Introduction

The current paper focuses on combination of information sources and prior knowledge relevant for perceptual grouping, more specifically applied to the integration of proximity and collinearity to a global orientation percept. We analyze grouping of local elements into lines as a problem of contour integration.

Historical and experimental paradigms in grouping

Two experimental paradigms have been popular in the psychophysical investigation of grouping mechanisms. In the contour detection paradigm, dots or Gabor patches which together define a virtual line of a certain length are embedded in a background of otherwise unrelated similar elements (Beck, Rosenfeld, & Ivry, 1989; Chinnis & Uttal, 1974; Field, Hayes, & Hess, 1993; Hansen & Hess, 2006; Kovác & Julesz, 1993; Uttal, 1975; Uttal, Bunnell, & Corwin, 1970). Participants in this line of experiments are to indicate whether or not they were able to discern a contour, for example, in a yes/no or two-alternative forced choice design. Other research groups made use of stimuli in which two or more candidate groupings were simultaneously available (Ben-Av & Sagi, 1995; Claessens & Wagemans, 2005; Gepshtein & Kubovy, 2000, 2005; Kubovy, Holcombe, & Wagemans, 1998; Kubovy & Wagemans, 1995; Kurylo, 1997; Oyama, 1961; Rock & Brosgole, 1964; Zucker, Stevens, & Sander, 1983).

Grouping under these conditions turns out to be a multistable phenomenon: prolonged or repeated presentation of one and the same stimulus potentially evokes different percepts. Note the difference with the contour detection paradigm: while contour detection measures detectability of a single contour in background noise, the multistable grouping paradigm measures relative preference among various candidates that are all valid perceptual solutions.

Foundations of Bayesian integration of grouping cues

The Bayesian framework has been valuable in explaining how information, from one or several sources, is combined with prior knowledge in perceptual
We will now outline the elements of a Bayesian theory which is related to an unobservable state of the world \( X \) observable variable (e.g., Wichmann, 2005). Secondly, assumptions about the \( \text{prior} \) that would otherwise be ambiguous (e.g., Mamassian, Knill, & Kersten, 1998; Willems & Wagemans, 2000). For example, prior constraints solve perceptual situations to an interpretation that is, a priori, more likely to be true. The reliability of a cue can be derived from its likelihood function, and indeed influences perceptual integration (e.g., Bu¨ lthoff & Mallot, 1988; Landy, Maloney, Johnston, & Young, 1995; Rosas, Wichmann, & Wagemans, 2007). Firstly, as in depth cue integration, it can be expected that some information sources are more reliable indicators of the presence of a contour than others. If \( X \) the naive observers follow a less optimal strategy: they meanfully be ordered, the MAP rule is the most advantage of the multistability that arises in a lattice in parametrized dot lattices (Kubovy, 1994) has successfullly been quantified in a “Pure Distance Law” (PDL) with only one free parameter (Gepshtein & Kubovy, 2000; Kubovy et al., 1998; Kubovy & Wagemans, 1995). The Bayesian models of other multistable perceptual phenomena (e.g., Mamassian & Landy, 1998; Triesch, Ballard, & Jacobs, 2002) assume a similar process to play a role in perceptual interpretation, as the variation in sensory measurements (internal noise) is generally not sufficient to explain the switches in perceptual interpretation of one and the same stimulus (Mamassian & Landy, 2001). In fact, together with Mamassian and his colleagues, we will assume that the distribution of perceptual choices directly reflects the probabilities that compose the posterior distribution.

When comparing the evidence for discrete hypotheses, say, \( \xi = \xi_1 \) versus \( \xi = \xi_2 \), it makes sense to consider their posterior odds \( \frac{p(\xi_1|x)}{p(\xi_2|x)} \). It is convenient to express the posterior odds in their logarithmic equivalent:

$$
\log \left( \frac{p(\xi_1|x_1, x_2)}{p(\xi_2|x_1, x_2)} \right) = \log \left( \frac{P(x_1|x_1)P(x_2|\xi_1)}{P(x_1|\xi_2)P(x_2|x_2)} \right)
$$

The quotient of likelihoods, \( \frac{p(x|x)}{p(x|x)} \), is an instance of a comparative evidence index known as the Bayes factor. The log of the Bayes factor, positive if a cue favors the hypothesis \( \xi = \xi_1 \) and negative if it favors the hypothesis \( \xi = \xi_2 \), is also referred to as weight of evidence. Recent evidence shows that the log-odds is a quantity represented by the nervous system in a situation where a decision has to be made in the face of a probabilistic outcome. For instance, the firing rate of LIP neurons in rhesus monkeys is roughly proportional to experimentally manipulated log-odds in a probabilistic classification task (Yang & Shadlen, 2007). As with humans, the rhesus monkeys did not deterministically choose their behavioral decision to match the most likely alternative, which would have optimized their chance of getting a reward.

**The stimuli: Zigzag dot lattices**

Schumann (1900) was the first to use a lattice of squares to illustrate that nearby elements are more likely to be grouped. This stimulus allowed him to present collinear structures in two orientations with different inter-element distances. Later on, experimental psychologists used the two-dimensional lattice structure to work on the quantification of perceptual organization laws (e.g., Oyama, 1961; Rock & Brosigole, 1964). Most of these studies take advantage of the multistability that arises in a lattice in which the inter-element distances along the two main orientations are not too different. Perceptual organization in parameterized dot lattices (Kubovy, 1994) has successfully been quantified in a “Pure Distance Law” (PDL) with only one free parameter (Gepshtein & Kubovy, 2000; Kubovy et al., 1998; Kubovy & Wagemans, 1995).
relevance of lattices for measuring relative strength of different linear organizations explains why they have mainly been used in the context of grouping in the presence of several grouping cues, e.g., proximity and shape or brightness similarity (Ben-Av & Sagi, 1995; Zucker et al., 1983), or proximity and element alignment (Claessens & Wagemans, 2005).

**Dot lattices**

A two-dimensional point lattice is a regular arrangement of points in the plane (see Figure 1). It can be thought of as a tiling of parallelograms. In the current paper, we base the construction of our stimuli on rectangular dot lattices, in which the constituent parallelograms are rectangles. Sides are determined by the vectors, \(a\) and \(b\). Apart from the inter-vector angle (here constrained to 90°), the most important construction parameter of a lattice is the ratio of inter-element distances along the two lattice vectors: \(|b|/|a|\). Our experimental methodology is largely based on the procedure introduced by Kubovy and Wagemans (1995): subjects reported the perceived global organization upon presentation of a dot lattice.

**Operationalization of discollinearity**

Distinguishing two main classes, discollinearity can emerge as a structural feature of the contour, as in a sawtooth pattern, or because of stochastic deviations (such as the small misalignments between different segments of the stem of a plant). To generate stimuli that mimic both types of collinearity perturbations, a zigzag pattern was imposed on the rectangular lattice geometry in two ways (see Figure 1). One stimulus class was generated by applying a fixed displacement along the \(a\)-orientation, with alternating sign on each dot. We will refer to these as fixed zigzag lattices (FZZ). In a second operationalization of discollinearity, stimuli were constructed by displacing the dots by a random amount along the \(a\)-vector. The magnitude of the displacement was determined by the absolute value of a normally distributed random deviate. We will refer to these constructions as stochastic zigzag lattices (SZZ). Both perturbation methods result in a structure with zigzag in the \(b\)-orientation, while dots remain collinear in the \(a\)-orientation.

As in Feldman (1996), we quantify discollinearity by the difference of the angle created by neighboring virtual lines with a straight line, a quantity that is also referred to as the turning angle \(\theta\). In the FZZ lattices, where the zigzag is fixed, there is a straightforward and constant relation between displacement \(d\) and turning angle \(\theta\). Note that row spacing \((b')\) has to undergo a compression to keep the inter-dot distance along \(b\) constant (Figure 1). For fixed zigzag lattices, inter-row spacing and displacement were calculated to match previously chosen levels for \(|b|/|a|\) and \(\theta\).

The technique for the construction of the stochastic zigzag lattices is somewhat more involved. As local displacements are independently drawn from a half-normal distribution, both dot spacing and turning angle are variable across the lattice. Inter-row spacing and the

![Figure 1](http://example.com/figure1.png)

**Figure 1.** Left: Principles of the construction of a regular rectangular lattice. Dots are placed along the orientations and distances according to basis vectors \(a\) and \(b\). In rectangular lattices, the angle between \(a\) and \(b\) is 90°, which constrains the basic parallelogram they form to a rectangle. Middle: Alternatingly displacing the dots left- and rightward by a fixed amount \(d\) defines a fixed zigzag lattice (FZZ) with turning angle \(\theta\) along \(b\). The introduction of the zigzag causes a slight compression of the inter-row distances \(b'\) compared to \(|b|\) in the original lattices. Right: Principles of the construction of a stochastically perturbed zigzag lattice (SZZ). The value of \(d\) is drawn from a half-normal distribution per dot. Therefore the local displacement \(d\), turning angle \(\theta\) and row spacing \(b'\) are variable, which is indicated by assigning location indices (e.g., \(i,j\) indicates \(i^{th}\) column, \(j^{th}\) row).
spread of the displacement distribution were determined adaptively. Average turning angle and dot distance thus approximated the same preestablished levels as for the fixed zigzag lattices.

