stochastically, with a probability governed by equation (1), whose solution for each of the four display types (Fig. 3A) is given by equations (34)–(37), Here we compare the performance of the Bayesian observer to that of human subjects tested with the same displays (Peterson and Salvagio, 2008).

3.1. Parameter Estimation

To determine the optimal values for the Bayesian observer's free parameters, we compared the model's performance with a range of parameter settings to that of the human subjects tested by Peterson and Salvagio (2008). For each (q, a) parameter setting, we calculated the probability that the observer would reproduce average human performance on the series of Bernoulli trials that constituted the set of experiments reported by Peterson and Salvagio (2008).

This joint (q, a) likelihood function is

$$p(\mathit{human\ data}|q, a) \propto \prod_i \prod_n p(q, a)_{in}^{h_{in}} \big(1 - p(q, a)_{in}\big)^{t_{in}-h_{in}}, \tag{38}$$

where i indexes display type (see Fig. 3A); n is the number of regions in the display; $p(q, a)$ is the observer's response probability that the convex region is figure (i.e., equations (34)–(37)); and h is the number of corresponding 'convex region is figure' responses made by the average human subject (mean proportion of human responses that the convex region is figure, multiplied by the number of trials, t, presented to the subject by Peterson and Salvagio, 2008: $t = 64$ for subjects tested on display type 1; $t = 72$ for subjects tested on the other display types). In calculating $p(q, a)$, we set N_c to the size of the color repertoire of the particular experiment, under the assumption that human subjects quickly learned (e.g., in practice trials) and incorporated that number into their perceptual inference (see Note 5).

We evaluated equation (38) over a wide range of (q, a) settings and took the maximum likelihood estimate as the observer's optimal parameter settings:

$$(q, a)_{best} = \underset{(q,a)}{\mathrm{Argmax}}\, p(\mathit{human\ data}|q, a). \tag{39}$$

The maximum likelihood estimate was $(q = 1.77, a = 1.44)$. The observer with these values expects 1.77 times as many objects to resemble the convex shapes as to resemble the concave shapes used by Peterson and Salvagio (2008). The observer further considers that two consecutive ground regions have 1.44 times the probability of matching in color as two objects do.

These somewhat subtle biases are sufficient to yield figure–ground perception that fits the human data reasonably well. Figure 4 shows the Bayesian observer's expected performance, with $(q = 1.77, a = 1.44)$, plotted together with the average human performance from Peterson and Salvagio (2008), for each display type and number of regions. Note that, like the average human subject, the Bayesian observer shows a clear convexity context effect for display types 1 and 2, but not for display types 3 and 4.

3.2. Understanding the Convexity Context Effect

Why does a convexity context effect occur for display types 1 and 2 but not for types 3 and 4? Qualitatively, display types 1 and 2 are consistent with convex fig-

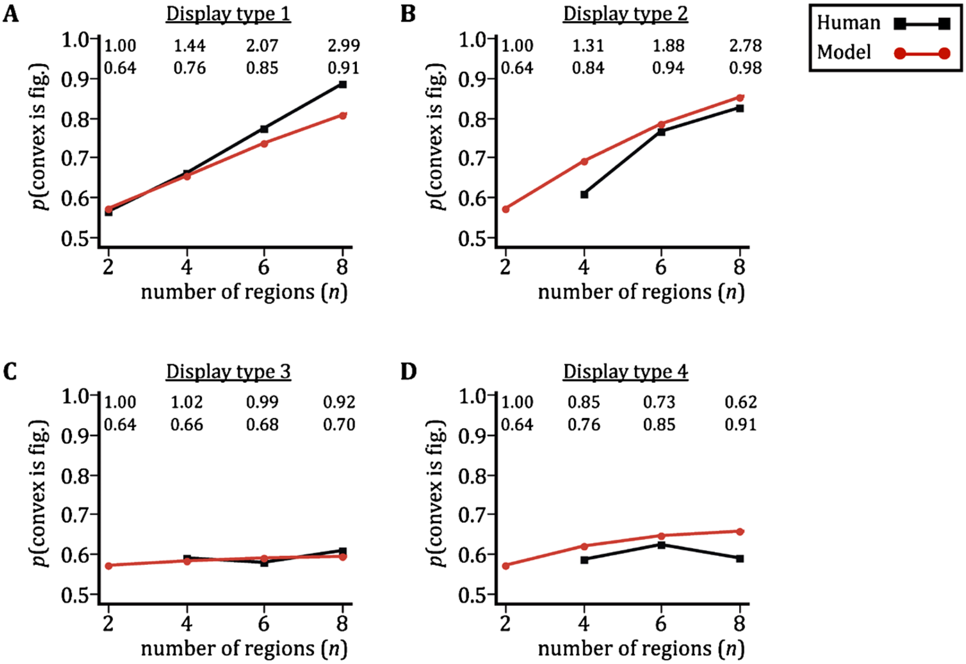

Figure 4. Mean performance of the Bayesian observer and human subjects on the four display types, as a function of number of display regions. (A) Display type 1: homogeneous convex, homogeneous concave. (B) Display type 2: heterogeneous convex, homogeneous concave. (C) Display type 3: homogeneous convex, heterogeneous concave. (D) Display type 4: heterogeneous convex, heterogeneous concave. Red curves and circles: Bayesian observer's performance (parameter settings: $q = 1.77$, $a = 1.44$). Black curves and squares: Average of human performance data from Peterson and Salvagio (2008). The upper row of numbers in each panel lists the color likelihood ratios for the corresponding display type and region number; the lower row of numbers lists the probability of Hypothesis 1 under the 3D scene interpretation. With increasing region number, the color likelihood ratio progressively favors the 3D interpretation (color likelihood ratios > 1) for display types 1 and 2 only. This figure is published in color in the online version.

ures against a homogeneously colored background, and as the number of regions increases, this interpretation become progressively more plausible; unlike display types 1 and 2, types 3 and 4 do not admit this interpretation.

Quantitatively, recall that equation (1) — and equivalently, equation (33) — governs the probability with which the Bayesian observer responds that the convex regions are figures. The convexity context effect occurs when that probability increases with increasing numbers of regions, n. The two factors in equation (33) that depend on n are the color likelihood ratio for the 3D relative to the 2D scene interpretation, and the posterior probability of Hypothesis 1 under the 3D interpretation. Whether the convexity context effect occurs, then, depends on the behavior of those two factors as n increases.

