Psychological Review
2015, Vol. 122, No. 4, 575–597

0033-295X/15/$12.00 http://dx.doi.org/10.1037/a0039540

Bayesian Hierarchical Grouping: Perceptual Grouping as Mixture Estimation

Vicky Froyen, Jacob Feldman, and Manish Singh
Rutgers University

We propose a novel framework for perceptual grouping based on the idea of mixture models, called Bayesian hierarchical grouping (BHG). In BHG, we assume that the configuration of image elements is generated by a mixture of distinct objects, each of which generates image elements according to some generative assumptions. Grouping, in this framework, means estimating the number and the parameters of the mixture components that generated the image, including estimating which image elements are "owned" by which objects. We present a tractable implementation of the framework, based on the hierarchical clustering approach of Heller and Ghahramani (2005). We illustrate it with examples drawn from a number of classical perceptual grouping problems, including dot clustering, contour integration, and part decomposition. Our approach yields an intuitive hierarchical representation of image elements, giving an explicit decomposition of the image into mixture components, along with estimates of the probability of various candidate decompositions. We show that BHG accounts well for a diverse range of empirical data drawn from the literature. Because BHG provides a principled quantification of the plausibility of grouping interpretations over a wide range of grouping problems, we argue that it provides an appealing unifying account of the elusive Gestalt notion of *Prägnanz*.

Keywords: perceptual grouping, Bayesian inference, hierarchical representation, visual perception, computational model

Perceptual grouping is the process by which image elements are organized into distinct clusters or objects. The problem of grouping is inherently difficult because the system has to choose the "best" grouping interpretation among an enormous number of candidates (the number of possible partitions of N elements, e.g., grows exponentially with N). Moreover, exactly what makes one interpretation "better" than another is a notoriously subtle problem, epitomized by the ambiguity surrounding the Gestalt term *Prägnanz*, usually translated as "goodness of form." Hence, despite an enormous literature (see Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012, for modern reviews), both the underlying goals of perceptual grouping—exactly what the system is trying to accomplish—as well as the computational mechanisms the human system employs to accomplish these goals, are poorly understood.

Many models have been proposed for specific subproblems of perceptual grouping. Models of contour integration, the process by which visual elements are grouped into elongated contours, are particularly well-developed (e.g., Ernst et al., 2012; Field, Hayes, & Hess, 1993; Geisler, Perry, Super, & Gallogly, 2001). However, these models often presuppose a variety of Gestalt principles, such as proximity, and good continuation (Wertheimer, 1923), whose motivations are themselves poorly understood. Figure-ground organization, in which the system interprets qualitative depth relations among neighboring surfaces, is also heavily studied, but again models often invoke a diverse collection of rules, including closure, symmetry, parallelism, and so forth (e.g., Kikuchi & Fukushima, 2003; Sajda & Finkel, 1995). Other problems of perceptual grouping have likewise been studied both computationally and empirically. Notwithstanding the success of many of these models in accounting for the phenomena to which they are addressed, most are narrow in scope and difficult to integrate with other problems of perceptual grouping. An overarching or unifying paradigm for perceptual grouping does not, as yet, exist.

Of course, it cannot be assumed that the various problems of perceptual grouping do, in fact, involve common mechanisms or principles. "Perceptual grouping" might simply be an umbrella term for a set of essentially unrelated though similar processes. (According to Helson, 1933, the Gestalt literature identified 114 distinct grouping principles; see also Pomerantz, 1986). Nevertheless, it has long been observed that many aspects of perceptual grouping seem to involve similar or analogous organizational preferences (Kanizsa, 1979), suggesting the operation of a common underlying computational mechanism—as reflected in Gestalt attempts to unify grouping via Prägnanz (Wertheimer, 1923). But attempts to give a concrete definition to this term have not converged on a clear account. Köhler (1950) sought an explanation of Prägnanz at the neurophysiological level, whereas van Leeuwen (1990a, 1990b) argued that it should reflect a theory of mental representation. Despite this long history, the idea of an integrated computational framework for perceptual grouping, in which each

This article was published Online First August 31, 2015.

Vicky Froyen, Jacob Feldman, and Manish Singh, Department of Psychology, Center for Cognitive Science, Rutgers University.

Vicky Froyen is now at the Graduate Center for Vision Research, SUNY College of Optometry. This research was supported by National Institutes of Health EY021494 to Jacob Feldman and Manish Singh, National Science Foundation DGE 0549115 (Rutgers IGERT in Perceptual Science), and a Fulbright fellowship to Vicky Froyen. We are grateful to Steve Zucker, Melchi Michel, John Wilder, Brian McMahan, and the members of the Feldman and Singh lab for their many helpful comments.

Correspondence concerning this article should be addressed to Vicky Froyen, 33 West 42nd Street, New York, NY 10036. E-mail: vfroyen@sunyopt.edu

of a variety of narrower grouping problems could be understood as a special case, remains elusive.

A number of unifying approaches have centered around the idea of simplicity, often referred to as the *minimum principle*. Hochberg and McAlister (1953) and Attneave (1954) were the first to apply ideas from information theory to perceptual organization, showing how the uncertainty or complexity of perceptual interpretations could be formally quantified, and arguing that the visual system chooses the simplest among available alternatives. Leeuwenberg (1969) developed a more comprehensive model of perceptual complexity based on the idea of the length of the stimulus description in a fixed coding language, now referred to as structural information theory. More recently, his followers have applied similar ideas to a range of grouping problems, including amodal completion (e.g., Boselie & Wouterlood, 1989; van Lier, van der Helm, & Leeuwenberg, 1994, 1995) and symmetry perception (e.g., van der Helm & Leeuwenberg, 1991, 1996). The concreteness of the formal complexity minimization makes these models a clear advance over the often vague prescriptions of the Gestalt theory. But they suffer from a number of fundamental problems, including the ad hoc nature of the fixed coding language adopted, the lack of a tractable computational procedure, and a variety of other problems (Wagemans, 1999).

Recently, a number of Gestalt problems have been modeled in a Bayesian framework, in which degree of belief in a given grouping hypothesis is associated with the posterior probability of the hypothesis conditioned on the stimulus data (e.g., Kersten, Mammasian, & Yuille, 2004). Contour integration, for example, has been shown to conform closely to a rational Bayesian model given suitable assumptions about contours (Claessens & Wagemans, 2008; Elder & Goldberg, 2002; Ernst et al., 2012; Feldman, 1997a, 2001; Geisler et al., 2001). Similarly, the Gestalt principle of good continuation has been formalized in terms of Bayesian extrapolation of smooth contours (Singh & Fulvio, 2005, 2007). But the Bayesian approach has not yet been extended to the more difficult problems of perceptual organization, and a unifying approach has not been developed. A more comprehensive Bayesian account of perceptual grouping would require a way of expressing grouping interpretations in a probabilistic language, and tractable techniques for estimating the posterior probability of each interpretation. The Bayesian framework is well known to relate closely to complexity minimization, essentially because maximization of the Bayesian posterior is related to minimization of the description length (DL; i.e., the negative log of the posterior; see Chater, 1996; Feldman, 2009; Wagemans, Feldman, et al., 2012). The Bayesian approach, too, is often argued to solve the bias-variance problem, giving a solution with optimal complexity given the data and prior knowledge (MacKay, 2003). Hence, a comprehensive Bayesian account of grouping promises to shed light on the nature of the minimum principle.