This operationalization of discollinearity seems to deviate substantially from measures popular in the contour interpolation and detection literature. When contour elements carry a local orientation, as is the case for Gabor patches or line fragments but not for circular dots, the angular relationship of two elements can be characterized by their co-circularity. Dots do not have local orientations, so it is, in principle, not possible to establish a co-circularity measure. Yet, one can reasonably assign a local orientation, namely the tangent to the visual spline connecting the dot to its neighbors. Under any smooth interpolation function, the tangents to the spline at the dot centers will be parallel with the global lattice orientation. The virtual orientations of dot dipoles thus have an angular deviation from co-circularity that exactly equals the turning angle. At least in zigzag patterns, this makes turning angles numerically comparable to discollinearity expressed as deviation from co-circularity.

A Bayesian model of integration of proximity, collinearity, and prior orientation distribution

Two stimulus variables contribute to the probability that a chain of dots is perceived to signal an existing, coherent contour: the distance between the dots, |v|, and the turning angle, \( \theta \). The larger the distance between contour elements, the smaller the likelihood that they belong to the same contour. Similarly, the more \( \theta \) deviates from 0, the smaller the likelihood that the elements belong to the same contour. These principles are not only dictated by common sense, but have also been measured in the study of the statistics of grouping principles in natural scenes (Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001; Sigman, Cecchi, Gilbert, & Magasco, 2001).

The case of grouping cues as independent sources of information

We are now ready for the translation of the general Bayesian model (Equation 3) to the case of independent combination of grouping cues:

\[
\log \left( \frac{P(b \mid d, \theta)}{P(a \mid d, \theta)} \right) = \log \left( \frac{P(d \mid b)}{P(d \mid a)} \right) + \log \left( \frac{P(b)}{P(a)} \right) + \log \left( \frac{P(\theta \mid b)}{P(\theta \mid a)} \right)
\]

This notation expresses how statistical properties of sampled contours are evaluated in order to establish the posterior probabilities, in log-odds ratio, of \( b \) rather than \( a \) being the “true” contour: the posterior odds equals the sum of the evidence that contour elements are to be found at a distance \( d \), if a contour \( v \) does exist, that contour elements are found at a turning angle \( \theta \), added to the prior log-odds ratio of \( b \) rather than \( a \) constituting a contour. Except for the stimulus manipulations, the only difference between \( a \) and \( b \) is their orientation. Thus, a prior for \( v \) represents the probability that a contour is to be found at the orientation of \( v, \theta \).

In order to test whether information sources are treated as being independent does not, in principle, require the further specification of a likelihood function. Based on observed posterior log-odds, one can make an estimate of the log Bayes factor for each level of distance ratio and each level of discollinearity, and perform a statistical test for (absence of) interaction. However, we will also attempt to capture the subjective likelihood function in a more formal way, for comparison with the ecological statistics literature, and in order to test for possible interactions with more statistical power.

Likelihood function and log Bayes factor for proximity

\( P(d \mid v) \) indicates the likelihood that a contour present in the \( v \)-orientation contains image elements with a distance \( d \). It is more likely that two nearby contour elements are connected by the same contour than two elements located further apart (e.g., Geisler et al., 2001). A monotonically decreasing function of distance, bounded to exclude negative values, is needed to capture this qualitative aspect. The likelihood that a subcontour of a given length \( d \) exists in a randomly picked contour is determined by \( P(l > d) \). This probability, the complement of the cumulative distribution function, is called the survival function of \( l \) evaluated in \( d \).

Length as a variable resulting from stochastic processes is modeled in survival analysis, where the exponential distribution as well as the power law (Pareto) distribution are popular and parsimonious choices. At a further point in the study, we will compare the relevance of each for our data. Here we will simply derive the log Bayes factor for each likelihood distribution. The exponential distribution is typical for length between Poisson events. An exponential distribution function of contour lengths can be written as \( p(l) = \lambda e^{-\lambda l} \). The probability that a contour without known length contains a subcontour of length \( d \), as based on the survival function, is \( e^{-\lambda d} \), and therefore, the log Bayes factor is simply

\[
\log \left( \frac{P(d \mid b)}{P(d \mid a)} \right) = -\lambda(|b| - |a|)
\]
is \( p(l) = q e^{-q l} \); the density starts at an offset \( l_0 > 0 \), below which it is 0. The complement of the cumulative density function, assuming that \( d_e \geq d_0 \), is given by \( \left( \frac{d_0}{d_e} \right)^{-q} \). The corresponding Bayes factor is

\[
\log \left( \frac{P(d_b|b)}{P(d_a|a)} \right) = -q \log \left( \frac{|b|}{|a|} \right)
\]  

(6)

**Likelihood function and log Bayes factor for good continuation**

Good continuation is manipulated through the turning angle of successive dot pairs. The more the “knee” formed by a dot triplet deviates from collinearity, the more severe the violation of good continuation, and the less likely that they indeed belong to one and the same contour. The question, then, is: how likely is it that a contour turns with an angle \( \theta \) between sampled contour elements? Feldman and Singh (2005) propose the most generic circular distribution to model the likelihood of a given turning angle, the von Mises density function. The likelihood of a turning angle in a contour is \( e^{\beta \cos(\theta)}(2\pi I_0(\beta))^{-1} \). Hereby we assume that the distribution is centered on 0°, and not shifted as in Feldman and Singh, who specifically discuss the case of closed contour figures. \( I_0(.) \) is a Bessel function needed to ensure that the distribution integrates to 1. \( \beta \) is a concentration parameter: a high \( \beta \) corresponds to a distribution with most of its density mass near the mean. A low \( \beta \) corresponds to a more spread density.

In the experiment presented here, dots are strictly collinear in the \( a \)-orientation. Knowing that \( \theta_a = 0 \), the likelihood equation for the \( a \)-contour candidate simplifies to \( e^{\beta \cos(\theta)}(2\pi I_0(\beta))^{-1} \). After simplification, the log Bayes factor is

\[
\log \left( \frac{P(\theta_b|b)}{P(\theta_a|a)} \right) = \beta (\cos \theta_b - 1)
\]  

(7)

As an alternative to a von Mises based likelihood function, we consider a Laplace distribution. The Laplace distribution is composed of an exponential distribution at the right side and its mirror image at the left side. It is much more peaked than a bell-shaped distribution as the von Mises density. The density function can be calculated as \( 0.5 \beta e^{-\beta |a|} \). Taking into account that in this experiment, \( \theta_a = 0 \), the log Bayes factor for discollinearity under a Laplace distribution is very simple:

\[
\log \left( \frac{\theta_b|b}{\theta_a|a} \right) = -\beta' |\theta_b|
\]  

(8)

If the log Bayes factor is proportional to a power of \(|\theta_b|\) rather than \(|\theta_b|^\gamma\) itself, then a generalized Laplacian distribution, as used by Elder and Goldberg (2002) in their good continuation metrics, is in order. The probability function of the generalized Laplace is proportional to \( e^{-e^{c|\theta|}(\gamma)} \), with \( c = \sqrt{\Gamma(3/\gamma)}/\Gamma(1/\gamma) \). The normalizing constant of this function depends on both \( \gamma \) and \( \sigma \), but for a chosen set of parameters, it does not matter in the log Bayes factor:

\[
\log \left( \frac{\theta_b|b}{\theta_a|a} \right) = - \left( \frac{c}{\gamma} \right) |\theta_b|^{\gamma}
\]  

(9)

**Prior distribution for contour orientations**

In a natural environment, there is considerable variation in the prevalence of contour orientations. Near-horizontal and near-vertical orientations are more frequent than oblique orientations, in urban as well as in natural scenes (Coppola, Purves, McCoy, & Purves, 1998). If \( \phi(\rho_a) \) is the prior density of the orientation of the \( a \)-vector, and \( \phi(\rho_b) \) of the \( b \)-vector, we will write the log prior odds as:

\[
\log \left( \frac{P(b)}{P(a)} \right) = \log \left( \frac{\phi(\rho_b)}{\phi(\rho_a)} \right)
\]  

(10)

In the current experiment, we present only rectangular lattices, and, hence, \( \rho_b \) always equals \( \rho_a + 90^\circ \). We will show later that the application of priors is somewhat more idiosyncratic, subject to qualitative variations across participants. Because this is a more empirical matter, we will discuss the appropriateness of alternative models for the log prior odds in **Results** section.