Table 1 shows for each display type the limiting behavior, as n approaches infinity, of both the color likelihood ratio (equations (29)–(32)) and the posterior

Table 1.
Limiting behavior, as the number of regions in each display type (column 1) approaches infinity, of the 3D *vs.* 2D scene interpretation color likelihood ratio (column 2), the posterior probability under the 3D interpretation that the convex regions are figures (column 3), and the frequency with which the observer will respond that the convex regions are figures (column 4)

$Display_i$	$\lim_{n\to\infty} clr_i$	$\lim_{n\to\infty} p(H_1\|3D, I_i)$	$\lim_{n\to\infty}(0.5 + \frac{p(H_1\|3D,I_i)-0.5}{1+(1/clr)})$
1	∞	1	1
2	∞	1	1
3	0, if $q > a$	1, if $q > b$	0.5, if $q > a$
	∞, if $q < a$	0, if $q < b$	0, if $q < a$
4	0	1	0.5

probability for Hypothesis 1 under the 3D scene interpretation (equations (9), (12), (16), (19)). For display types 1 and 2, both factors grow with n; thus, as the number of regions increases, the observer becomes increasingly confident that the scene is 3D and that, given the scene is 3D, the convex regions are figures. Therefore, in the limit that n approaches infinity, the observer's convex-region-is-figure response probability approaches one.

For display type 3, in contrast, the behavior of the two factors as n approaches infinity depends on the values of q and a, with the result that the observer's convex-region-is-figure response probability approaches either 0.5 (when $q > a$) or zero (when $q < a$). This interesting divergent behavior reflects the opponent-nature of the observer's two beliefs, as applied to display type 3. On the one hand, the belief that objects tend to be convex (q-parameter) biases the observer to perceive the convex regions as figures against a heterogeneously colored background. On the other hand, the belief that backgrounds tend not to change color (a-parameter) biases the observer to perceive the concave regions as figures against a homogeneously colored background. Under these circumstances, the observer's perception depends on which parameter, q or a, is largest. If q is largest, the observer concludes that the scene is in fact two-dimensional (color likelihood ratio approaches zero), because a 3D scene with a heterogeneously colored background is less probable than a 2D scene. The observer thus responds randomly that the convex region is figure (response probability 0.5). If a is largest, the observer concludes that the scene is three-dimensional with concave figures (color likelihood ratio approaches infinity), and never responds that the convex region is figure (response probability zero).

Finally, for display type 4, as n approaches infinity, although the posterior probability of Hypothesis 1 under the 3D interpretation grows towards one, the color likelihood ratio diminishes towards zero because the background is heterogeneously colored under both 3D hypotheses. Convinced that the scene is two-dimensional, the observer thus responds randomly that the convex region is figure on 50% of trials.

3.3. Mean Performance and Variability

Because the observer responds stochastically by posterior sampling (see equation (1)), its responses show binomial variability. Thus, if repeatedly presented with t_{in} trials of display type i and number of regions n, for which the corresponding solution of equation (1) (see equations (34)–(37)) is probability p_{in}, then the number of times the observer responds that the convex region is figure will be distributed with expected mean and standard deviation:

$$\begin{aligned} \mu_{in} &= t_{in} p_{in}, \\ \sigma_{in} &= \sqrt{t_{in} p_{in}(1 - p_{in})}. \end{aligned} \tag{40}$$

Using the same t_{in} as Peterson and Salvagio (2008), we compared the Bayesian observer's performance standard deviation, expressed as proportion of trials (obtained by dividing sigma in equation (40) by t_{in}) to the (between-subject) standard deviation of the human data reported by Peterson and Salvagio.

Figure 5 plots the Bayesian observer's mean performance curves along with its ±1SD performance variability. Also plotted is the mean performance ±1SD of the human subjects' data. In order to illustrate long-range trends in the Bayesian observer's performance, the plots extend to hypothetical displays with up to 14 regions, although humans were tested on displays of up to 8 regions only. The Bayesian observer's performance falls squarely within the range of performance of the human subjects, and does not differ significantly from that of the average human subject (one-tailed binomial p-value comparing model to average human performance: $p > 0.05$ for every display type and number of regions at which human data were collected). Interestingly, the Bayesian observer's (within-subject) variability is less than the variability observed in the human data. We expect that the variability in the human data owes to two sources: within-subject noise due to the stochastic nature of each subject's responses, and some between-subject differences in parameter settings (see Note 6).

3.4. Range of Parameter Settings Yielding a Good Fit

Up to this point, we have considered the performance of the Bayesian observer with optimal parameter settings; we now investigate the performance of the observer over a range of parameter settings.

Figure 6 plots, for (q, a) values ranging from (1.0, 1.0) to (2.5, 2.0), the probability that the Bayesian observer would perform equally to the average human subject on the four display types (equation (38)). The figure reveals that a Bayesian observer lacking the object convexity and/or ground color homogeneity expectations ($q = 1$, $a = 1$) yields a very poor match to the human perceptual data. Good fits require that $q > a > 1$. The model fits the human data well over a range of (q, a) settings close to (1.77, 1.44), the maximum likelihood estimate.