In what follows, we introduce a unified Bayesian framework for perceptual grouping, called Bayesian hierarchical grouping (BHG), based on the idea of mixture models. Mixture models have been used to model a wide variety of problems, including motion segmentation (Gershman, Jäkel, & Tenenbaum, 2013; Weiss, 1997), visual short-term memory (Orhan & Jacobs, 2013). and categorization (Rosseel, 2002; Sanborn, Griffiths, & Navarro, 2010). But other than our own work on narrower aspects of the problem of perceptual grouping (Feldman et al., 2013; Feldman, Singh, & Froyen, 2014; Froyen, Feldman, & Singh, 2010, 2015), mixture models have yet to be applied to this problem in general. BHG builds on our earlier work but goes considerably farther in encompassing a broad range of grouping problems and introducing a coherent and consistent algorithmic approach. BHG uses agglomerative clustering techniques (in much the same spirit as Ommer & Buhmann, 2003, 2005, in computer science) to estimate the posterior probabilities of hierarchical grouping interpretations. We illustrate and evaluate BHG by demonstrating it on a diverse range of classical problems of perceptual organization, including contour integration, part decomposition and shape completion, and also show that BHG generalizes naturally beyond these problems. In contrast to many past computational models of specific grouping problems, BHG does not take Gestalt principles for granted as premises, but attempts to consider, in a more principled way, exactly what is being estimated when the visual system decomposes the image into distinct groups. To preview, our proposal is that the visual system assumes that the image is composed of a combination (mixture) of distinct generative sources (or objects), each of which generates some visual elements via a stochastic process. In this view, the problem of perceptual grouping is to estimate the nature of these generating sources, and thus to decompose the image into the coherent groups, each of which corresponds to a distinct generative source.

In what follows we will first outline the computational framework of BHG. We then show how BHG accounts for a number of distinct aspects of perceptual grouping, relying on a variety of data drawn from the literatures of contour integration, part decomposition and shape completion, as well as some "instant psychophysics" (i.e., perceptually natural results on some simple cases).

The Computational Framework

In recent years, Bayesian models have been developed to explain a variety of problems in visual perception. The goal of these models, broadly speaking, is to quantify the degree of belief that ought to be assigned to each potential interpretation of image data. In these models, each possible interpretation $\mathbf{c_j} \in \mathbf{C} = \{\mathbf{c_1} \ldots \mathbf{c_J}\}$ of an image D is associated with posterior probability $p(\mathbf{C}|D)$, which, according to Bayes' rule, is proportional to the product of a prior probability $p(\mathbf{C})$ and likelihood $p(D|\mathbf{C})$ (for introductions, see Feldman, 2014; Kersten et al., 2004; Mamassian & Landy, 2002). Similarly, we propose a framework in which perceptual grouping can be viewed as a rational Bayesian procedure by regarding it as a kind of *mixture estimation* (see also Feldman et al., 2014; Froyen et al., 2015). In this framework, the goal is to use Bayesian inference to estimate the most plausible decomposition of the image configuration into constituent "objects."[1] Here, we give a brief overview of the approach, sufficient to explain the applications that follow, with mathematical details left to the Appendixes.

[1] Note that a full decision-theoretic treatment would add a utility to each hypothesis (a loss function) in order to decide among actions (Maloney & Mamassian, 2009). In this article we focus more narrowly on the determination of belief in grouping hypotheses, and defer the broader problem of action selection to future research.

An Image Is a Mixture of Objects

A mixture model is a probability distribution that is composed of the weighted sum of some number of distinct component distribution or sources (McLachlan & Basford, 1988). That is, a mixture model is a combination of distinct data sources, all mixed together, without explicit labels—like a set of unlabeled color measurements drawn from a sample of apples mixed with a sample of oranges. The problem for the observer, given data sampled from the mixture, is to decompose the data set into the most likely combination of components (e.g., a set of reddish things and a set of orange things). Because the data derive from a set of distinct sources, the distribution of values can be highly multimodal and heterogeneous, even if each source has a simple unimodal form. The fundamental problem for the observer is to estimate the nature of the sources (e.g., parameters such as means and variances) while simultaneously estimating which data originate from which source.

The main idea behind our approach is to assume that the *image itself is a mixture of objects*.[2] That is, we assume that the visual elements we observe are a sample drawn from a mixture model in which each *object* is a distinct data-generating source. Technically, let $D = \{x_1 \ldots x_N\}$ denote the image data (with each x_n representing a two-dimensional vector in $\mathbb{R}^2$, e.g., the location of a visual element). We assume that the image (really, the probability distribution from which image elements are sampled) is a sum of K components

$$p(x_n|\phi) = \sum_{k=1}^{K} p(x_n|\theta_k)p(c_n = k|\mathbf{p}). \quad (1)$$

In this expression, $c_n \in \mathbf{c} = \{c_1 \ldots c_N\}$ is the assignment of data x_n to source components, $\mathbf{p}$ is a parameter vector of a multinomial distribution with $p(c_n = k|\mathbf{p}) = p_k$, θ_k are the parameters of the k-th object, and $\phi = \{\theta_1, \ldots, \theta_K, \mathbf{p}\}$. The observer's goal is to estimate the posterior distribution over the hypotheses $\mathbf{c_j}$, assigning a degree of belief to each way of decomposing the image data into constituent objects—that is, to group the image.

The framework derives its flexibility from the fact that the objects (generating sources) can be defined in a variety of ways depending on assumptions and context. In a simple case like dot clustering, the image data might be assumed to be generated by a mixture of Gaussian objects, that is, simple clusters defined by a mean μ_k and covariance matrix Σ_k (see Froyen et al., 2015). Later, we introduce a more elaborate object definition appropriate for other types of grouping problems, such as contour integration and part decomposition. For contour integration, for example, we define an image as a mixture of contours, with contours formalized as elongated mixture components. For part decomposition, we define a shape as a "mixture of parts." In fact, all the object definitions we introduce in this article are variations of a single flexible object class, which, as we will show, can be tailored to generate image data in the form of contours, clusters, or objects parts. So notwithstanding the differences among these distinct kinds of perceptual grouping, in our framework, they are all treated in a unified manner.

To complete the Bayesian formulation, we define priors over the object parameters $p(\theta|\beta)$ and over the mixing distribution $p(\mathbf{p}|\alpha)$. The first of these defines our expectations about what "objects" in our context tend to look like. The second is more technical, defining our expectations about how distinct components tend to "mix." (It is the natural conjugate prior for the mixing distribution, the Dirichlet distribution with parameter α.) Using these two priors, we can rewrite the mixture model (Equation 1) to define the probability of a particular grouping hypothesis $\mathbf{c_j}$. The likelihood of a particular grouping hypothesis is obtained by marginalizing over the parameters (θ_k, $p(D|\mathbf{c_j}, \beta) = \int \Pi_{n=1}^{N} p(x_n|\theta_{c_n}) \prod_{k=1}^{K} p(\theta_k|\beta) d\theta$). This yields the posterior of interest

$$p(\mathbf{c_j}|D, \alpha, \beta) \propto p(D|\mathbf{c_j}, \beta)p(\mathbf{c_j}|\alpha), \quad (2)$$

where $p(\mathbf{c_j}|\alpha) = \int p(\mathbf{c_j}|\mathbf{p})p(\mathbf{p}|\alpha)d\mathbf{p}$ is a standard Dirichlet integral (see Rasmussen, 2000, for derivation). This equation defines the degree of belief in a particular grouping hypothesis.

Note that the posterior can be decomposed into two intuitive factors: the likelihood $p(D|\mathbf{c_j}, \beta)$, which expresses how well the grouping hypothesis $\mathbf{c_j}$ fits the image data D, and a prior, which, in effect, quantifies the *complexity* of the grouping hypothesis $\mathbf{c_j}$. As in all Bayesian frameworks, these two components trade off. Decomposing the image into more groups allows each group to fit its constituent image data better, but at a cost in complexity; decomposing the image into fewer, larger groups is simpler, but does not fit the data as well. In principle, Bayes' rule allows this tradeoff to be optimized, allowing the observer to find the right balance between a simple grouping interpretation and a reasonable fit to the image data.

Unfortunately, as in many Bayesian approaches, computing the full posterior over grouping hypotheses $\mathbf{c_j}$ can become intractable as N increases, even for a fixed number of components K (Gershman & Blei, 2012). Moreover, we do not generally know the number K of clusters, meaning that K must be treated as a free parameter to be estimated, making the problem even more complex. To accommodate this, we extend the finite mixture model into a so-called Dirichlet process mixture model, commonly used in machine learning and statistics (Neal, 2000). Several approximation methods have been proposed to compute posteriors for these models, such as Markov-Chain Monte Carlo (McLachlan & Peel, 2004) or variational methods (Attias, 2000). In the current article, we adopt a method introduced by Heller and Ghahramani (2005), called Bayesian hierarchical clustering (BHC), to the problem of perceptual grouping, resulting in the framework we refer to as BHG.