**Methods**

**Stimulus set**

The \(|b|/|a| \) parameter was manipulated from 4/5 (0.8) to 5/4 (1.250) in five steps along a logarithmic scale (0.8, \( \sqrt{0.8} \approx 0.8944, 1, \sqrt{1.250} \approx 1.118, 1.25 \)). In a crossed-factorial design, the discollinearity \( \theta \) also took five different values, in a linear range from \( 0^\circ \) to \( 20^\circ \) (\( 0^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ \)). \( 0^\circ \) corresponds to collinearity. We did not go beyond \( 20^\circ \) to stay clear from grouping orientations that are not part of the response alternatives. The discollinearity manipulation always applied to the \( b \)-vector. Dots along the \( a \)-vector were always collinear.

Except for dot distance ratio and discollinearity, based on variations in prior orientation probabilities, one can expect that global orientation of the stimulus has an effect on grouping. The ‘oblique effect’ (Appelle, 1972), i.e. a preference of the visual system for cardinal axes, has been...
demonstrated in a large number of phenomena, including grouping in function of texture segregation (Olson & Attneave, 1970). A failure to investigate the role of global orientation would lead to misestimation of the contributions of proximity and discollinearity. To obtain correct estimates for prior probabilities as well as for the likelihood functions, we manipulated stimulus orientation. The orientation of the $a$-vector was set from 0° to 170° in intervals of 10°.

The five levels of distance ratio (0.8, 0.8944, 1, 1.118, 1.25), the five levels of discollinearity (0° to 20° in 5° intervals), the two stimulus types (FZZ and SZZ), and the 18 global orientations (0° to 170° in 10° intervals) combine to 900 different stimuli in a crossed factorial design. Example stimuli are shown in Figure 2.

**Participants**

One author and six naive observers participated in the experiment. The naive subjects were undergraduate and graduate students who participated for course credits. All had normal or corrected-to-normal acuity.

**Procedure**

By means of a paper print of a common lattice inducing a strong grouping percept, subjects were told what we meant by a global grouping, and shown that grouping can occur in more than one orientation for a given stimulus. Then they went through a practice session with a subset of the experimental stimulus set.

The participants were then subjected to four experimental sessions each, separated by at least half a day. Two sessions contained FZZ stimulus trials, the other two SZZ lattices. The order of the sessions was counterbalanced across subjects. Before being confronted with a new stimulus type, participants went through a new practice block.

The experimental part of each session was subdivided in five blocks of 180 trials each. In each session, each of the 450 stimuli per zigzag type (5 distance ratios × 5 discollinearities × 18 orientations) was presented twice in a random order. Subjects typically needed about 45 minutes to go through the 900 trials in a session. Taking the sessions together, the participants saw four presentations of each stimulus, amounting to 72 presentations per distance-discollinearity combination, for each of the two discollinearity types (FZZ and SZZ).

A trial was initiated with the presentation of an empty blue disk in a black surround, a visually induced circular viewing window with a diameter somewhat smaller than the height of the viewable area of the CRT (see Figure 3, left). After 800 ms, a red fixation dot was shown in the center for 500 ms. The lattice presentation lasted 300 ms. The response screen was composed of three blue disks on a black background (see Figure 3, right). Two of these disks were linked to a lattice vector. The corresponding orientation was represented by a straight bisecting line with a superimposed wiggly line. We used this particular layout so as not to bias the subject toward responding the straight ($a$-) or the zigzag ($b$-) vector based on the resemblance to the icons. During the instructions, the participants were told that the wiggly line meant that the dots did not have to be perfectly collinear to constitute a valid organization. A third disk was empty. Participants could use it to indicate that they did not perceive any global grouping at all, or that their organization did not correspond to the options offered in the response screen. After the participant had clicked on one of the disks with a

Figure 2. *Left:* Stimulus examples for the FZZ stimuli. From left to right, $|b|/|a|$ increases from 0.8, over 1, to 1.25. The upper row contains the lattices with $\theta = 20^\circ$, the middle row with $\theta = 10^\circ$, while the lower row shows the control condition without zigzag manipulation ($\theta_b = 0^\circ$). *Right:* Stimulus examples for the SZZ lattices, in the same arrangement.
red dot indicating the position of the mouse pointer, the circular viewing window appeared on the screen, and the trial cycle restarted.

This type of response screen departs from the one used by Kubovy and Wagemans (1995), who included two more orientations in the response options and did not display a blank “wildcard” option. We chose to offer only two orientations in this study, because response confusion is higher with four orientations. Note that lattice orientations other than a or b are less important as potential grouping orientations in rectangular lattice patterns. The blank disk accommodates for the few expected “neither-a-nor-b” groupings. At the same time, the blank option has the function of removing lapse trials, avoiding the estimation of a lapse rate post hoc (as was done by, e.g., Bleumers, De Graef, Verfaillie, & Wagemans, 2008).2

**Results**

First we will present a visualization of the raw response frequencies for the different experimental conditions (Raw data section). Then we discuss the statistical analyses of the model in which cues and prior distributions are combined independently according to a Bayesian strategy (Independent combination of cues section). We focus on parametric models for the influence of cues on grouping (Formalization of likelihood functions and priors section), after which we explore possible deviations from Bayesian
independent combination (Grouping cues and priors interact through global orientation section).

**Raw data**

The raw data consist of the proportions $a$, $b$, and ‘blank’ choices given by each subject for each stimulus. In qualitative terms, the pattern that we expect contains three components:

1. For a distance ratio $|b|/|a| < 1$ the number of $b$-responses is larger than the number of $a$-responses when no discollinearity is involved. The number of $b$-responses declines as the distance ratio increases.
2. Since the discollinearity $\theta$ is only applied to the $b$-vector, a larger $\theta$ will weaken grouping along the $b$-orientation, resulting in fewer $b$-responses.
3. Response frequencies are modulated by global lattice orientation.

As stated earlier, an oblique effect should be interpreted here as a difference between the attractiveness of the (near-) horizontal and the (near-) vertical orientation.

The relative number of $a$, $b$, and blank responses for each combination of distance ratio and discollinearity are shown for two representative subjects in Figure 4. Within one set of axes, the levels of distance ratio increase along the abscissa, while the levels of discollinearity increase along the vertical dimension. The results for FZZ (fixed zigzag) and SZZ (stochastic zigzag) stimuli are presented next to each other.

To verify whether prediction (1) regarding proximity is confirmed, it is easiest to concentrate on the lowest row with collinear lattices ($\theta = 0^\circ$). From left to right, the relative proportion of $b$-responses (in green) decreases in favor of the $a$-responses (in red). This is the pattern of data expected on the basis of grouping by proximity. The particular condition with $|a| = |b|$, is a starting point to explore the main effect of discollinearity (prediction 2). In principle, proximity should not play a role, since dots in the $a$- and $b$-vector orientations are equidistant. Thus the increase in $a$-responses one can observe moving up along the middle column in each set of axes (where $|b|/|a| = 1.0000$), is purely the influence of discollinearity. In the margins, below the horizontal and to the left of the vertical axes, one can inspect the evolution of response frequencies as a function of distance and discollinearity respectively, collapsed over the other experimental factors. Comparing the modulation in frequencies, it is obvious that, at least in the range in which the variables were manipulated, distance is much more important for grouping than discollinearity. The global orientation preference (prediction 3) can be read from the radial line graph in the corner of the data layout. Each of the radial lines is drawn in the orientation $\rho$ in which the $a$-vector could be oriented ($0^\circ$ to $170^\circ$). The length of a line corresponds to the proportion of $a$-responses relative to the sum of $a$- and $b$-responses. The pattern at the right, for subject PF, for instance, is clearly indicative of a large preference for vertical over horizontal grouping.

---

Figure 4. Left: Graphical representation of the response frequencies for subject BN. The stack bars show, per type of zigzag (FZZ vs. SZZ) the relative portions of $a$, $b$ and blank responses (color legend in lower left corner). From left to right within each of the inner axes, the value of $|b|/|a|$ increases. From the lower to the upper row, discollinearity increases from 0 to $20^\circ$. The arrangement is related to the layout of stimulus examples in Figure 2. Outside the inner axes, marginal frequencies are shown. Right: Graph as the left panel, for a subject with a much stronger orientation preference. The radial lines in the corner plot the number of $a$-responses for each orientation of $a$, relative to all trials without blank response.
Independent combination of cues

We now turn to the evaluation of Bayesian strategies in the combination of grouping cues. In the first stage of this analysis, we did not impose any constraint on the relation between the variation in the independent variables and the observed log-odds of the responses. One can think of this as a nonparametric version of independent combination of grouping cues. In this model, the expected log-odds for the combination of a given distance ratio, discollinearity and global orientation amounts to a sum of estimates for each level of the stimulus variables. Adding indices to Equation 4, we can write

$$\log \left( \frac{P(b|d_{i}^{(i)}, \theta_{b}^{(j)})}{P(a|d_{i}, \theta_{a})} \right)^{(k)} = \log \left( \frac{P(d_{i}^{(i)}|b)}{P(a|d_{i})} \right) + \log \left( \frac{P(\theta_{b}^{(j)}|b)}{P(\theta_{a}|a)} \right) + \log \left( \frac{P(b|a)}{P(a)} \right)^{(k)} \tag{11}$$

—that is, the log of responses for the $i^{th}$ distance and $j^{th}$ discollinearity along $b$ is simply the sum of their log Bayes factors within each level ($k$) of global orientation. The $a$-vector is not indexed, because its inter-dot distances and collinearity are held constant in the experiment. Each level of each stimulus variable is represented by a separate log Bayes factor, except for the logical constraints that the evidence provided by the proximity cue equals 0 if $d_{i} = d_{m}$, as for the discollinearity cue if $\theta_{b} = \theta_{a} = 0$. With 4 degrees of freedom for the influence of distance ratio, 4 degrees of freedom for discollinearity, and 18 degrees of freedom for global orientation, there are 26 parameter estimates for $5 \times 5 \times 18 = 450$ combinations of the variables involved.