Figure 7 shows the performance of the observer with four parameter settings that result in half-maximum likelihoods. Note the generally consistent nature of

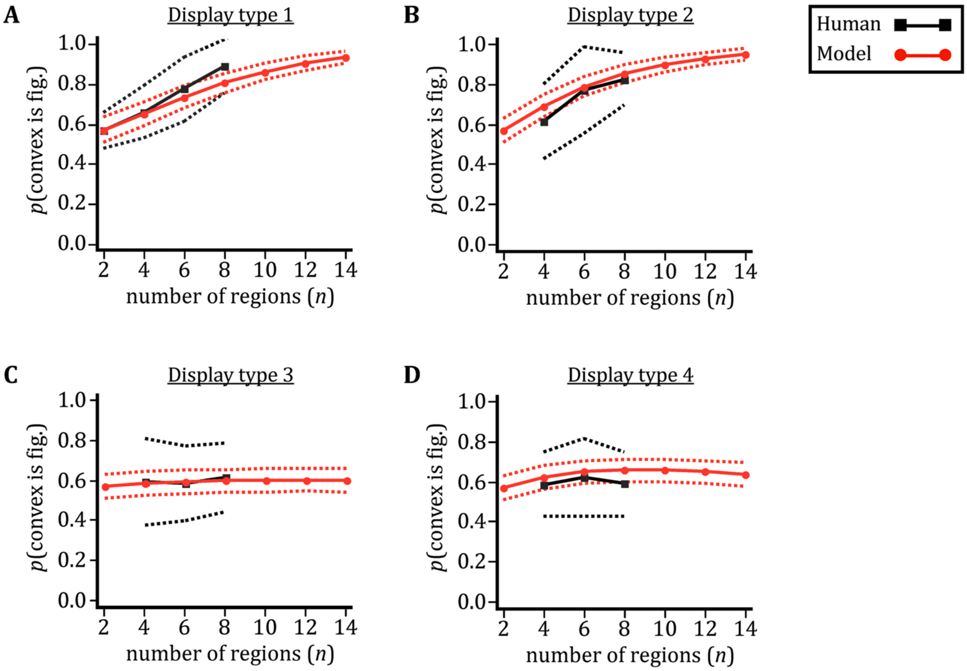

Figure 5. Mean performance and variability of the Bayesian observer and of human subjects. (A) Display type 1: homogeneous convex, homogeneous concave. (B) Display type 2: heterogeneous convex, homogeneous concave. (C) Display type 3: homogeneous convex, heterogeneous concave. (D) Display type 4: heterogeneous convex, heterogeneous concave. Red curves and circles: Bayesian observer's mean performance (parameter settings: $q = 1.77$, $a = 1.44$; solid curve) ±1SD (dotted curves), plotted out to 14 regions in order to illustrate long-range performance trends. Black curves and squares: average of human performance data from Peterson and Salvagio (2008) (solid curve) ±1SD (dotted curves).

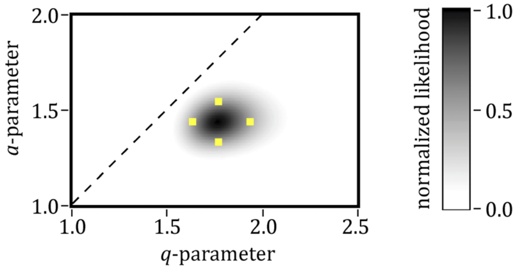

Figure 6. Joint (q, a) likelihood function. For each (q, a) parameter setting, pixel darkness represents the probability that the Bayesian observer would produce the same data as the average of human subjects tested by Peterson and Salvagio (2008) (equation (38)). Scale bar: probability normalized to the maximum likelihood setting: $q = 1.77$, $a = 1.44$. Yellow squares indicate four parameter settings with half-maximum-likelihoods: $(q = 1.77, a = 1.55)$, $(q = 1.77, a = 1.33)$, $(q = 1.63, a = 1.44)$ and $(q = 1.93, a = 1.44)$. Dashed line: $q = a$.

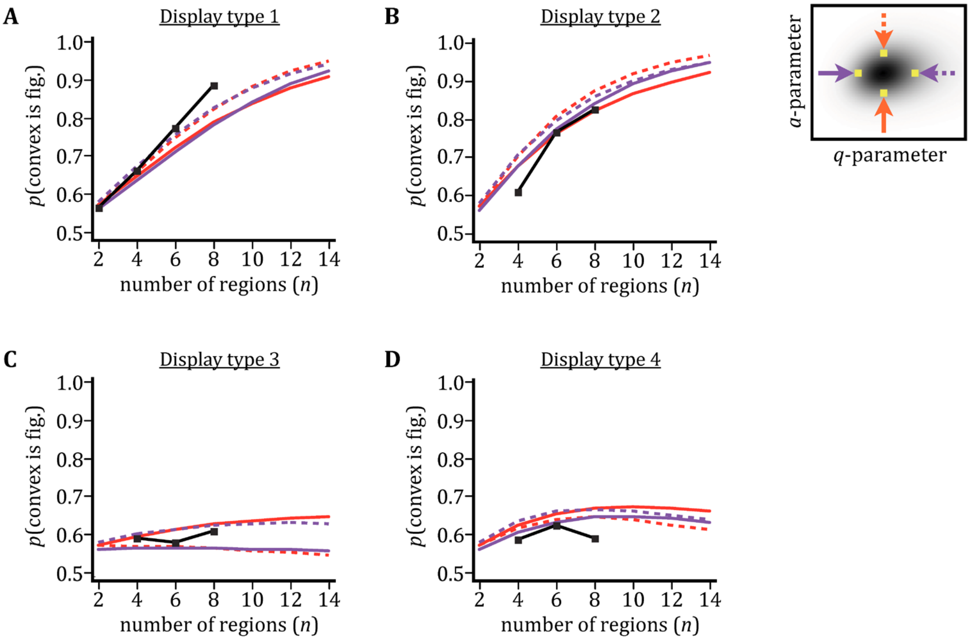

Figure 7. Effect of (q, a) variation on Bayesian observer's performance. (A) Display type 1: homogeneous convex, homogeneous concave, (B) Display type 2: heterogeneous convex, homogeneous concave. (C) Display type 3: homogeneous convex, heterogeneous concave. (D) Display type 4: heterogeneous convex, heterogeneous concave. The colored curves in each plot show the Bayesian observer's performance at the four half-maximum-likelihood parameter settings specified in Fig. 6 (inset: central portion of joint likelihood function from Fig. 6; arrows indicate the color and style of the corresponding performance curves). Black curves and squares: average of human performance data from Peterson and Salvagio (2008).

these curves, all showing strong convexity context effects for display types 1 and 2 but not for display types 3 and 4. Note also that an increase in q consistently elevates the convex-region-is-figure response frequency (compare solid and dotted purple curves), whereas an increase in a exerts this effect with display types 1 and 2, but produces the opposite effect with display types 3 and 4 (compare solid and dotted orange curves). For display type 3, the posterior probability of Hypothesis 1 under the 3D interpretation (equation (16)) drops towards zero with increasing a, as the observer confidently attributes the homogeneously colored convex regions to the background. For display type 4, the color likelihood ratio (equation (32)) drops towards zero with increasing a, as the absence of homogeneously colored alternating regions becomes incompatible with the uniform background expected of a 3D scene.