One of the main features of BHG is that it produces a *hierarchical* representation of the organization of image elements. Of course, the idea that perceptual organization tends to be hierarchical has a long history (e.g., Baylis & Driver, 1993; Lee & Mumford, 2003; Marr & Nishihara, 1978; Palmer, 1977; Pomerantz, Sager, & Stoever, 1977). Machine learning, too, has often employed hierarchical representations of data (see Murphy, 2012, for a useful overview). Formally, a hierarchical structure corresponds to a tree in which the root node represents the image data at the most global level, that is, the grouping hypothesis that postulates that all image data are generated by one underlying object. Subtrees then describe finer and more local relations between image data, all the way down to the

[2] It should be noted that by "objects," we mean data-generating sources in the image. In this article we address the problem of perceptual grouping in two dimensions. Our framework, however, is readily extended to 3D (El-Gaaly, Froyen, Elgammal, Feldman, & Singh, 2015).

leaves, which each explain only one image datum x_n. Although this formalism has become increasingly popular (Amir & Lindenbaum, 1998; Feldman, 1997b, 2003; Shi & Malik, 2000), no general method currently exists for actually building this tree for a particular image. In this article we introduce a computational framework, BHG, that creates a fully hierarchical representation of a given configuration of image elements.

Bayesian Hierarchical Clustering

In this section we give a slightly technical synopsis of BHG (details can be found in the Appendices A and B). Like other agglomerative hierarchical clustering techniques, BHC (Heller & Ghahramani, 2005) begins by assigning each data point its own cluster, and then progressively merges pairs of clusters to create a hierarchy. BHC differs from traditional agglomerative clustering methods, in that it uses a Bayesian hypothesis test to decide which pair of clusters to merge. The technique serves as a fast approximate inference method for Dirichlet process mixture models (Heller & Ghahramani, 2005).

Given the data set $D = \{x_1 \ldots x_N\}$, the algorithm is initiated with N trees T_i each containing one data point $D_i = \{x_n\}$. At each stage, the algorithm considers merging every possible pair of trees. Then, by means of a statistical test, it chooses two trees T_i and T_j to merge, resulting in a new tree T_k, with its associated merged data set $D_k = D_i \cup D_j$ (Figure 1A). Testing all possible pairs is intractable, so to reduce complexity, BHG considers only those pairs that have "neighboring" data points. Neighbor relations are defined via adjacency in the Delaunay triangulation, a computa-

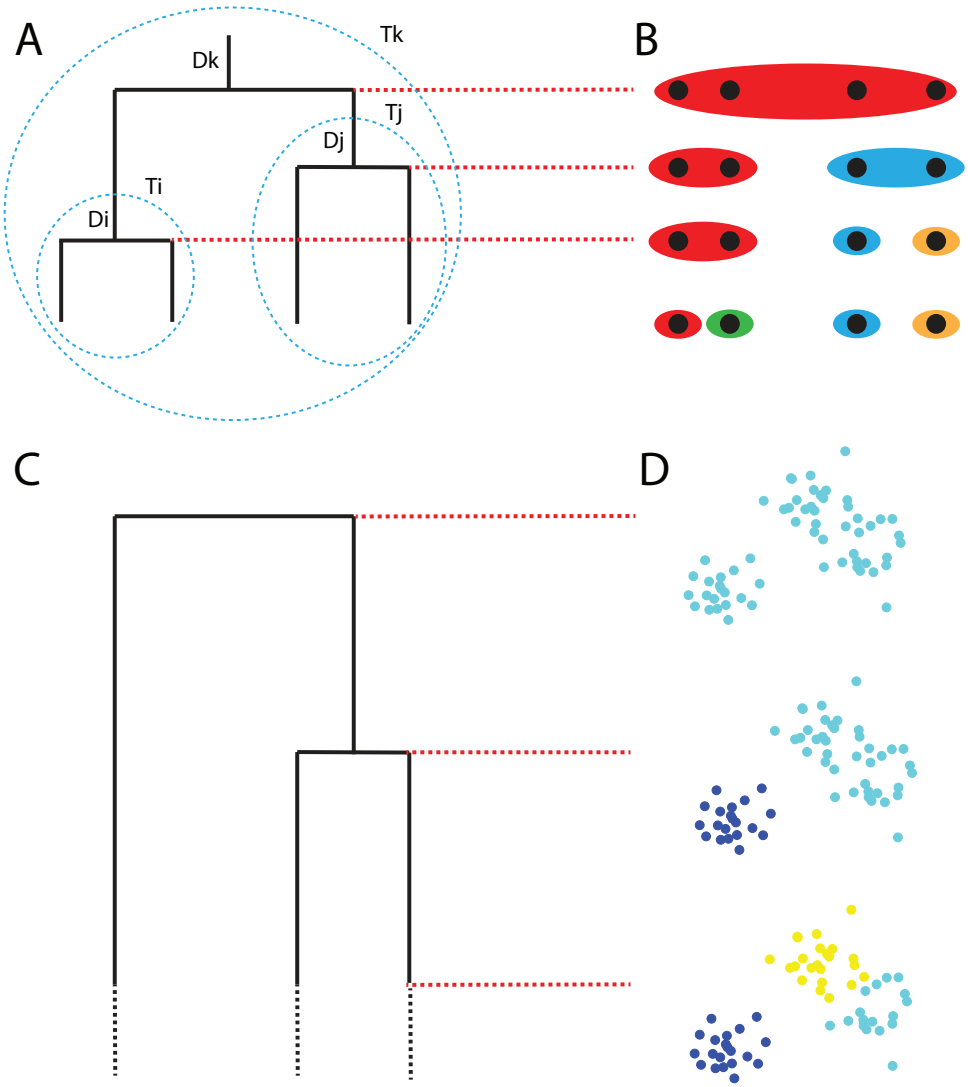

Figure 1. Illustration of the Bayesian hierarchical clustering process. (A) Example tree decomposition (see also Heller & Ghahramani, 2005) for the 1D grouping problem on the right. (B) Tree slices, that is, different grouping hypotheses. (C) Tree decomposition as computed by the clustering algorithm for the dot clusters on the right, assuming bivariate Gaussian objects. (D) Tree slices for the dot clusters. See the online article for the color version of this figure.

tionally efficient method of defining spatial adjacency (de Berg, Cheong, van Kreveld, & Overmars, 2008) that has been found to play a key role in perceptual grouping mechanisms (Watt, Ledgeway, & Dakin, 2008). Considering only neighboring trees substantially reduces computational complexity, from $\mathcal{O}(N^2)$ in the initial phase of Heller & Ghahramani's original algorithm to $\mathcal{O}(N\log(N))$ in ours (see Appendix A).

In considering each merge, the algorithm compares two hypotheses in a Bayesian hypothesis testing framework. The first hypothesis $\mathcal{H}_0$ is that all the data in D_k are generated by only one underlying object $p(D_k|\theta)$, with unknown parameters θ. In order to evaluate the probability of the data given this hypothesis $p(X|\mathcal{H}_0)$, we introduce priors $p(\theta|\beta)$ over the objects as described earlier in order to integrate over the to-be-estimated parameters θ,

$$p(D_k|\mathcal{H}_0) = \int_\theta \prod_{x_n \in D_k} p(x_n|\theta)p(\theta|\beta)d\theta. \quad (3)$$

For simple objects such as Gaussian clusters, this integral can be computed analytically, but with more complex objects, it becomes intractable and approximate methods must be used (see Appendix B).