The model summarized in Equation 11 was contrasted with a model in which interaction between the grouping cues was allowed, in the spirit of Equation 1. In the latter case, the likelihood function, and therefore the log Bayes factor, of the combined cues does not equal the sum of the individual log Bayes factors anymore:

$$\log \left( \frac{P(d_{i}^{(i)}, \theta_{b}^{(j)}|b)}{P(d_{i}, \theta_{a}|a)} \right) \neq \log \left( \frac{P(d_{i}^{(i)}|b)}{P(d_{i}|a)} \right) + \log \left( \frac{P(\theta_{b}^{(j)}|b)}{P(\theta_{a}|a)} \right) \tag{12}$$

This means that the most generic model allows for a separate log-odds estimate per combination of distance and discollinearity. Again, there is an exception for the combination $d_{i} = d_{m}$ with $\theta_{a} = \theta_{b} = 0$, because the $a$- and the $b$-vector cannot be distinguished in this condition. Logically, the log Bayes factor for this case is 0, and grouping odds purely reflect the prior odds. This generic model thus has 24 degrees of freedom for the specification of likelihood function: one for each $(i, j)$ combination, minus one. Together with the 18 levels of the prior odds, this amounts to 42 parameters to be estimated.

As further detailed in Appendix A, statistical tests comparing models are based on the likelihood-ratio statistic. In this case, maximum binomial likelihoods were established for the model with and without the parameters modeling cue interaction. The difference between minus twice the log of these likelihoods (i.e., the deviances) is approximately $\chi^{2}$ distributed with $42 - 26 = 16$ degrees of freedom. However, due to the high number of free parameters in this specific case, we opted to rely on bootstrapping rather than asymptotic theory to approximate the distribution of the test statistic and the corresponding $p$-value. In addition, we combined the individual test statistics to obtain a “global” $p$-value under the appropriate distribution. This meta-analytical procedure allows for the quantification of the statistical merit of competing models as evaluated across observers (see Appendix A for details).

For the FZZ model fits, the likelihood ratio test yields $p$-values $>.50$ for 4 subjects and one $p$-value between .25 and .50. One $p$-value was nearly significant, with $p \approx .06$; the likelihood ratio statistic reached significance at .05 level in a single case, with $p \approx .02$. For the subset of SZZ stimuli, for two observers we found $p > .50$, and for the others $.25 < p \leq .50$. Summed deviance differences, both across and separated per stimulus type, all produced global $p$-values near .50. Based on this analysis there is no evidence that we should reject the simpler independent model in favor of the generic model with more parameters: including interaction only trivially improves model fit. Therefore, we can conclude that observers combine grouping cues as independent pieces of information. In the parametrically guided analysis (Formalization of cue interactions section), we will return to the possibility that the volunteer with a .02 $p$-value for interaction with the FZZ stimuli (BN) is an exception in the group.

Formalization of likelihood functions and priors

Proximity likelihood

We now investigate the applicability of Equations 5 and 6, which reflect an exponential or power-law structure for proximity information in grouping. Predictions are compared with the original Pure Distance Law, PDL (Kubovy et al., 1998; Kubovy & Wagemans, 1995), which is not based on the idea of grouping as Bayesian inference.

The exponential model, Equation 5, predicts a log-odds that is linear as a function of $|b| - |a|$. Under the power law, expressed in Equation 6, response odds are a linear function of $\log(|b| - |a|) - \log(|a|)$. The Pure Distance Model states that the log-odds are a linear function of $(|b| - |a|)\min(|a|, |b|)$. In the range of tested inter-element distances, the difference between these predictors...
is numerically small. The validity of the parametrical models according to the exponential model, the power law, and the PDL can be evaluated statistically by comparison with the generic independence model of Equation 11 in a likelihood ratio test. The models according to exponential, power law, and the PDL are not mutually nested. Therefore, comparisons among these have to be based on how well each fits the data empirically, as well as on theoretical considerations.

The graphs illustrating the data with median fit (Figure 5) show how close predictions for the original Pure Distance Law and the power-based principle are. Nevertheless, they do differ slightly in the quality of fit to the data. Among this pair of models, the power law provides the better fit in 13 of the 14 cases. The purely exponential function performs somewhat better than the power law for 4 FZZ sets and 1 SZZ set, while the power law “wins” for 2 FZZ and 5 SZZ sets. Apparently, the power law performs better when zigzag is stochastic than when it is deterministic. Global $p$-values of aggregated deviances indicate significant deviations from all three models. When the most extreme deviance was excluded from the calculation, $p$ exceeded .15 for the power law model. This suggests that the tested data set (FZZ for subject CSE), with a deviance much larger than average, is atypical. Selectively excluding data sets (“jackknifing”, see the section on meta-analytical principles in Appendix A) did not increase the $p$-value for the exponential model. Therefore, unlike the power law, it fails globally.

When data are not fitted well by the power law (and therefore neither by the resembling PDL), the corresponding global curvature does not match the scale invariance property. If scale invariance does apply, inversion of values of $|b|/|a|$ (e.g., from 0.8 to 1.25) leads to sign inversion of the response log-odds (e.g., from a logit of 6 to a logit of $-6$). The deviation of data from scale invariance is then most easily evaluated by how the nonparametric log-odds for the most extreme distance ratios differ in absolute value. Deviations from the power law are unsystematic, generally moderate and rarely significant. Predictions of the exponential model, on the other hand, tend to underestimate the curvature of the data. The CSE/FZZ data set is the only set with curvature still lower than exponential. Subject CSE’s pattern of FZZ data points toward a global preference for a convex likelihood function.

Besides the better data fit, some more theoretical arguments also favor the power law among its competitors. To the extent that scale invariance is a requirement for perceptual grouping—and Elder and Goldberg (2002), among others, suggest that it should be—an exponential model should in principle be excluded. The Pure Distance Law depends on the division by the shortest inter-element distance for scale invariance to apply. This is a disadvantage: the visual system needs to determine which of the lattice vectors is the shortest in order to evaluate the relative likelihoods for grouping. This adds an algorithmic step to the grouping process that is hard to explain from a Bayesian viewpoint. In contrast, grouping based on the power law distribution is scale-invariant and allows for parallel one-step calculations for the likelihood of each lattice vector.

Figure 6 summarizes the estimates for the power law exponents. To allow for comparison with other data on grouping by proximity we included a panel with estimates for the weight of proximity for responses modeled without a prior. For most observers, the estimates are clearly larger in absolute value than the exponent estimates of Elder and Goldberg (2002) or Oyama (1961) (dashed lines). It is important to realize the magnitude of this effect. An exponent of $-20$ means that, upon doubling the distance between image elements, the probability that they belong to the same contour drops with a ratio of one in a million! Note that the order across observers and even the value of the exponent are largely preserved between stimulus types (Spearman rank order correlation in the order of 0.9, significant for $n = 7$ at $a = .05$).

**Collinearity likelihood**

As shown in Equations 7 and 8, there are at least two candidates for the formalization of the effect of discollinearity on grouping. One is based on a von Mises distribution of turning angles, and predicts log Bayes factors that are linear, without intercept, as a function of
the cosine of the turning angle minus one. A discollinearity likelihood based on a Laplace distribution of angles predicts log Bayes factors that directly reflect a linear transformation of the turning angle. For comparison with Elder and Goldberg (2002), we also fit the two-parameter generalized Laplace distribution (see Equation 9). The generalized Laplace distribution includes the regular Laplace distribution ($\gamma = 1$), as well as a very good approximation to the von Mises distribution ($\gamma = 2$) as special cases. We will analyze these alternatives in the same way as we did with proximity, by comparing their model fits and predictions with the free-form, nonparametric model of independent combination of grouping cues (Equation 11), as well as among each other.

Figure 7 shows the results of the model-fitting procedure for the participant with the median fit among observers. Across the range of the discollinearity manipulation, effect size is only a fraction of the large shifts in response frequencies seen during manipulation of distance ratio (note the different axis scaling when comparing with Figure 6). The maximum effect was a log-odds estimate of $-3.07$. While this amounts to an appreciable reduction of the response odds to a ratio of approximately 1/20, the magnitude of maximal effect for distance— as it happens to be the case, in the same observer, JW—was the square of this, a factor of 400. The performance of the one-parameter models for discollinearity is more heterogeneous than was the case for distance. Interestingly, the best model depends on the type of stimulus. While for FZZ stimuli, the von Mises based likelihood provided the best fit (in 6/7 data sets), the Laplace likelihood function was best for 5 out of the 7 SZZ data sets. The $p$-values of added deviance differences confirmed this pattern; the “losing” model for each stimulus type deviated from the

Figure 6. Estimates of power law exponents for all observers. SZZ estimates are plotted along the ordinate against the corresponding FZZ estimate. We fitted the model without the prior orientation preferences and plotted the resulting estimates together with the $-2.92$ estimate reported in Elder and Goldberg (2002) (dashed lines).