3.5. Sensitivity of Parameter Estimation to Human Subject Error

We now consider the effect of human subject error on our estimation of the Bayesian observer's parameter values. In any perceptual study, it is plausible that subjects will on occasion lose concentration. Since we have fit the model parameters (q, a) to the performance of human subjects recorded by Peterson and Salvagio (2008), we would like to assess the extent to which our estimates could be corrupted by attention lapses among those subjects. Our strategy is to model the expected effect of attention lapses, then back-adjust the human data to obtain hypothetical lapse-free data on which to perform parameter estimation.

We begin by postulating that the typical subject experiences attention lapses on a randomly occurring proportion, λ, of trials, and that on trials when an attention lapse occurs the subject pushes either response button with equal frequency. Under this scenario, the recorded data (convex-region-is-figure response probability, y) relate to the subject's hypothetical (uncontaminated by lapses) perceptual data, y', according to:

$$y = y'(1 - \lambda) + 0.5\lambda. \tag{41}$$

It follows that:

$$y' = \frac{y - 0.5\lambda}{1 - \lambda}. \tag{42}$$

To explore the extent to which lapses might have affected our parameter estimates, we set λ to 0.1, a value at the upper extreme of what we would consider the plausible range of lapse rates for a typical human subject. We then fit the observer's parameters to the hypothetical lapse-free human data (equation (42)). Compared to the original parameter estimate ($q = 1.77$, $a = 1.44$), the new maximum likelihood estimate ($q = 1.96$, $a = 1.54$) had slightly higher q- and a-values (11 and 7% higher, respectively), resulting in a slightly more pronounced convexity context effect for the observer viewing display types 1 and 2 (Fig. 8). We conclude that our parameter estimates are reasonably robust against the effects of occasional attention lapses.

3.6. Sensitivity of Parameter Estimation to Value of Model Constants

Finally, we explore the effect on our (q, a) parameter estimation of varying two factors that up to this point we have treated as constants: N_c and z. We have set the Bayesian observer's N_c for each display type to the number of colors used in that display type by Peterson and Salvagio (2008). This procedure presupposes that human subjects learned (i.e., during practice and perhaps early on in the experiment) the number of colors in the displays they were viewing, and set their N_c accordingly; we expect that this learning occurred over just a few trials, because the number of colors was small. An alternative possibility, however, is that human subjects instead set N_c in accord with their visual experience obtained outside the laboratory setting. In such case, subjects might have applied a single (and possibly large) value of N_c to all display types. To explore the effect that this alternative N_c setting would have on our estimate of the observer's parameters, we obtained maximum likelihood (q, a) estimates for a series of different N_c values. These parameter estimates deviated only slightly from our original estimates ($q = 1.77$,

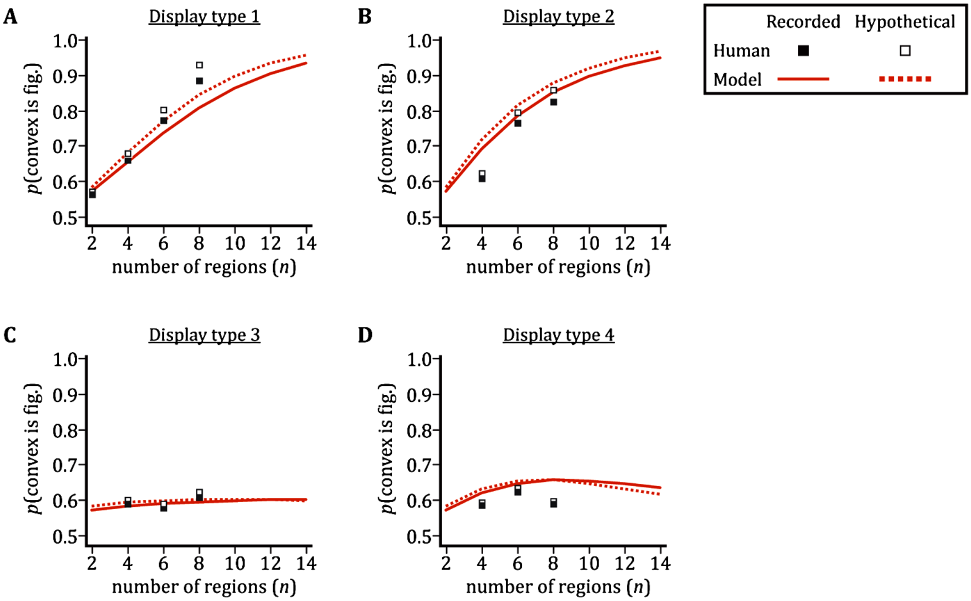

Figure 8. Exploration of the effect of attention lapses on parameter estimates. Open squares: hypothetical lapse-free data of a human subject that, combined with a 10% lapse rate, would give rise to the average data observed by Peterson and Salvagio (2008) (filled squares). The expected effect of lapses is to shift the convex-region-is-figure response frequency towards 50%. Thus, the hypothetical data lie above the observed data. Solid red curve: Bayesian observer's performance with $q = 1.77$, $a = 1.44$. Dotted red curve: Bayesian observer's performance with parameters fit to the hypothetical data: $q = 1.96$, $a = 1.54$. Panels A–D: display types 1–4. This figure is published in color in the online version.

$a = 1.44$) (Fig. 9A). Thus, our (q, a) estimates are rather robust against variation in the value of N_c. Similarly, we obtained the maximum likelihood (q, a) estimates that would have resulted if z, a factor set to 1 in the model (see equation (24)), were set instead to values ranging from 0.5 to 2.0. Again, the (q, a) estimates did not deviate markedly from the original estimates (Fig. 9B), with the exception of a growing upward trend as z dropped to low values.