The second hypothesis, $\mathcal{H}_1$, is the sum of all possible partitionings of D_k into two or more objects. However, again, exhaustive evaluation of all such hypotheses is intractable. The BHC algorithm circumvents this problem by considering only partitions that are consistent with the tree structure of the two trees T_i and T_j to be merged. For example, for the tree structure in Figure 1A, the possible tree-consistent partitionings are shown Figure 1B. As is clear from the figure, this constraint eliminates many possible partitions. The probability of the data under $\mathcal{H}_1$ can now be computed by taking the product over the subtrees $p(D_k|\mathcal{H}_1) = p(D_i|T_i)p(D_j|T_j)$. As will become clear, $p(D_i|T_i)$ can be computed recursively as the tree is constructed.

To obtain the marginal likelihood of the data under the tree T_k, we need to combine $p(D_k|\mathcal{H}_0)$ and $p(D_k|\mathcal{H}_1)$. This yields the probability of the data integrated across all possible partitions, including the one-object hypothesis $\mathcal{H}_0$. Weighting these hypotheses by the prior on the one-object hypothesis $p(\mathcal{H}_0)$ yields an expression for $p(D_i|T_i)$,

$$p(D_k|T_k) = p(\mathcal{H}_0)p(D_k|\mathcal{H}_0) + [1 - p(\mathcal{H}_0)]p(D_i|T_i)p(D_j|T_j). \quad (4)$$

Note that $p(\mathcal{H}_0)$ is also computed bottom up as the tree is built, and is based on a Dirichlet process prior (Equation 6; for details, see Heller & Ghahramani, 2005; we discuss this prior in greater depth in text that follows). Finally, the probability of the merged hypothesis $p(\mathcal{H}_0|D_k)$ can be found via Bayes' rule,

$$p(\mathcal{H}_0|D_k) = \frac{p(\mathcal{H}_0)p(D_k|\mathcal{H}_0)}{p(D_k|T_k)}. \quad (5)$$

This probability is then computed for all Delaunay-consistent pairs, and the pair with the highest merging probability is merged. In this way, the algorithm greedily[3] builds the tree until all data are merged into one tree. Given the assumed generative models of objects, this tree represents the most "reasonable" decomposition of the image data into distinct objects.

Examples of such trees are shown in Figure 1. Figure 1A illustrates a tree for the classical one-dimensional grouping problem (Wertheimer, 1923; Figure 1B). The algorithm first groups the closest pair of image elements, then continues as described earlier until all image elements are incorporated into the tree. Figure 1C and D similarly illustrate two-dimensional dot clustering (Froyen et al., 2015). The results shown in the figure assume objects consisting of bivariate Gaussian distributions of image elements.

Tree slices and grouping hypotheses. During the construction of the tree, one can greedily find an at-least local maximum posterior decomposition by splitting the tree once $p(\mathcal{H}_0|D_k) < .5$. However, we are more often interested in the distribution over all the possible grouping hypotheses rather than choosing a single winner. In order to do so, we need to build the entire tree, and subsequently take what are called *tree slices* at each level of the tree (Figure 1B and D), and compute their respective probabilities $p(\mathbf{c_j}|D, \alpha, \beta)$. Because the current algorithm proposes an approximation of the Dirichlet process mixture model (see Heller & Ghahramani, 2005, for proof), we make use of a Dirichlet process prior (Rasmussen, 2000, and independently discovered by Anderson, 1991) to compute the posterior probability of each grouping hypothesis (or *tree slice*). This prior is defined as

$$p(\mathbf{c_j}|\alpha) = \frac{\Gamma(\alpha)\alpha^K \prod_{k=1}^K \Gamma(n_k)}{\Gamma(N+\alpha)}, \quad (6)$$

where n_k is the number of data points explained by object with index k. When $\alpha > 1$, there is a bias toward more objects, each explaining a small number of image data; whereas when $0 < \alpha < 1$, there is a bias toward fewer objects, each explaining a large number of image data. Inserting Equation 6 into Equation 2, we can compute the posterior distribution across all tree-consistent decompositions of data D:

$$p(\mathbf{c_j}|D, \alpha, \beta) \propto p(\mathbf{c_j}|\alpha) \prod_{k=1}^K p(D_k|\beta), \quad (7)$$

where $p(D_k|\beta)$ is the marginal likelihood over θ for the data in cluster k of the current grouping hypothesis.

Prediction and completion. For any grouping hypothesis, BHG can compute the probability of a new point x^* given the existing data D, called the posterior predictive distribution $p(x^*|D, \mathbf{c_j})$.[4] As in Equation 1, the new datum is generated from a mixture model consisting of the K components comprised in this particular grouping hypothesis. More specifically, new data are generated as a weighted sum of predictive distributions $p(x^*|D_k) = \int p(x^*|\theta)p(\theta|D_k, \beta)d\theta$, where D_k is the data associated with object k),

$$p(x^*|D, \mathbf{c_j}) = \sum_{k=1}^K p(x^*|D_k)\pi_k. \quad (8)$$

Here, π_k is the posterior predictive of the Dirichlet prior defined as $\pi_k = (\alpha + n_k)/\sum_{i=1}^K (\alpha + n_i)$ (see Bishop, 2006, p. 478, for a derivation).

The posterior predictive distribution has a particularly important interpretation in the context of perceptual grouping: It allows the model to make predictions about missing data, such as how shapes

[3] A greedy algorithm is one that makes a choice at each stage of computation based on available information, and does not revisit choices once made.

[4] In standard Bayesian terminology, the posterior distribution assigns probability to hypotheses, and the posterior predictive distribution assigns probability to unobserved data.

continue behind occluders (amodal completion). Several examples are shown in the Results section.

The Objects

In the BHG framework, objects (generating sources) can be defined in a way suitable to the problem at hand. For a simple dot clustering problem, we might assume Gaussian objects governed by a mean μ_k and a covariance matrix Σ_k. We have previously found that such a representation accurately and quantitatively predicts human cluster enumeration judgments (Froyen et al., 2015). Figure 1C to D shows sample BHG output for this problem. In this problem, because the prior and likelihood are conjugate, the marginal $p(D|H_0)$ can be computed analytically.

However, more complex grouping problems, such as part decomposition and contour integration, call for a more elaborate object definitions. As a general object model, assume that objects are represented as B-spline curves $G = \{g_1 \ldots g_K\}$ (see Figure 2), each governed by a parameter vector $\mathbf{q_k}$. (B-splines [see Farouki & Hinds, 1985] are a convenient and widely used mathematical representation of curves). For each spline, data points x_n are sampled from univariate Gaussian distributions perpendicular to this curve,

$$p(x_n|\theta_k) = \mathcal{N}(d_n|\mu_k, \sigma_k), \tag{9}$$

where $d_n = \|x_n - g_k(n)\|$ is the distance between the data point x_n and its perpendicular projection to the curve $g_k(n)$, also referred to as the *rib length*.[5] μ_k and σ_k are, respectively, the mean rib length and the rib length variance for each curve. Put together, the parameter vector for each component is defined as $\theta_k = \{\mu_k, \sigma_k, \mathbf{q_k}\}$. Figure 2 shows how this object definition yields a wide variety of forms, ranging from contour-like objects (when $\mu_k = 0$; Figure 2A) to axial shapes (when $\mu_k > 0$, making "wide" shapes; Figure 2B). Objects with $\mu_k = 0$ but larger variance σ_k will tend to look like dots generated from a cluster. For elongated objects, note that the generative function is symmetric along the curve (see Figure 2), constraining objects to obey a kind of local symmetry (Blum, 1973; Brady & Asada, 1984; Feldman & Singh, 2006; Siddiqi, Shokoufandeh, Dickinson, & Zucker, 1999).