Figure 7. Model predictions and free estimates for the effect of discollinearity in $b$ on the grouping odds. The shown data are the ones associated to the median fit (BN).
data to generate $p < 0.01$, versus $p \approx 0.44$ (von Mises, FZZ) and $p \approx 0.14$ (Laplace, SZZ). The latter value seems to signal a somewhat poor fit, but 40% of the global deviance was contributed by one participant (JW).

Not surprisingly, the generalized Laplace distribution gave the best overall fit. The large improvement in global model fit, compared to the one-parameter models, supported the inclusion of an extra degree of freedom. The only data set which kept defying our modeling efforts was JW/SZZ—jackknifing these data showed that model fit for the other data was in fact excellent, with a global $p > .5$.

In Figure 8 parameter estimates are plotted in log-coordinates, with base 2, such that the estimates of the exponent can easily be compared to the values 1 ($\log_2(1) = 0 – \text{Laplace}$) and 2 ($\log_2(2) = 1 – \text{von Mises}$). SZZ results tend to cluster near $\gamma = 1$, FZZ either near or above $\gamma = 2$. The right panel gives an idea of how the parameters translate into subjective probability distributions. Except for the large variability among observers, it is obvious that, in general, the SZZ discollinearity distributions are more kurtotic than the FZZ distributions. The lowest estimates we obtained for $\gamma$ coincide with the corresponding co-circularity estimate as reported in Elder and Goldberg (2002) ($\log_2(\gamma) \approx -.136$). On the other hand, the inferred standard deviations $\sigma$ of the distributions are 2.8 to 6.7 times larger than the standard deviation of Elder and Goldberg’s co-circularity distribution, even when orientation priors are excluded from the model for comparability.

### Formalization of cue interactions

Combining the parametric models for independent distance and discollinearit y grouping, we obtain the following formulation to predict log-odds of grouping for the $k^{th}$ global orientation:

$$\log \left( \frac{P(b|d_0, \theta_0^{(k)})}{P(a|d_0, \theta_0^{(k)})} \right) = -\beta \log \left( \frac{|b|}{|a|} \right) - \left( \frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)} \right)^{\frac{\gamma}{\gamma}} \sigma^{-\gamma} \theta_0^{b} + \log \left( \frac{P(b)}{P(a)} \right)^{(k)}$$

(13)

The second term involves gamma functions which originate from the parameterization of the generalized Laplace distribution. While it seems to complicate the equation, what matters is that, given a hypothesized distribution for distances and discollinearities in real contours, the log-odds are linearly dependent on the distance ratio and a power of the turning angle. While it was not tested in this experiment, our prediction is that with non-zero discollinearity in the $a$-orientation, the log-odds of response frequencies would be linearly dependent on $\theta_0^{b} - \theta_0^{a}$. Compared to the model with a completely free likelihood function (Equation 11), this model performed very well, with 11 out of 14 fits where the increase in deviance was not statistically significant. In the three other cases, two $p$-values were situated between .01 and .05, (CLS/FZZ and JW/SZZ) and one $p = .0015$ (CSE/FZZ), for the reasons mentioned before.

Having obtained the more succinct parametric version of the grouping model (Equation 13), we can quantify interactions between grouping cues in a more specific way. Notwithstanding the exact cause of the interplay between grouping principles, either type of interaction can be modeled as a dependence of the value of $\beta$ on turning angle ($\beta^{(i)}$), or dependence of the value of $\sigma$ or $\gamma$ on the distance ratio ($\sigma^{(i)}, \gamma^{(i)}$). The only interaction that emerged

Figure 8. Left: Estimates of spread and exponent of the generalized von Mises distribution, for FZZ and SZZ data, in log coordinates with base 2. Exponent estimates near the horizontal lines are typical for likelihood functions approximating the one-parameter Laplace distribution (at 0) and the von Mises distribution (at 1). Right: Visual representation of probability distributions of turning angle, which underlie likelihood functions. FZZ curves are more curved, with shapes that either remind of Gaussians, or only deflect from a plateau at high discollinearities. SZZ curves are, in general, much more leptokurtic.
in a consistent manner occurred in the BN/FZZ data set. As one can verify in Figure 9, the modulation of distance grouping by discollinearity is ordinally coherent: the larger the discollinearity along the $b$-vector, the stronger response frequencies are determined by the distance ratio ($p = .002$, for log-likelihood ratio test of the model with $\beta^{(\cdot)}$ versus fixed distance weight $\beta$). Apparently, when $b$ is already at a “disadvantage” because $|b|$ is larger than $|a|$, discollinearity gets “punished” more, as if collinearity and proximity mutually enforce each other. If, on the other hand, $b$ is short compared to $a$, discollinearity does not play any role of importance.

Formalization of prior orientation odds

Observers use proximity and collinearity information in a rather consistent way in contour integration. The same cannot be said of their preference for global orientation, which reflects variability in their prior tendencies to group along certain orientations. Orientation anisotropies in grouping tasks have been documented before (e.g., Olson & Attnave, 1970). Remember that in dot lattices, the tendency to choose one or the other vector as a function of lattice orientation reflects relative preference of one versus the other global orientation. From a Bayesian point of view, that is, supposing that one knows the ecological statistics of contour orientations, it is possible to develop normative guidelines about how relative orientation probabilities should be evaluated. According to the data of Coppola et al. (1998), horizontal and vertical contours are more prevalent than other orientations in all types of scenes (indoor, outdoor, and forest scenes). The way orientation frequency drops off with deviation from the cardinal axes and whether vertical or horizontal orientations are more frequent depends on the type of scene. Analysis of the data of Coppola et al., digitized from their histograms, revealed that orientation distributions can be thought of as discrete mixtures of at least three important components: horizontal and near-horizontal orientations, vertical and near-vertical orientations, and a set of orientations uniformly distributed between $0$ and $180^\circ$. The weights of these three components and the shape of the near-vertical and near-horizontal distributions are scene-specific. Orientation frequencies in urban scenes decrease from cardinal axes in a way that is reminiscent of the Laplace distribution and are dominated by vertical orientations. In natural scenes, the drop-off is more gradual and has the characteristics of a von Mises distribution, and horizontal orientations are slightly more prevalent than vertical ones. Also Hansen and Essock (2004) arrive at a horizontal preponderance conclusion for scenes devoid of manufactured structures.

Any variation in spread or weight of vertical and horizontal orientations will influence the curve of response log-odds as a function of global lattice orientation. Specific combinations of parameter values in the mixture distribution lead to very different curves. It is not easy to infer parameter values from the data without additional constraints, because of sizeable intercorrelations in parameter estimates. At a more fundamental level, many observers do not follow the normative strategy: the log-odds of their orientation preferences are asymmetric around the cardinal axes (see Figure 10). This means that, for these observers, cardinally symmetric orientations—e.g., $20^\circ$ clockwise and $20^\circ$ counterclockwise away from vertical—are associated with quite different frequencies. Internal models of orientation statistics appear to be rather idiosyncratic, and are subject to individual biases.

Grouping cues and priors interact through global orientation

There is yet another level at which observers’ orientation preferences deviate from what one would
expect. From the fact that prior probabilities are not affected by stimulus information, the level of prior support for a certain orientation, say, 10°, should be independent of whether this orientation is the orientation of the a- or the b-vector. In rectangular lattices, the orientations involved in a stimulus and a stimulus rotated 90° are exactly the same—e.g., 10° and 100°. Lattices with $\rho_a = 10°$ ($\rho_b = 100°$) and $\rho_a = 100°$ ($\rho_b = 10°$) involve identical prior probabilities, with differently assigned labels. Therefore, it would make sense to include the following constraint in the model:

$$P(\beta_a > \beta_b) = P(\beta_a + 90° > \beta_b + 90°),$$

or

$$\log P(\beta_a > \beta_b) / P(\beta_a > \beta_b) = -\log P(\beta_a > \beta_b + 90°) / P(\beta_a + 90° > \beta_b + 90°).$$

This means that the log-prior ratio for 10° should be the inverse of the one for 100°, for 0° and 110°, etc. In general, this is not the case, even if taking variability of estimates into account. This can be verified statistically by comparing model fits with and without the inversion constraint. The constraint would reduce the model by 9 degrees of freedom, but model fit is affected negatively, with significant ($p < .05$) reductions in the goodness-of-fit for more than half of the data sets (4/7 FZZ and 4/7 SZZ sets).