4. Discussion

We have shown that a Bayesian observer model that incorporates two plausible expectations regarding the statistics of natural scenes replicates the convexity context effects demonstrated by Peterson and Salvagio (2008). The Bayesian observer interprets ambiguous visual images by bringing to bear the expectations that objects tend to be convex and that backgrounds tend (more than foreground objects) to be homogeneously colored. The results support the hypothesis that human vision achieves figure–ground perception by interpreting sensory inputs in light of these and other expected environmental regularities.

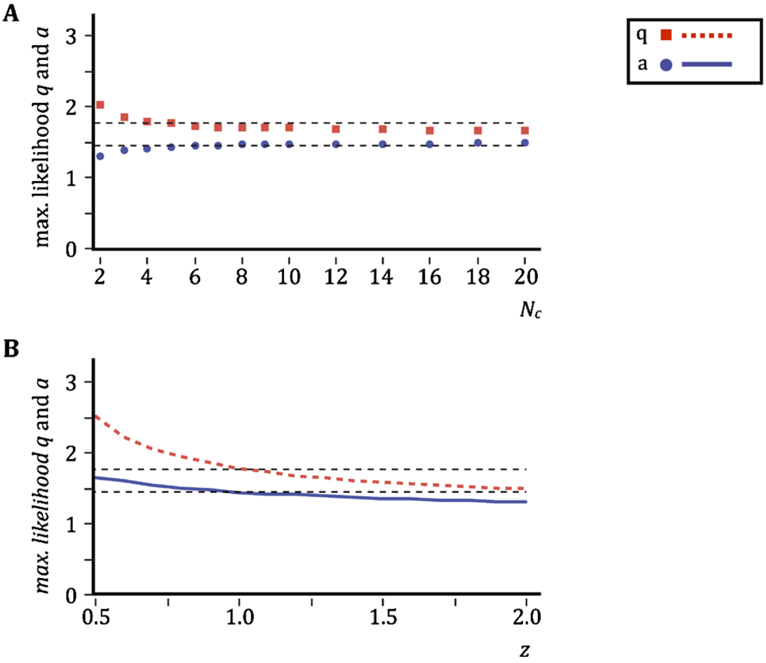

Figure 9. Effect of varying factors N_C and z. (A) Maximum likelihood (q, a) parameter settings when N_C takes on values from 2 to 20, applied by the Bayesian observer uniformly to the four display types. (B) Maximum likelihood (q, a) parameter settings when z takes on values from 0.5 to 2.0. Horizontal dotted lines in both panels: upper line, $q = 1.77$; lower line, $a = 1.44$. This figure is published in color in the online version.

4.1. Comparison to Previous Work

Several studies support the view that visual perception incorporates an object convexity expectation, along with other expectations that presumably reflect the statistics of natural scenes (Adams and Mamassian, 2004; Burge *et al.*, 2010; Fowlkes *et al.*, 2007; Langer and Bülthoff, 2001; Mamassian and Landy, 1998; O'Shea *et al.*, 2010). For instance, using computer-generated displays of complex surfaces, Langer and Bülthoff (2001) showed that subjects' perception of shape from shading was consistent with the incorporation of priors for object convexity, light-from-above, and viewpoint-from-above. Mamassian and Landy (1998) showed that a Bayesian observer replicated human perception of otherwise ambiguous line drawings, provided the observer incorporates three prior constraints: the expectation that surfaces tend to be convex, tend to be viewed from above, and tend to have contours that follow the principal lines of curvature in the drawings. Burge *et al.* (2010) showed in natural scenes a statistically convincing tendency for convex regions to be closer to the viewer than adjacent concave regions, suggesting that most objects are convex. They showed that human subjects perceive consistently with a Bayesian observer that incorporates an understanding of this association, optimally combining depth information obtained from convexity and binocular disparity cues (see also Burge *et al.*, 2005). The present study complements these others by demonstrating that priors for object convexity and background color homogeneity are sufficient to replicate observed context effects in figure–ground perception.

To the extent that our Bayesian observer provides an accurate account of figure–ground perception, the parameter values that result in good fits to the human data should show reasonable agreement with the statistical properties of natural scenes. With respect to the q-parameter, Fowlkes *et al.* (2007) studied convex and concave regions centered on 50 000 contour points labeled from 200 images of natural scenes. They found (Fig. 4 of Fowlkes *et al.*, 2007) that the more convex region was figure (as determined by consensus among human observers) in 60% of cases. Burge *et al.* (2010) reported the depth distributions of the convex *vs.* concave sides of approximately 450 000 contour points taken from images of natural scenes. Integrating across these data (Fig. 3(b) of Burge *et al.*, 2010), we find that 65% of regions nearer to the observer (i.e., figures) were convex (see Note 7). If the convex and concave shapes used by Peterson and Salvagio (2008) are equally representative of convex and concave real-world objects, then our Bayesian observer's best-fit q value, 1.77, equates to the belief that 64% of objects are convex, a value in rather good agreement with those provided by Fowlkes *et al.* (2007) and Burge *et al.* (2010) (see Note 8). With respect to the a-parameter, natural scene statistics for ground color homogeneity have not, to our knowledge, been reported. Our model is at present under-constrained in this regard, among others.

4.2. Testable Predictions

In its current form, our Bayesian observer model makes several predictions regarding the performance of human subjects viewing alternating convex-concave region displays such as those in Fig. 3. These predictions will serve as the basis for future empirical studies. In Peterson and Salvagio (2008), each subject was tested only on a single display type of a single number of regions, so the data from that study speak to across-subject, rather than within-subject trends. The model predicts that, if tested in a repeated-measures experiment on multiple display types and region numbers, an individual subject's convex-region-is-figure response frequency will conform to the equations of the Bayesian observer, with a fixed (q, a) setting applying to that subject for all conditions (see Note 9). A further testable prediction is that, if asked whether the entire display depicts a 3D scene or a 2D scene, subjects' response frequencies will conform to the posterior probabilities described by equations (22), with the color likelihood ratios described by equations (29)–(32). Indeed, the model predicts that, if asked to respond both whether the convex regions are figures and whether the scene is 3D, an individual subject's answers to those two questions will show trial-by-trial correspondence, as predicted by equation (1). Finally, a prediction from the model is that future studies of color statistics in natural scenes will find a value for the a-parameter similar to the best-fit value reported here.