In the Bayesian framework, the definition of an object class requires a prior $p(\theta|\beta)$ on the parameter vector θ governing the generative function (Equation 9). In the illustrations in this article we use a set of priors that create a simple but flexible object class. First, we introduce a bias on the "compactness" of the objects by adopting a prior on the squared arc length F_{k1} (also referred to as *elastic energy*) of the generating curve k, $F_{k1} \sim \exp(\lambda_1)$, resulting in a preferences for shorter curves. Similarly, we introduce a prior governing the "straightness" of the objects by introducing a prior over the squared total curvature F_{k2}, also referred to as bending energy, of the curve k (Mumford, 1994), $F_{k2} \sim \exp(\lambda_2)$, resulting in a preference for straighter rather than curved objects (for computations of both F_{k1} and F_{k2}, see Appendix B). We also assume a normal-inverse-chi-squared prior (conjugate of the normal distribution) over the function that generates the data points from the curves (Equation 9), with parameters $\{\mu_0, \kappa_0, \nu_0, \sigma_0\}$. μ_0 is the expectation of the rib length and κ_0 defines how strongly we believe this; σ_0 is the expectation of the variance of the rib length, and ν_0 defines how strongly we believe this. Taken together, these priors and their hyperparameter vector $\beta = \{\lambda_1, \lambda_2, \mu_0, \kappa_0, \nu_0, \sigma_0\}$ induce a simple but versatile class of objects suitable for a wide range of spatially defined perceptual grouping problems.

In summary, the BHG framework makes a few simple assumptions about the form of objects in the world, and then estimates a hierarchical decomposition of the image data into objects (or, more correctly, assigns posterior probabilities to all potential decompositions within a hierarchy). In what follows, we give examples of results drawn from a variety of grouping problems, including contour integration, part decomposition, and shape completion, and show how the approach explains a number of known perceptual phenomena and fits data drawn from the literature.

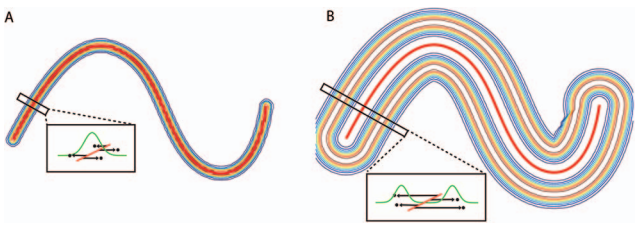

Figure 2. The generative function of our model depicted as a field. Ribs sprout perpendicularly from the curve (red), and the length they take on is depicted by the contour plot. (A) For contours, ribs are sprouted with a μ close to zero, resulting in a Gaussian fall-off along the curve. (B) For shapes, ribs are sprouted with $\mu > 0$, resulting in a band surrounding the curve. See the online article for the color version of this figure.

Results

We next show results from BHG computations for a variety of perceptual grouping problems, including contour integration, contour grouping, part decomposition, and shape completion.

Contours

In what follows, we show several examples of how BHG can be applied to contour grouping. In the first three examples, the problem in effect determines the set of alternative hypotheses, and we show how BHG gauges the relative degree of belief among the available grouping interpretations. In the rest of the examples, we illustrate the capabilities of BHG more fully by allowing it to generate the hypotheses themselves rather than choosing from a given set of alternatives.

To apply BHG to the problem of grouping contours, we need to define an appropriate data-generating model. To model contours, we use the generic object class defined earlier with the hyperparameters set in a way that expresses the idea that contours are narrow, elongated objects ($\mu_0 = 0$ and $\kappa = 1 \times 10^4$, $\sigma_0 = .01$, and $\nu_0 = 20$) that do not extend very far or bend very much ($\lambda_1 = 0.16$ and $\lambda_2 = 0.05$). Finally we set the parameter of the Dirichlet process prior to $\alpha = .1$, expressing a modest bias toward keeping objects together rather than splitting them up. Naturally, alterna-

[5] Random deviations from the central axis are called "ribs" because they sprout laterally from a "skeleton"; see fuller discussion of Feldman and Singh's (2006) skeletal framework in the text that follows.

tive setting of these parameters may be appropriate for other contexts.

Dot lattices. Ever since Wertheimer (1923), researchers have used dot lattices (see Figure 3) to study the Gestalt principle of grouping by proximity (e.g., Kubovy & Wagemans, 1995; Oyama, 1961; Zucker, Stevens, & Sander, 1983). A dot lattice is a collection of dots arranged on a grid-like structure, invariant to at least two translations a (with length $|a|$), and b (with length $|b| \geq |a|$). These two lengths and the angle γ between them defines the dot lattice, and influence how the dots in the lattice are perceived to be organized. A number of proposals have been made about grouping in dot lattices, with one of the few rigorously defined mathematical models being the pure-distance law (Kubovy, Holcombe, & Wagemans, 1998; Kubovy & Wagemans, 1995), which dictates that the tendency to group one way or the other will be determined solely by the ratio $|v|/|a|$. Here, $|v|$ can be $|b|$ or the length of the diagonals of the rectangle formed by $|a|$ and $|b|$.

In the computational experiments that follow, the angle γ was set to 90°, while also keeping the orientation of the entire lattice fixed, allowing us to restrict our attention to the two dominant percepts of rows and columns (see Figure 3). In these lattices, $|a|$ (interdot distance in the horizontal direction) is fixed, and $|b|$ (interdot distance in the vertical direction) is varied such that $|b|/|a|$ ranges from 1 to 2. The BHG framework allows grouping to be modeled as follows.

Following Equation 7, the posterior distribution over the two hypotheses, rows (c_h) and columns (c_v), can then be computed as

$$p(\boldsymbol{c_v}|D) = \frac{p(D|\beta, \boldsymbol{c_v})p(\boldsymbol{c_v}|\alpha)}{p(D|\beta, \boldsymbol{c_h})p(\boldsymbol{c_h}|\alpha) + p(D|\beta, \boldsymbol{c_v})p(\boldsymbol{c_v}|\alpha)}. \quad (10)$$

The results for a 5 × 5 lattice are shown in Figure 3A. We plotted the log of the posterior ratio log $p(\boldsymbol{c_v}|D)/p(\boldsymbol{c_h}|D)$ as a function of the ratio $|b|/|a|$ (similar to Kubovy et al., 1998). Figure 3A to B shows how the posterior ratio decreases monotonically with the ratio $|b|/|a|$, consistent with human data (Kubovy & Wagemans, 1995). In other words, the posterior probability of the horizontal contours (rows) interpretation increases monotonically with $|b|/|a|$. The exact functional form of the relationship depends on the details of the object definition. If our object definition includes an error on arc length, then this implies a linear relationship between the ratio $|b|/|a|$, and the log posterior ratio is precisely as predicted by the pure-distance law (Figure 3A). The examples in this article use a quadratic penalty of arc length (see Appendix B) for computational convenience, in which case we get a slightly nonlinear prediction depicted in Figure 3B.

Good continuation. Good continuation, the tendency for collinear or nearly collinear sets of visual items to group together, is another Gestalt principle that has been extensively studied (Feldman, 1997a, 2001; Smits & Vos, 1987; Smits, Vos, & Van Oeffelen, 1985). Here, we show that BHG can make quantitatively accurate predictions about the strength of tendency toward collinear grouping by comparing its predictions to measured judgments of human subjects. Feldman (2001) conducted an experiment in which subjects were shown dot patterns consisting of six dots (e.g., Figure 4A and B) and then