If a 90° shift of lattice orientation does not produce inversion of log-prior ratios, it means that, to the visual system, it does make a difference whether a certain vector is an a or a b-vector; in other words, the experimental manipulations of proximity and collinearity “interact” with the global orientations when converging toward a grouping percept. To understand better the origin of this effect, we re-fitted models with orientation-dependent estimates for $\beta$, $\gamma$ and $\sigma$, while including the inversion constraint, forcing global orientation preferences to behave as reflecting prior probabilities. In order to obtain stable estimates and to allow the model-fitting algorithm to converge, we tested inclusion of orientation-specific $\beta$, $\gamma$ and $\sigma$ separately. In addition, we partitioned the stimulus presentations into near-horizontal lattices ($\rho_a = 170°$, 0°, 10°), near-vertical lattices ($\rho_a = 80°$, 90°, 100°), and oblique sets near 30°, near 60°, near 120° and near 150°.

![Figure 10](image-url) Prior grouping odds, $\log(\frac{P(b)}{P(a)})$, as a function of the a-vector orientation, for the observer with the median influence of global lattice orientation. The preference for vertical over horizontal orientations is common to all observers. Asymmetries in the curves reflect idiosyncrasies and divergence from realistic orientation distributions.

![Figure 11](image-url) Variation of $\beta$ and $\log_2(\gamma)$ as a function of $\rho_a$. Diagonal lattices are associated with larger (absolute) $\beta$ but smaller $\gamma$, which reflects a stronger role for proximity in oblique orientations, and more weight for small disollinearities in cardinal orientations.
Some results of the model-fitting procedures are shown in Figure 11. The absolute value of $\beta$ is larger for obliquely oriented lattices than for lattices with main vectors near the cardinal axes. Proximity is treated as if it were less informative at horizontal and vertical orientations. The opposite is the case for $\gamma$ and $\sigma$ (not shown). Discollinearity affects grouping more at vertical, and especially at horizontal orientations. With $\gamma$ inversely related to kurtosis, the implied distribution of turning angles is clustered more tightly near 0° for horizontal lines. When $\gamma$ is held constant, $\sigma$ follows a similar pattern: the spread of the implied discollinearity distribution is tighter for horizontal and vertical lines than for diagonal ones. In summary, observers seem to use the assumption that horizontal and vertical contours are longer and straighter than oblique lines.

**Discussion**

**Summary**

We used dot lattice stimuli, varying distance ratio and introducing zigzag, to determine how human observers combine proximity and collinearity information into a single consistent grouping percept. The results were interpreted in a Bayesian paradigm for contour integration (for a diagram, see Figure 12). We can summarize our findings as follows:

1. In general, proximity and collinearity are treated as independent sources of information for grouping.
2. Global orientation of a grouping candidate strongly influences its salience, reflecting prior assumptions about orientations of contours. The observers in this study preferred vertical orientations over horizontal ones, although with large inter-individual variation in the effect size.
3. The most successful model for the subjective likelihood function of distance is based on a power law. However, predictions from the Pure Distance Law are practically indistinguishable.
4. The most successful model for the subjective likelihood function of discollinearity is based on the generalized Laplace distribution. Parameter estimates point toward higher kurtosis in the subjective distribution of angles along contours with stochastic zigzag than with regular zigzag.
5. Observers diverge significantly in the weight of evidence assigned to different grouping cues. Nevertheless, there is a remarkable consensus in the principle of independence of cues in their combination,

---

**Figure 12.** Schematic representation of the conceptual framework. Stochastic properties of the external world determine global orientations and geometrical statistics of contour elements. In a Bayesian inference framework, the visual system uses these stochastic regularities in likelihood functions and prior distributions to arrive at a global orientation judgment. Likelihoods for each lattice vector are determined on the basis of their inter-element distances and discollinearities, modulated by global orientation. Evidences from grouping cues are summed and added to prior orientation probability of the response candidates to establish posterior log-odds of the response options. The response frequencies are a consequence of a probability matching decision rule.
as well as in the importance of evidence, with a 25% increase in distance being much more detrimental to grouping than a zigzag of 20°.

6. Further analysis revealed that grouping by proximity is weakened for cardinal lattice orientations, while discollinearity exerts a stronger influence for these orientations. This is consistent with an internal model in which horizontal and vertical contours tend to be longer and straighter than contours in oblique orientations.

A normative versus a descriptive Bayesian framework

All volunteers in this experiment have undoubtedly been raised in similar environments, with presumably little variation in ecological contour statistics. Yet, the sizeable differences in the weight of evidence derived from various grouping cues show that we are certainly not using the same model of stochastic variations in contours. This implies that we do not follow actual contour statistics. From a Bayesian point of view, we combine conditionally independent cues as they should be combined, but with weights that are not well tuned to statistical properties of the world; we do use Bayesian logic in perceptual inference, but with an input composed of beliefs rather than probabilities in the frequentist sense. The term already implies the way in which beliefs might be different from learned veridical probabilities: they are moldable. Research has indeed shown that, given novel environments, observers are able to quickly grasp the changes and use them in perceptual judgments. On the other hand, they do not change the structural logic of cue combination (Michel & Jacobs, 2007).

Inter-individual differences might also arise from stable or momentary variations in sensitivity to stimulus attributes. In a bottom-up approach, perceptual organization can only be influenced by a grouping factor in as much as it is detectable. The extent to which grouping and detectability are coupled, is an interesting issue that will have to be tackled in an experiment combining signal detection and grouping task methodologies.

Likelihoods for strings of contour elements

Subjective distributions of distance and discollinearity of contour fragments seem to be narrower than the physical distributions measured in the literature on ecological statistics. Part of the explanation may lie in the number of dot dipoles used in the perceptual inference process. It has been proposed that grouping in dotted lines occurs within a four-dot window (Feldman, 1997). This finding was corroborated by a similar conclusion for interpolation (Warren, Maloney, & Landy, 2002, 2004). What is the consequence of this in probabilistic inference?

If a single realization of an observable variable leads to a certain likelihood function, say, \( P(x_1|\xi) \), then four independent realizations of the same variable will have a joint likelihood of \( P(x_1, x_2, x_3, x_4|\xi) = P(x_1|\xi) P(x_2|\xi) P(x_3|\xi) P(x_4|\xi) \). If the four realizations of \( X \) have the same value, as do four distances along a chain of equidistant dots, \( P(x_1, x_2, x_3, x_4|\xi) = P(x_1|\xi)^4 \). Generalizing, for \( n \) identical measurements, the joint likelihood function gets raised to the \( n^{th} \) power, or the log-likelihood multiplied with \( n \), compared to the case of a single measurement.

As to the distance cue, jointly evaluating \( n \) dipoles as independent sources of information rather than \( 1 \) raises the probability distribution for the distance in each orientation to the \( n^{th} \) power. This would mean that the estimates obtained for \( \beta \) in Results section are an \( n \)-fold of Elder and Goldberg’s 2.92 estimate (see Figure 6). The estimates suggest that subjects might use 6–7 dipoles per orientation to evaluate grouping odds based on distance. Similarly, while the co-circularity cue in Elder and Goldberg has a standard deviation of approximately 1.34 radians, or about 0.42 in \( \log_2 \) units, a joint set of \( n \) discollinearities will generate a downward shift of the estimated standard deviation. This is somewhat harder to evaluate numerically, but in any case, variations like these might help to explain where inter-subject variability originates from.

Independent cue combination

Elder and Goldberg (2002) extensively discuss the independence of grouping cues. We found that, in this regard, subjects do behave normatively: with a single exception, the contributions of proximity and collinearity were additive in log-odds, which points toward Bayesian independent cue combination. This is not the first time that grouping laws have been shown to operate additively in the response logit. In an earlier study, we obtained the same result for proximity and alignment in Gabor lattices (Claessens & Wagemans, 2005). In a recent publication, Kubovy and van den Berg (2008) discuss the independence in the combination of proximity and brightness similarity. The linear increase or decrease in grouping log-odds provoked by manipulations of brightness consistency is largely compatible with the generalized Laplace fitted by Elder and Goldberg (in fact, with \( \gamma = .89 \) it is an almost pure Laplace distribution), and with the independence of brightness and proximity in ecological data. Results in various dot lattice experiments thus converge, possibly not in the exact numerical parameter values but at least in the qualitative principles, toward an optimal Bayesian combination of different sources of grouping information.

The principle of addition of “weights of evidence” in the perceptual grouping task is in line with the findings in the probability learning task by Yang and Shadlen (2007).
Even when sources of information were not statistically independent, the neural populations kept treating them as if they were. This suggests that in certain situations of probabilistic evaluation, we might behave as what is known in machine learning as naive Bayesian classifiers. It has been found that, even violating normative combination rules, naive Bayesian classifiers are surprisingly accurate in many situations (e.g., Domingos & Pazzani, 1997). There are important advantages to this “suboptimal” algorithm: memory requirements increase linear rather than exponentially with the number of stimulus characteristics, and the straightforwardness of the calculation favors speed, learning rate, and simplicity of the computational architecture compared to the normative model.