4.3. Limitations and Extensions

We have purposefully put forth the simplest model we could imagine to replicate the convexity context effects described by Peterson and Salvagio (2008). We expect the model, in its present form resting on several simplifying assumptions and limited in scope to the perception of alternating convex-concave region displays, to provide a scaffold for more sophisticated models to come. Here, we discuss some simplifying features of our model, and we consider means by which the model could be modified and extended.

One simplifying feature of our Bayesian observer is that it responds stochastically by sampling from its posterior distributions (equation (1)). This feature is attractive in the context of the current model, because it endows the observer with stochastic responses characteristic of human figure–ground perception, without the addition of free parameters that would otherwise be needed to elicit stochastic responses through image or neural noise (Mamassian *et al.*, 2002). Posterior sampling does not maximize expected utility under standard cost functions; in this sense, it is a suboptimal strategy. Nevertheless, a growing body of research has begun to consider the advantages to the organism of posterior sampling (Gershman *et al.*, 2012; Moreno-Bote *et al.*, 2011) and to identify perceptual tasks in which humans apparently follow this strategy (Battaglia *et al.*, 2011; Moreno-Bote *et al.*, 2011; Vul *et al.*, 2009; Wozny *et al.*, 2010). Empirical studies will be needed to determine whether humans viewing the displays considered here (Figs 2 and 3) are engaging in posterior sampling or are using another strategy. Although we consider posterior sampling to be an attractive simplifying feature of our model, future extensions could implement a maximum *a posteriori* readout, coupled with internal noise to produce stochastic responses.

Our Bayesian observer incorporates several simplifying assumptions regarding scene statistics that are not at present constrained by empirical scene studies. One of these assumptions is that the generative model for ground color is a Markov chain, the probability assigned to the color of a ground region depending only on the color of the preceding ground region (see Note 10) (see equation (3)). A more sophisticated generative model might encode dependencies across multiple ground regions. Implicit to the observer's generative model is the expectation that foreground objects tend to be smaller than background structures (so that the same background structure tends to extend behind two or more foreground objects). A more sophisticated generative model might encode this expectation explicitly, incorporating a distribution for the dimensions of ground structures relative to the dimensions of figures.

Our Bayesian observer assumes that a figure can take on any color, with probability unrelated to the colors of other figures in the scene. One small but interesting discrepancy between the average human data and the performance of the Bayesian observer suggests that this color independence assumption may need to be revised: the human data show somewhat (non-significantly) greater convex-region-is-figure response probabilities for display type 1 than for display type 2, whereas the Bayesian observer shows a trend in the opposite direction (Fig. 4A, B). In display type 2, the color heterogeneity of the convex regions argues against their interpretation as ground regions; consequently, for each number of regions (greater than 2) in display type 2, the posterior probability of Hypothesis 1 under the 3D scene interpretation (equation (12)) is slightly larger than the corresponding posterior probability calculated for display type 1 (equation (9)). The effect of this trend on the observer's performance is only partially offset by the fact that display type 2 yields lower color likelihood ratios than does display type 1 (compare equations (29) and (30)) (see Note 11). Thus, the Bayesian observer responds more frequently that the convex region is figure for display type 2. The opposite trend shown in the average human data may be due simply to random variability (noise) in the data. However, an alternative possibility is that human observers consider figures to have a slight tendency to share the same color, as do the convex regions in display type 1 but not display type 2. To explore this possibility, the model could be extended to incorporate an expectation for color dependency across figures.

Our Bayesian observer assumes that the probability of encountering a 3D scene with the shapes shown (Figs 2 and 3) is equivalent to the probability of encountering a 2D scene with those shapes (i.e., $z = 1$). Consequently, if it were to consider only the shapes of the regions, and not their colors, the observer would favor neither the 3D nor the 2D scene interpretation (see Note 12). A z-value of approximately one may be reasonable for the nondescript shapes considered here, but we certainly do not expect this value to hold generally. The shapes in some images, for instance photographs of scenes containing trees, houses, or other familiar objects, will clearly favor a 3D interpretation ($z > 1$), whereas those in other images (e.g., Fig. 10) will favor a 2D interpretation ($z < 1$). To equip the Bayesian observer to handle a range of natural images, a necessary first step would be to endow the observer with the knowledge of scene statistics required to set z appropriately for each image (equation (24)) (see Note 13).

Finally, in making figure–ground judgments, our Bayesian observer considers only the shapes and colors of regions in the visual image. A more sophisticated model, applicable to a greater variety of scenes, would of course make use of additional depth cues, such as linear perspective and binocular disparity.

5. Conclusion — Towards a Bayesian Approach to Perception

The results of the present study and others (Burge *et al.*, 2010; Langer and Bülthoff, 2001; Mamassian and Landy, 1998) suggest that the Gestalt convexity cue is productively viewed as reflecting a Bayesian perceptual bias grounded in natural scene statistics. Together with others (Elder and Goldberg, 2002; Geisler and Diehl, 2003; Knill *et al.*, 1996; Mamassian, 2006), we suggest that all Gestalt perceptual principles — and indeed perception in general — will be most usefully investigated within the Bayesian framework. Indeed, the explanatory power of the Bayesian approach extends beyond the visual domain into nonvisual perception (Goldreich,

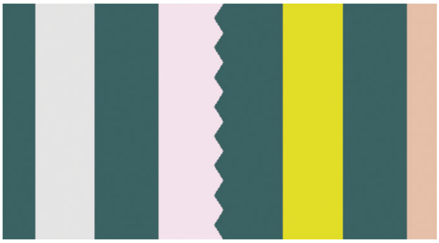

Figure 10. A display with alternating homogeneously-heterogeneously colored regions that does not evoke figure–ground context effects. Asked when viewing displays such as this to indicate whether the region to the left or right of the central zigzag edge was a *figure*, subjects responded at near-chance (50%) levels (Peterson and Salvagio, 2008, Experiment 4). Although the homogeneously colored regions in this display type are consistent with a uniform-background 3D scene interpretation, we suspect that the repeating rectangular shapes are overwhelmingly more compatible with a 2D scene interpretation. In other words, for this display type, we suspect that human observers would estimate $p(shapes|3D)/p(shapes|2D) \ll 1$. A z-value close to zero (equation (24)), strongly favoring the 2D scene interpretation, would then drive the observer's response frequency towards 50% (see equation (33)).