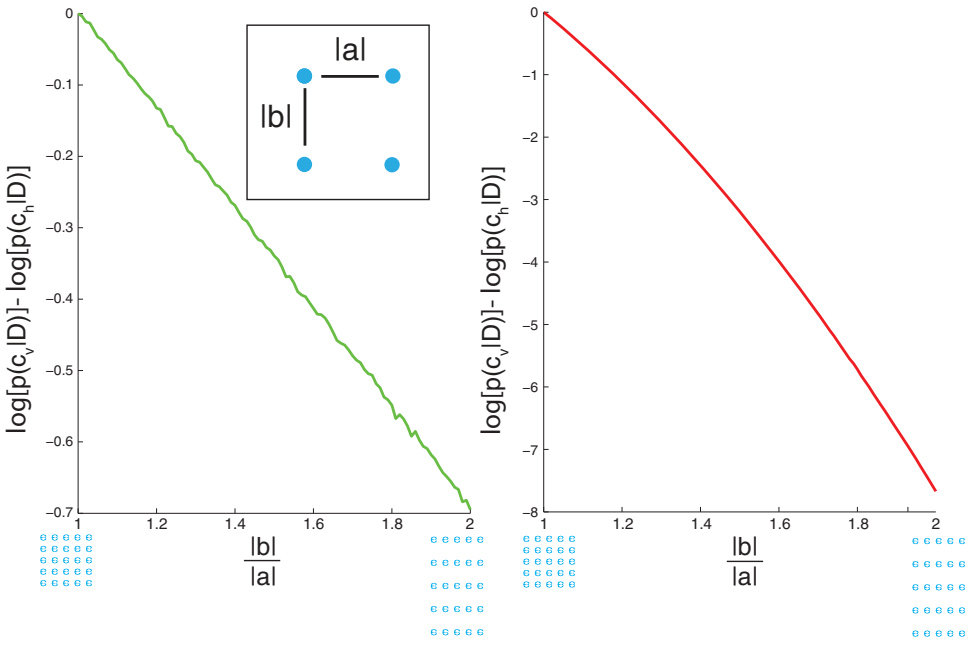

Figure 3. Bayesian hierarchical grouping predictions for simple dot lattices. As input, the model received the location of the dots as seen in the bottom two rows, in which the ratio of the vertical |b| over the horizontal |a| dot distance was manipulated. The graph on top shows the log posterior ratio of seeing vertical versus horizontal contours as a function of the ratio |b|/|a|. Left plot: Object definition included error on arc length. Right plot: Object definition included a quadratic error on arc length. See the online article for the color version of this figure.

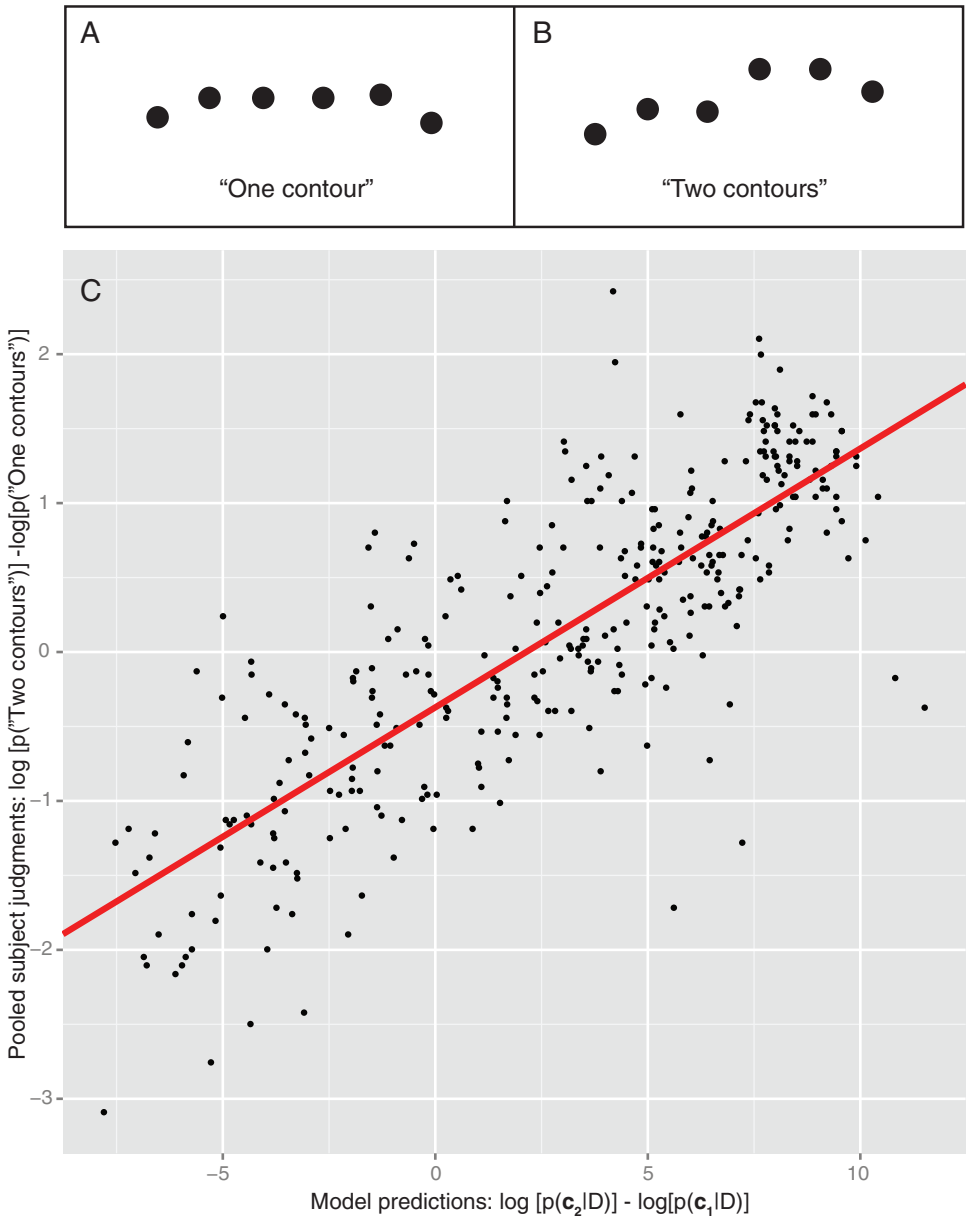

Figure 4. Bayesian hierarchical grouping's performance on data from Feldman (2001). (A, B) Sample stimuli with likely responses (stimuli not drawn to scale). (C) Log odds of the pooled subject responses plotted as a function of the log posterior ratio of the model log $p(\mathbf{c}_2|D) - \log p(\mathbf{c}_1|D)$, in which each point depicts one of the 343 stimulus types shown in the experiment. Both indicate the probability of seeing two contours $p(\boldsymbol{c}_2|D)$. Model responses are linearized using an inverse cumulative Gaussian. See the online article for the color version of this figure.

asked to indicate if they saw *one* or *two* contours, corresponding to a judgment about whether the dots were perceived as grouping into a single smooth curve or two distinct curves with an "elbow" (nonsmooth junction) between them. To model these data, we ran BHG on all the original stimuli and computed the probability (Equation 10) of the two alternative grouping hypotheses: $\boldsymbol{c}_1$ (all dots generated by one underlying contour) and $\boldsymbol{c}_2$ (dots generated by two distinct contours).[6] We then compared this posterior to pooled human judgments for all 343 dot configurations considered in Feldman, and found a strong monotonic relationship between model and data (see Figure 4). Because of the binomial nature of both variables, we took the log odds ratio of each. The log odds of the human judgments and log posterior ratio of the model predictions were highly correlated (Likelihood

[6] The hypothesis $\boldsymbol{c}_2$ comprises several subhypotheses listing all the possible ways that these six dots could be subdivided into two contours. Here, we only took into account those hypotheses that would not alter the ordering of the dots (i.e., hypotheses {(1), (2, 3, 4, 5, 6)}, {(1, 2), (3, 4, 5, 6)}, {(1, 2, 3), (4, 5, 6)}, {(1, 2, 3, 4), (5, 6)}, and {(1, 2, 3, 4, 5), (6)}). Because these are disjoint, $p(\boldsymbol{c}_2|D)$ is the sum over all these hypotheses.

Ratio Test [LRT] = 229.3, degrees-of-freedom [*df*] = 1, $p < .001$, $R^2 = 0.6650$; Bayes Factor [BF] = $2.93775e + 77$).[7]

Association field. The interaction between good continuation and proximity is often expressed as an "association field" (Field et al., 1993; Geisler et al., 2001), and is closely related to the idea of cocircular support neighborhoods (Parent & Zucker, 1989; Zucker, 1985). Given the object definition, the BHG framework automatically introduces a Bayesian interaction between proximity and collinearity without any special mechanism. To illustrate this effect, we manipulated the distance D and relative orientation θ of two segments each composed of five dots (Figure 5A). To quantify the degree of grouping between the two segments, we considered two hypotheses: one $\boldsymbol{c_1}$, in which all 10 dots are generated by a single underlying contour, and another $\boldsymbol{c_2}$, in which the two segments are each generated by distinct contours. The posterior probability of the former $p(\boldsymbol{c_1}|D)$, computed using Equation 10, expresses the tendency for the two segments to be perceptually grouped together in the BHG framework (Figure 5B). As can be seen in the figure, the larger the angle between the two segments and/or the further they are from each other, the less probable $\boldsymbol{c_1}$ becomes. The transition between ungrouped and grouped segments can be seen in Figure 5B as a transition from blue (low $p[\boldsymbol{c_1}|D]$) to red (high $p[\boldsymbol{c_1}|D]$).