**Modulation of grouping cues**

Yet, in the results of the current experiment, this independence has to be nuanced: when candidate orientations are near vertical and horizontal, observers treat proximity as less informative and collinearity as more informative than in obliquely oriented lattices. At first sight, this makes sense. With construction and morphogenesis under the influence of gravity, objects consisting of long stretches of straight parts are naturally rotated toward cardinal axes. Under realistic viewing conditions, longer contours in the center of our visual field tend to be either horizontal or vertical. This seems to be related to the analysis of contour orientation statistics by Keil and Cristóbal (2000), who found that anisotropies are more pronounced for lower spatial frequencies. More research in ecological statistics is needed to establish an empirical basis to allow for quantitative predictions of exactly how much global orientation influences the information value of grouping cues. With bending contours and, ideally, the inclusion of 3-D considerations, this will certainly be an operational challenge.

This complication issues a warning. If the visual system is able to conditionalize the distributions underlying proximity and collinearity cues on orientation, it might do so for other cues as well, accounting for dependencies on specific stimulus characteristics. For example, Bleumers et al. (2008) found that the weight assigned to proximity in perceptual organization of dot lattices is modulated by their eccentricity. This might reflect different statistics for contours in the visual periphery. From Coppola et al. (1998) we know that the prior orientation distributions for urban and forest scenes have quite different characteristics. One might expect that the visual system, if it has a way of “storing” this information, will change its global preference and, for example, favor horizontal over vertical orientations. This has been discussed by Yuille and Bulthoff (1996) as the concept of competitive priors. Context mediates the information value of cues in the same way as it determines the correct prior probabilities. In typical urban settings, most objects are man-made. Man-made objects have on average more leptokurtic contours: the bends a contour makes tend to be more concentrated around perfect collinearity than natural contours, with “heavy tails” representing higher prevalence of corners (Elder, personal communication). Especially in cases in which the gains in terms of correctness of inference are considerable, it is possible that the visual system uses this information rather than behaving as a naive Bayesian. For this reason, if it were put to the test, one might observe a correlation between the prior density function of orientation and the parameters of the likelihood function for good continuation. More experiments are needed to clarify the extent to which observers are tuned to contextual information.

Similarly, the information value of grouping cues depends on other stimulus aspects that do not necessarily enter the grouping process in a more direct way. It is useful to remember at this point that contour likelihood dropped faster with increasing discollinearity for stochastic than for systematically induced zigzag. Grouping processes thus turn out to be sensitive to subtle changes in the contour candidates. Based on the properties of the stimulus, the visual system selectively adjusts the influence of some cues (good continuation) but not of others (proximity) in the perceptual inference outcome.

For the specific case of local discollinearity, the importance of global interpretation has convincingly been demonstrated by Strother and Kubovy (2006). Imposing curvilinearity on a vector of a dot lattice, they found that the resulting curved pattern was in fact a more salient grouping than the grouping along the orthogonal orientation, along which dots were collinear. This forms an interesting contrast with the results reported here; after all, curvilinear dotted lines are locally composed of discolinear dot triplets. Yet, curvilinearity strengthens grouping, while zigzag weakens it. It might not be too surprising that curvilinear dotted lines are easier to group than zigzag lines, but it is rather counterintuitive that curvilinear patterns also should be easier to group than collinear patterns. Contextual information, whether local or global, can irresistibly push the visual system to prefer curvilinear rather than rectilinear configurations. For instance, van Assen and Vos (1999) have shown how flanking dots, even when instructed to ignore, biased interpolation between two dots away from a straight line, toward a smoothly curved one. Strother and Kubovy rely on the influence of non-accidental properties to explain their result. In the context of Bayesian contour integration, Feldman and Singh (2005) give a clue in deriving the “surprisal” value of dotted contours on a shifted von Mises distribution rather than one that is centered on 0° turning angles. The contours that Feldman and Singh analyze have a natural curvature tendency because of closure. An analysis of the Strother and Kubovy data along the lines of the current paper, with an adjustment in
the discollinearity likelihood function to account for closure, might explain higher salience of curvilinear than of ortholinear lines.

In a complete Bayesian account, this would require the specification of a continuous mixture of exactly how many contours curve how much because of closure. This involves an extension of the hierarchical structure in the model. Comparing and combining results from different research groups, it becomes obvious that the study of hierarchical probabilistic reasoning, as formalized in Bayes nets (e.g., Glymour, 2003) will gain in importance. Bayesian models will need to grow in structure, but with significant gains in how much we understand about cue combination, in grouping, depth perception, and other fields, context effects, competitive priors, etc. In our opinion, it is an important direction for future research that promises to provide a formal framework for mechanisms that have been documented before, such as accumulation, cooperation, disambiguation and vetoing (Bülthoff & Mallot, 1988), competitive priors (Yuille & Bülthoff, 1996), principles of modified weak fusion (Landy et al., 1995) and other robust integration frameworks (Knill, 2007), and interacting prior constraints (Mamassian & Landy, 2001).

Conclusions

In the current paper, the issue of combination of grouping cues into a multistable percept was approached from the Bayesian integration framework. We drew upon previous work in the ecological statistics literature to develop a psychophysical counterpart; in doing so, we intended to provide an empirical evaluation of how well qualitative and quantitative aspects of grouping in human subjects match expectations from the statistical analysis of contours in natural scenes.

As to the qualitative comparison, it can be concluded that distance and discollinearity are, in general, combined in an independent, non-interactive manner. This was tested independently from the exact form of the likelihood functions involved, as well as with the best-fitting parametric likelihood functions. In the comparative evaluation of which of three likelihood functions best describes the influence of each grouping cue, results are again in line with ecological statistics: grouping probability decays with distance according to a power law, while generalized Laplace distributions apply for both patterned and stochastic zigzag. On a cautionary note, it is not always possible to clearly distinguish between two likelihood functions, as was the case here for the power law and the Pure Distance Law. Parameter estimates diverge from their counterparts in scene statistics for most observers. In the discussion, we show how this might be attributed to integration of information across a sample of the stimulus.

The prior distribution of global orientation was harder to capture formally, as it is subject to inter-individual differences and inconsistencies. A preference for vertical over horizontal groupings emerged, albeit with varying magnitude among observers. More importantly, we have shown that global orientation correlates with the strength of grouping cues in a specific way. The lower influence of distance for cardinal orientations and higher influence for discollinearity, compared to oblique orientations, suggests that the visual system expects contours near horizontal and vertical to be longer and straighter. While our Bayesian analysis to grouping was greatly inspired by the ecological statistics literature, this new finding in psychophysics, in turn, calls for new input from ecological statistics.

Appendix A

Some aspects of model fitting and evaluation

Maximum likelihood estimation

The process of fitting the model and parameters is based on a standard maximum likelihood procedure. Each trial corresponds to a Bernoulli event that leads to a $b$-response with a chance that can be calculated by the inverse of the logit:

$$P(b - \text{response}) = \frac{e^{\log \left( \frac{p_{ab}}{p_{ba}} \right)}}{1 + e^{\log \left( \frac{p_{ab}}{p_{ba}} \right)}}$$

$$P(a - \text{response}) = 1 - P(b - \text{response}) \tag{A1}$$

The model makes no predictions regarding the occurrence of blank alternatives, and the according responses are discarded from the model-fitting procedure. The actual likelihood of a $b$-response, per trial, depends on the stimulus at hand; in terms of lattice parameters, summarized in $d_i, \theta_a, \theta_b, \rho_a, \rho_b$. Let us write that, for the $i^{th}$ trial, the log-odds are $\log(P(b|d_i,\theta_b,\rho_b)/P(a|d_i,\theta_a,\rho_a))$; the inverse logit, as given above, transforms this to what we will write as $P(a\text{-response})$, and $P(b\text{-response})$. The exact way in which the log-odds are determined by the stimulus parameters depends on the model used and the values of model parameters. The maximum likelihood principle dictates that the “best” estimates for values of model
The likelihood ratio test, Akaike’s and Bayes information criterion

Wilks (1938) has shown that the decrease in minimal deviance from a certain model to a trivially expanded version is approximately $\chi^2$ distributed, given certain regularity conditions. This is the core of the likelihood ratio test. Imagine that we were to test whether the threshold $\mu$ of a certain psychometric function $p = F(x, \mu, \sigma)$ equals 0. We would fit the model $M_1$, without constraints, leading to maximum likelihood $l_1$, with maximum likelihood estimates $\hat{\mu}$ and $\hat{\sigma}$. Fitting of model $M_0$ against the data would yield estimate $\hat{\sigma}_0$ and likelihood $l_0$, with $\mu_0 = 0$. The likelihood ratio, $l_0/l_1$, summarizes how much support we have for the null-hypothesis $\mu = 0$ versus the model without constraints. A value near 1 shows that the constraint does not greatly affect model fit, and that $\mu$ can be discarded as a free parameter without much loss. The actual statistic used to evaluate whether the constraint should be maintained or rejected is minus twice the log likelihood ratio ($-2 \log LR$) $-2\log(l_0/l_1)$—note that this equals the absolute differences between deviances. If the sample is sufficiently large, and if the null hypothesis is true, $-2\log LR$ is distributed as $\chi^2$, with as degrees of freedom the number of parameters held fixed by the hypothesized constraint. In our example, to test statistically whether $\mu$ might be 0, at significance level $\alpha$, we would look up the $1 - \alpha$ quantile in the $\chi^2$ distribution with $df = 1$. Only if the log-likelihood ratio as obtained for a sample of data exceeds this criterion, we would reject $\mu = 0$, at least in the classical Neyman-Pearson framework of hypothesis testing. The likelihood ratio test procedure is described in many classical and standard statistical textbooks (e.g., Bain & Engelhardt, 1992), especially in those covering categorical data analysis (e.g., Agresti, 2002).