2007; Kording *et al.*, 2004; Norris and McQueen, 2008) and multisensory integration (Angelaki *et al.*, 2009; Deneve and Pouget, 2004; Ernst, 2006; Ma and Pouget, 2008).

Acknowledgements

We thank Elizabeth Salvagio for providing display images and individual subjects' data, and Joe Austerweil and two anonymous reviewers for their thoughtful and insightful comments on the manuscript. We acknowledge support from the Natural Sciences and Engineering Research Council of Canada (DG) and from NSF BCS-0960529 (MAP).

Notes

1. From here on in, we will use the terms 'convex regions' and 'concave regions' to refer respectively to regions with convex parts and regions with concave parts. When two convex parts abut one another, they create a concave cusp. Thus, by a convex region we mean a region that consists of convex parts delimited by concave cusps. The regions in the displays used by Peterson and Salvagio (2008) had approximately 5–18 parts (convexity context effects did not vary with the number of parts).
2. In Peterson and Salvagio (2008), region number was a between-subject factor. The trends described are therefore across-subject trends, not within-subject trends.
3. We assume that the subjects' perception was independent of the task instruction. Peterson and Salvagio (2008) instructed subjects to identify (by button

press) one of the regions as a figure; the subjects did not have the option to respond ‘none’, i.e., that they perceived the display as a flat (2D) surface lacking figures. Nevertheless, we do not believe that this forced-choice procedure compelled the subject to actually perceive a 3D scene.

4. Peterson and Salvagio (2008) used equal numbers of displays in which the leftmost regions were concave and convex. For a display whose leftmost region is concave, our focus in the derivations to follow would simply change from Hypothesis 1 to Hypothesis 2, with results identical to those we describe here for Hypothesis 1.
5. Peterson and Salvagio (2008) used two colors (black and white) for display type 1, five colors (cyan, magenta, yellow, orange, gray) for display types 2 and 3, and four colors for display type 4 (cyan, magenta, yellow, gray). We therefore set the value of N_c for each display type accordingly, when evaluating equations (35)–(37). Note that the value of N_c is in fact irrelevant to the observer’s perception of display type 1, because only q and a (not b) enter into equation (34).
6. Conceivably, one source of between-subject variation in parameter settings is between-subject variation in visual experience. We expect an individual’s parameter settings to reflect the statistics of the visual scenes to which that person has been previously exposed; these experienced scene statistics surely vary somewhat from person to person. A second possible source of between-subject variation in parameter settings is that individuals may differ in the accuracy with which they encode the experienced scene statistics.
7. The frequency histogram (Fig. 3(b)) of Burge *et al.* (2010) discretizes depth into 2-m-wide bins. We used GraphClick v. 2.9 (Arizona Software) to extract these data, and calculated the proportion of figures that were convex and the proportion that were concave by summing the corresponding plots (blue and red, respectively) across all positive-depth bins (i.e., excluding the bin centered on zero depth). Thus, the values (65 and 35%) pertain to figure–ground depth separations of greater than 1 m.
8. Recall (equation (2)) that q is the ratio of p_1 to p_2, where

$$p_1 = p(convex_{P\&S}|object),$$
$$p_2 = p(concave_{P\&S}|object).$$

Here the subscript *P&S* refers to the particular convex and concave shapes of the regions used by Peterson and Salvagio (2008). We may rewrite these probabilities as

$$p_1 = p(convex_{P\&S}|convex, object)\,p(convex|object),$$
$$p_2 = p(concave_{P\&S}|concave, object)\,p(concave|object).$$

If in the displays used by Peterson and Salvagio the convex shapes were as representative of convex objects as the concave shapes were of concave objects, then

$$p(convex_{P\&S}|convex, object) = p(concave_{P\&S}|concave, object)$$

and q becomes

$$q = \frac{p(convex|object)}{p(concave|object)}.$$

Assuming further that any object is either convex or concave,

$$p(convex|object) + p(concave|object) = 1$$

we have that

$$p(convex|object) = \frac{q}{1+q},$$
$$p(concave|object) = \frac{1}{1+q}.$$

For $q = 1.77$, these probabilities are 64 and 36%, respectively.

9. This prediction rests upon the assumption that there are no contaminating effects of priming, response bias, and/or practice in this hypothetical repeated-measures experiment. In addition, such experiments would best be performed with scenes all containing the same numbers of colors, to avoid possible complications associated with changes to the observer's N_c.
10. Our observer also ignored several idiosyncratic features of the actual displays used by Peterson and Salvagio (2008), including the facts that alternating regions varied in luminance and that some colors were not placed randomly but rather occupied fixed positions within the displays.
11. For a given number of regions, $p(H_1|shapes, 3D)$, $p(H_2|shapes, 3D)$, and $p(colors|H_1, shapes, 3D)$ are identical for the two display types, but $p(colors|H_2, shapes, 3D)$ is greater for display type 1 than it is for display type 2, because display type 1 is consistent with homogeneously colored convex ground regions. Therefore, (see equation (26)) $p(colors|shapes, 3D)$ is larger for display type 1. This results in a greater color likelihood ratio (equation (23)) for display type 1.
12. If the observer were to consider just the shapes shown, but not their colors, equations (22) (Section 2.5) would reduce to

$$p(3D|1) = \frac{1}{1+(1/z)},$$
$$p(2D|I) = \frac{1}{z+1},$$

where z is as defined in equation (24). Assumption 4 (Section 2.3) is equivalent to the statement that $z = 1$. Therefore, $p(3D|I) = p(2D|I) = 0.5$ in the absence of color information.

13. The observer would also need to set the q-parameter to a value appropriate for the shapes in the images. Recall that q is the probability that an object would resemble the convex regions in the display divided by the probability that it would resemble the concave ones. The value of q, then, is specific to the actual shapes in the image. We have every reason to suspect that q would be greater for a person viewing a photograph of a series of faces (regions with convex shapes) and the spaces between them (regions with concave shapes) than for the same person viewing one of the displays shown in Fig. 3A. In such cases, $q \gg 1$.