Contour grouping. The preceding examples illustrate how BHG gauges the degree to which a given configuration of dots coheres as perceptual group. But the larger and much more difficult problem of perceptual grouping is to determine how elements should be grouped in the first place—that is, which of the vast number of qualitatively distinct ways of grouping the observed elements is the most plausible. Here, we illustrate how the framework can generate grouping hypotheses by means of its hierarchical clustering machinery and estimate the posterior distribution over those hypotheses. We first ran BHG on a set of simple edge configurations (see Figure 6). One can see that the framework decomposes these into intuitive segments at each level of the hierarchy. Figure 6A gives an illustration of how the MAP (*maximum a posteriori*) hypothesis is that all edges are generated by the same underlying contour, whereas the hypothesis one step down the hierarchy segments it into two intuitive segments. The latter hypothesis, however, has a lower posterior. Another example of an intuitive hierarchy can be seen in Figure 6D, in which the MAP estimate consists of three segments. The decomposition one level up (the two-contour hypothesis) joins the two segments that are abutting together, but has a lower posterior. These simple cases show that the model can build up an intuitive space of grouping hypotheses and assign them posterior probabilities.

Figure 7 shows a set of more challenging cases in which dots are arranged in a variety of configurations in the plane. The figure shows the MAP interpretation drawn from BHG (color coded), meaning the single interpretation assigned highest posterior. In most cases, the decomposition is highly intuitive. In Figure 7B, the long vertical segment is broken into two parts. Although this might seem less intuitive, it follows from the assumptions embodied in the object definition. Specifically, the object definition includes a penalty for the length of the curve in the form (λ_1), which is applied to the entire curve. Exactly how dot spacing influences contour grouping is part of a more complex question of how spacing and regularity of spacing influence grouping (e.g., Feldman, 1997a; Geisler et al., 2001).

Parts of Objects

The decomposition of whole objects into perceptually distinct parts is an essential aspect of shape interpretation with an extensive literature (see, e.g., Biederman, 1987; Cohen & Singh, 2006; De Winter & Wagemans, 2006; Hoffman & Richards, 1984; Palmer, 1977; Singh & Hoffman, 2001). Part decomposition is known to be influenced by a wide variety of specialized factors and rules, such as the minima rule (Hoffman & Richards, 1984), the short-cut rule (Singh, Seyranian, & Hoffman, 1999), and limbs and necks (Siddiqi & Kimia, 1995). But each of these rules has known exceptions and idiosyncrasies (see De Winter & Wagemans, 2006; Singh & Hoffman, 2001), and no unifying account is known. Part decomposition is conventionally treated as a completely separate problem from the grouping problems considered in this article. But as we now show, it can be incorporated elegantly into the BHG framework by treating it as a kind of grouping problem, in which we aim to group the elements that make up the object's boundary into internally coherent components. When treated this way, part decomposition seems to follow the same basic principles—for example, evaluation of the strength of alternate hypotheses via Bayes' rule—as do other kinds of perceptual grouping.

Feldman and Singh (2006) proposed an approach to shape representation in which shape parts are understood as stochastic mixture components that generate the observed shape. Their approach is based on the idea of the *shape skeleton*, a probabilistic generalization of Blum's (1967, 1973) medial axis representation. In this framework, the shape skeleton is a hierarchically organized set of axes, each of which generates contour elements in a locally symmetric fashion laterally from both sides. As mentioned earlier, the random lateral vector is referred to as a *rib* (because it sprouts sideways from a skeleton), and the distance between the contour element and the axis is called the *rib length*. In the Feldman and Singh (2006) framework, the goal is to estimate the skeleton that, given an observed set of contour elements, is mostly likely to have generated the observed elements (the MAP skeleton). Critically, each distinct axial component of the MAP skeleton "explains" or "owns" (is interpreted as having generated) distinct contour elements, so a skeleton estimate entails a decomposition of the shape into parts. In this sense, part decomposition via skeleton estimation is a special case of mixture estimation, in which the shape itself is understood as "a mixture of parts." Skeleton-based part decomposition seems to correspond closely to intuitive part decompositions for many critical cases (Feldman et al., 2013). But the original approach (Feldman & Singh, 2006) lacked a principled and tractable estimation procedure. Here, we show how this approach to part decomposition can be incorporated into the BHG mixture estimation framework, yielding an effective hierarchical decomposition of shapes into parts.

To apply BHG, we begin first by sampling the shape's boundary to create a discrete approximation consisting of a set of edges $D = \{x_1 \ldots x_N\}$. Figure-ground assignment is assumed known (that is, we know which side of the boundary is the interior; the skeleton is assumed to explain the shape "from the inside"). Next,

[7] Both the BF and the LRT were computed by comparing a regression model in which the log odds of the model predictions were taken as a predictor versus an unconditional means model containing only an intercept.

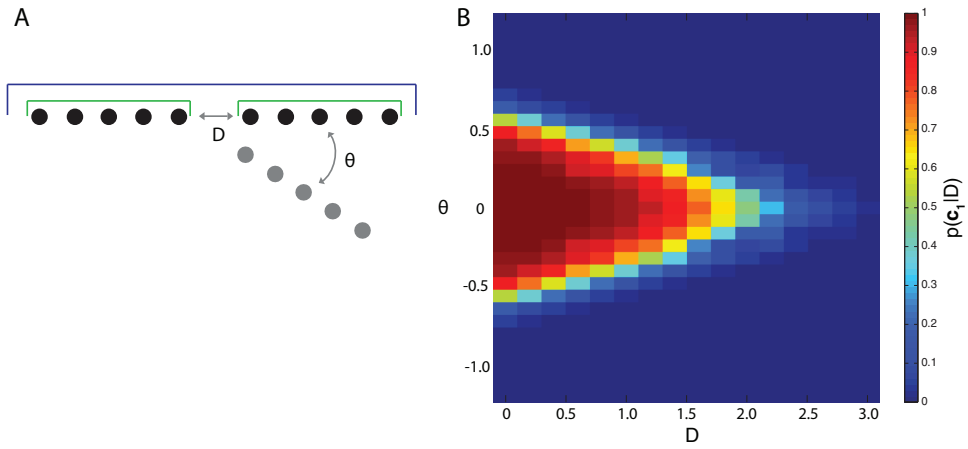

Figure 5. The "association field" between two line segments (each containing five dots) as quantified by Bayesian hierarchical grouping. (A) Manipulation of the distance and angle between these two line segments. The blue line depicts the one-object hypothesis, and the two green lines depict the two-object hypothesis. (B) The association field reflecting the posterior probability of $p(\boldsymbol{c_1}|D)$ of the one-object hypothesis. See the online article for the color version of this figure.

we choose hyperparameters that express our generative model of a part (i.e., reflect our assumptions about what a part looks like). The main difference compared to contours is that we now do not want the mean rib length to be zero. That is, in the case of parts, we assume that the mean rib length can be assigned freely with a slight bias toward shorter mean rib lengths to incorporate the idea that parts are more likely to be narrow ($\mu_0 = 0$; $\kappa_0 = .001$). The remaining generative parameters were set to reflect the assumption that parts should preferably have smooth boundaries ($\sigma_0 = .001$; $\nu_0 = 10$). The hyperparameters biasing the shape of the axes themselves were set to identical values as in the contour integration case ($\lambda_1 = .16$; $\lambda_2 = .05$). Finally, the mixing hyperparameter was set to $\alpha = .001$.