Note that the likelihood ratio test applies to nested models, a framework particularly apt for tests of the relevance of parameters to explain the data. When the two models under consideration are not nested, the distribution of the likelihood ratio statistic is not specified. It usually will not be possible to derive a distribution useful for testing. In these situations, there are at least two diagnostic statistics popular in model comparison: Akaike’s information criterion (AIC) and the Schwartz or Bayesian information criterion (BIC). If we call the number of free parameters in a model $k$, and the number of observations $n$, then AIC is the deviance “corrected” with $2 \times k$, while BIC corrects the deviance with $\log(n) \times k$. Notwithstanding their obvious similarity, AIC and BIC have very different theoretical motivations. Yet both can be seen as a deviance that is penalized for relying on estimation to fit the data. BIC is considerably more severe in this penalty, which explains why model selection based on AIC and BIC can give quite different results.

Bootstrapped likelihood ratio distributions

In order to design a statistical test of the independent combination hypothesis explored in this paper, we compared pairs of nested models after maximum likelihood fitting using the likelihood ratio statistic. In most tests, we relied on the asymptotic $\chi^2$ distribution of deviance differences. However, due to the high number of estimated parameters involved as well as a considerable number of expected 0-frequency observations in the test of independent combination of grouping cues, we deemed it safer, in this particular case, to rely on bootstrapping techniques to have access to more accurate $p$-values. Specifically, we proceeded along the lines of the following scheme:

1. Establish the maximum likelihood estimates (MLE) under each of the competing models. Calculate the deviance for each: $D_{M_0}$, $D_{M_1}$. The test statistic is the absolute difference between the deviances.
2. Use the MLE of the parameters to generate artificial data sets, sampling simulated response frequencies from binomial distributions under the assumption that $M_0$, the constrained model, is true.
3. Perform a MLE procedure for both $M_0$ and $M_1$ on the simulated data sets. Calculate $D_{M_0} - D_{M_1}$ for each.
4. Fit a gamma distribution to the bootstrapped distribution of deviance differences.

While it is possible to approximate $p$-values with the raw bootstrapped distribution, the gamma distribution provides a smoothing that is theoretically related to the asymptotically expected $\chi^2$ distribution.
For MLE fitting of models to simulated data sets, 2001 per subject-stimulus (FZZ/SZZ) combination, we ran the SAS nlmixed procedure, as we did for the observed data sets. This yielded a \( p \)-value for each observer–stimulus type combination.

**Some meta-analytical principles**

In a simultaneous-testing situation, it is not unlikely that at least once \( p \) will be under a .05 significance level for purely statistical reasons. An isolated case of significance should not lead to automatic rejection of the null hypothesis. The models described in this paper were fitted at subject-level, and therefore we have seven \( p \)-values for each model comparison within each type of discollinearity. Several scenarios are possible. The most convenient situation, at least from a data-analytic point of view, is that either all \( p \)-values are reasonably near 0.5, without any significance, or all \( p \)-values are below 0.05. These are strong cases for absence or presence of the tested effect, respectively. But this is rather exception than rule. What to do if all \( p \)-values approach 5% without individually signaling significance? What if one or more \( p \)-values reach significance, but most do not? How to conclude whether there is an effect, whether we are confronted with an outlier or an entirely different group of subjects, whether the significance is a normal statistical fluctuation, or whether yet another stochastic mechanism is at work?

Ideally, hypotheses could be tested in random-effects models, in which not the individual effect sizes, but their distributions in the population are estimated. Unfortunately, this approach, which involves numerical evaluation of integrals or Markov Chain Monte Carlo simulation (e.g., Tuerlinckx, Rijmen, Verbeke & De Boeck, 2006), would be unreasonably computationally expensive for the models with a larger number of parameters. We will rely on some other meta-analytical principles instead. First, consider the hypothesis of before that a threshold parameter \( \mu \) in a psychometric function is 0. In the testing procedure, we would fit both the unconstrained and the constrained (\( \mu = 0 \)) model, per subject \( j \), where \( j = 1 \) to \( J \). The maximum likelihoods for both models would be approximately the same.

Under the null hypothesis, the differences between these, \( \Delta \mathcal{D}_j = \mathcal{D}_{M0,j} - \mathcal{D}_{M1,j} \), approximately follow a \( \chi^2_{df=1} \) distribution. Distribution theory states that the sum of \( \chi^2 \)-distributed variables is also \( \chi^2 \) distributed, with as shape parameter the sum of degrees of freedom of the constituent distributions. In short, \( \Delta \mathcal{D} \sim \sum_{j=1}^J \Delta \mathcal{D}_j \sim \chi^2_{df=Jk} \). This is a very useful result to establish a “global effect” criterion: if the aggregated deviance exceeds the .95 quantile in a \( \chi^2_{df=Jk} \) distribution, a “global effect” exists. Except for serving as a very welcome summary of the global pattern of observers’ data, the combination of deviance differences provides a large increase in statistical power compared to the individual fits. On the other hand, this does not allow us to discard the individual deviance differences or \( p \)-values from our discussion. It is very well possible that one of our volunteers is an outlier and is subject to a completely idiosyncratic effect. In this case, the global deviance difference will be largely constituted by the deviance difference of one single individual; in other words, while most of the deviance differences would be in the “body” of the \( \chi^2_{df=k} \) distribution, one would find one deviance difference far in the “tail”. A case like this is easily discovered by jackknifing: leaving out each participant in turns, re-calculating the aggregated deviance and re-evaluating global significance. Another useful tool to diagnose this type of situation is the quantile plot, in which obtained deviance differences are plotted against their ordered expected values in a sample of size \( J \) from a \( \chi^2_{df=Jk} \) distribution. One can also plot the \( p \)-values, which should approximately be uniformly distributed in the interval \([0, 1]\) if the null-hypothesis is true. If all deviance-difference points cluster very much toward one side, there is a global effect. If all points are scattered tending, in median, toward the middle, there is a global absence of effect. If all points are reasonably close to the middle of the theoretical distribution, but one approaches boundary or is situated far in the tail, we have an outlier. If a group is near median, and a group near the extreme, the participants are subdivided in two groups.

**Acknowledgments**

This research was supported by a grant from the Fund for Scientific Research-Flanders (G.0095.03) and from the Research Fund of the University of Leuven (GOA 2005/03). The preparation and execution of the reported experiment took place at the University of Leuven, Belgium, as part of the doctoral research of the first author under the supervision of the second author. Some of the theoretical development, analysis, and manuscript writing took place at the Universidade de São Paulo, Brazil, which is the current affiliation of the first author (supported by fellowship 06/50189-9 from the Fundação de Amparo à Pesquisa do Estado de São Paulo). The final version was prepared with support from the Methusalem grant by the Flemish Government awarded to the second Author (METH/08/02). We would like to thank Dirk Smit and Astrid Janssens for their assistance in conducting pilot experiments. We also thank Francis Tuerlinckx, Karl Verfaillie, James Elder, and two anonymous reviewers for their valuable remarks on a previous draft.

Commercial relationships: none.

Corresponding author: Johan Wagemans.

Email: johan.wagemans@psy.kuleuven.be.

Address: Laboratory of Experimental Psychology, Department of Psychology, University of Leuven, Belgium.
Footnotes

1In principle, the common Laplace distribution is not a circular distribution. The wrapped Laplace distribution, which could be bounded between $-180^\circ$ and $180^\circ$, has been proposed (Jammalamadaka & Kozubowski, 2003), but the corresponding log Bayes factor is exactly the same in the range $-180^\circ$ to $180^\circ$.

2We assumed that our response variable of interest, the relative proportions of $a$- versus $b$-responses, would not depend on the number of response alternatives offered. The independence of the relative preference between two choice alternatives (here, $a$ and $b$) from the presence or absence of other choice options (here, the blank) is an axiomatic idea in choice theory known as irrelevance of independent alternatives (Luce, 1959). Although substantial differences between participants in the tendency to use the blank option do exist, the results support our reliance on the irrelevance assumption: the pattern of $a$- and $b$-response frequencies for the subset of perfectly collinear stimuli ($\theta = 0^\circ$) is identical to the pattern obtained in the analyses by Kubovy et al. (1998) for the same stimuli with four response alternatives.

References