References

Adams, W. J. and Mamassian, P. (2004). Bayesian combination of ambiguous shape cues, *J. Vision* **4**, 921–929.

Angelaki, D. E., Klier, E. M. and Snyder, L. H. (2009). A vestibular sensation: probabilistic approaches to spatial perception, *Neuron* **64**, 448–461.

Battaglia, P. W., Kersten, D. and Schrater, P. R. (2011). How haptic size sensations improve distance perception, *PLoS Comput. Biol.* **7**, e1002080.

Burge, J., Fowlkes, C. C. and Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception, *J. Neurosci.* **30**, 7269–7280.

Burge, J., Peterson, M. A. and Palmer, S. E. (2005). Ordinal configural cues combine with metric disparity in depth perception, *J. Vision* **5**, 534–542.

Deneve, S. and Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links, *J. Physiol. Paris* **98**, 249–258.

Elder, J. H. and Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours, *J. Vision* **2**, 324–353.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration, in: *Human Body Perception From The Inside Out*, G. Knoblich, I. M. Thornton, M. Grosjean and M. Shiffrar (Eds), Ch. 6, pp. 105–131. Oxford University Press, New York, NY, USA.

Fine, I., MacLeod, D. I. and Boynton, G. M. (2003). Surface segmentation based on the luminance and color statistics of natural scenes, *J. Optic. Soc. Amer. A Opt. Image. Sci. Vis.* **20**, 1283–1291.

Fowlkes, C. C., Martin, D. R. and Malik, J. (2007). Local figure–ground cues are valid for natural images, *J. Vision* **7**, 1–9.

Geisler, W. S. and Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems, *Cognit. Sci.* **27**, 379–402.

Gershman, S. J., Vul, E. and Tenenbaum, J. B. (2012). Multistability and perceptual inference, *Neural Comput.* **24**, 1–24.

Goldreich, D. (2007). A Bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions, *PLoS One* **2**, e333.

Hulleman, J. and Humphreys, G. W. (2004). A new cue to figure–ground coding: top–bottom polarity, *Vision Research* **44**, 2779–2791.

Ing, A. D., Wilson, J. A. and Geisler, W. S. (2010). Region grouping in natural foliage scenes: image statistics and human performance, *J. Vision* **10**, 11–19.

Kanizsa, G. and Gerbino, W. (1976). Convexity and symmetry in figure–ground organization, in: *Vision and Artifact*, M. Henle (Ed.), pp. 25–32. Springer, New York, NY, USA.

Kennedy, J. M. (1974). *A Psychology of Picture Perception*. Jossey-Bass, San Francisco, USA.

Knill, D. C., Kersten, D. and Mamassian, P. (1996). Implications of a Bayesian formulation of visual information for processing for psychophysics, in: *Perception as Bayesian Inference*, D. C. Knill and W. Richards (Eds), pp. 239–286, Cambridge University Press, New York, NY, USA.

Kording, K. P., Ku, S. P. and Wolpert, D. M. (2004). Bayesian integration in force estimation, *J. Neurophysiol.* **92**, 3161–3165.

Langer, M. S. and Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading, *Perception* **30**, 403–410.

Ma, W. J. and Pouget, A. (2008). Linking neurons to behavior in multisensory perception: a computational review, *Brain Research* **1242**, 4–12.

Mamassian, P. (2006). Bayesian inference of form and shape, *Prog. Brain. Res.* **154**, 265–270.

Mamassian, P. and Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings, *Vision Research* **38**, 2817–2832.

Mamassian, P., Landy, M. and Maloney, L. T. (2002). Bayesian modelling of visual perception, in: *Probabilistic Models of the Brain: Perception and Neural Function*, R. P. N. Rao, B. A. Olshausen and M. S. Lewicki (Eds), pp. 13–36. MIT Press, Cambridge, MA, USA.

Moreno-Bote, R., Knill, D. C. and Pouget, A. (2011). Bayesian sampling in visual perception, *Proc. Natl Acad. Sci. USA* **108**, 12491–12496.

Norris, D. and Mcqueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition, *Psychol. Rev.* **115**, 357–395.

O'Shea, J. P., Agrawala, M. and Banks, M. S. (2010). The influence of shape cues on the perception of lighting direction, *J. Vision* **10**, 1–21.

Palmer, S. E. and Brooks, J. L. (2008). Edge-region grouping in figure–ground organization and depth perception, *J. Exper. Psychol. Hum. Percept. Perform.* **34**, 1353–1371.

Palmer S. E. and Ghose T. (2008). Extremal edges: a powerful cue to depth perception and figure–ground organization, *Psychol. Sci.* **19**, 77–84.

Peterson, M. A. (2003). On figures, grounds, and varieties of amodal surface completion, in: *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, R. Kimchi, M. Behrmann and C. Olson (Eds), pp. 87–116. LEA, Mahwah, NJ, USA.

Peterson, M. A. and Gibson, B. S. (1994). Must figure–ground organization precede object recognition? An assumption in peril, *Psychol. Sci.* **5**, 253–259.

Peterson, M. A. and Salvagio, E. (2008). Inhibitory competition in figure–ground perception: context and convexity, *J. Vision* **8**, 1–13.

Peterson, M. A., Harvey, E. H. and Weidenbacher, H. L. (1991). Shape recognition input to figure–ground organization: which route counts? *J. Exper. Psychol. Hum. Percept. Perform.* **17**, 1075–1089.

Pizlo, Z. (2001). Perception viewed as an inverse problem, *Vision Research* **41**, 3145–3161.

Vecera, S. P., Vogel, E. K. and Woodman, G. F. (2002). Lower-region: a new cue for figure–ground assignment, *J. Exper. Psychol. Gen.* **131**, 194–205.

Vul, E., Hanus, D. and Kanwisher, N. (2009). Attention as inference: selection is probabilistic; responses are all-or-none samples, *J. Exper. Psychol. Gen.* **138**, 546–560.

Wozny, D. R., Beierholm, U. R. and Shams, L. (2010). Probability matching as a computational strategy used in perception, *PLoS Comput. Biol.* **6**, e1000871.