Figure 8 shows an example of BHG applied to a multipart shape. The model finds the most probable (MAP) part decomposition (Figure 8C) and the entire structural hierarchy (Figure 8B). In other words, BHG finds what we argue is the "perceptually natural description" of the shape at different levels of the structural hierarchy (Figure 8D and E). The MAP part decomposition for several shapes of increasing number of parts is shown in Figure 9. Note that each axis represents a part.

Figure 10 shows some sample results. BHG correctly handles difficult cases such as a leaf on a stem (Figure 10A) and dumbbells (Figure 10B), while still maintaining the hierarchical structure of the shapes (Siddiqi et al., 1999). Figure 10D shows a typical animal. In particular, Figure 10C shows the algorithm's robustness to contour noise. This shape, the "prickly pear" from Richards, Dawson, and Whittington (1986) is especially interesting because a qualitatively different type of noise is added to each part of the shape, which cannot be correctly handled by standard scale-space techniques. Furthermore, a desirable side effect of our approach is the absence of ligature regions. A *ligature* is the "glue" that binds two or more parts together (e.g., connecting the leaf to its stem in Figure 10A). Such regions have been identified (August, Siddiqi, & Zucker, 1999) as the cause of internal instabilities in the conventional medial axis transform (Blum, 1967), diminishing their usefulness for object recognition. Past medial axis models had to cope with this problem by explicitly identifying and deleting such regions (e.g., August et al., 1999). In our approach, by contrast, they do not appear at all in the BHG part decompositions, so they do not have to be dealt with separately.

Shape complexity. As a side effect, the BHG model also provides a natural measure of shape complexity, a factor that often arises in experiments but which lacks a principled definition. For example, we have found that shape complexity diminishes the detectability of shapes in noise (Wilder, Singh, & Feldman, 2015). Rissanen (1989) showed that the negative of the logarithm of probability—log p, the DL, provides a good measure of complexity because it expresses the length of the description in an optimal coding language (see also Feldman & Singh, 2006). To compute the DL of a shape, we first integrate over the entire grouping hypothesis space $\boldsymbol{C} = \{\mathbf{c_1} \ldots \mathbf{c_J}\}$:

$$p(D|\alpha, \beta) = \frac{1}{J}\sum_{j=1}^{J} p(D|\beta, \boldsymbol{c_j})p(\boldsymbol{c_j}|\alpha). \tag{11}$$

The DL is then defined as DL $= -\log p(D|\alpha, \beta)$. Figure 9 shows how, as we make a shape perceptually more complex (here, by increasing the number of parts), the DL increases monotonically. This metric is universal to our framework and can be used to express the complexity of any image given any object definition. In general, the DL expresses the complexity of any stimulus, given the generative assumptions about the object classes that are assumed to have generated it.

Part salience. Hoffman and Singh (1997) proposed that representation of object parts is graded, in that parts can vary in the degree to which they appear to be distinct perceptual units within the shape. This variable, called *part salience*, is influenced by a number of geometric factors, including the sharpness of the region's boundaries, its degree of protrusion (defined as the ratio of the part's perimeter to the length of the part cut), and its size relative to the entire shape. Within BHG, we can define part

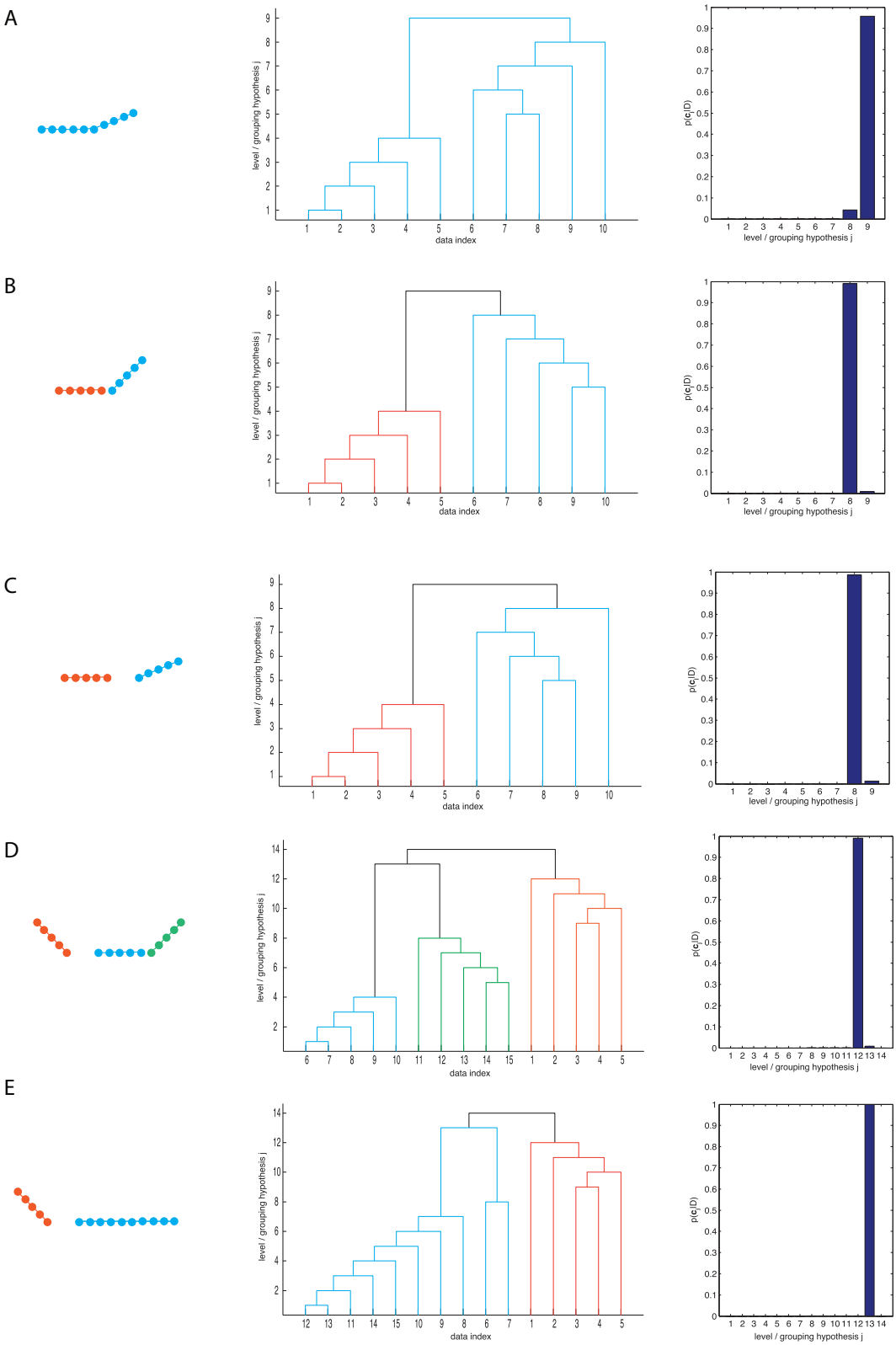

Figure 6. Bayesian hierarchical grouping (BHG) results for simple dot contours. The first column shows the input images and their MAP segmentation. Input dots are numbered from left to right. The second column shows the tree decomposition as computed by BHG. The third column shows the posterior probability distribution over all tree-consistent decompositions (i.e., grouping hypotheses). See the online article for the color version of this figure.

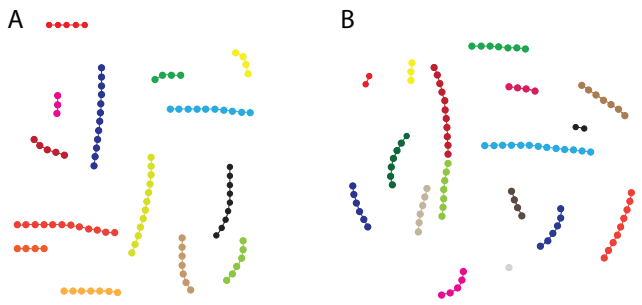

Figure 7. (A, B) The MAP grouping hypothesis for a more complex dot configurations. Distinct colors indicate distinct components in the MAP. (B) An example illustrating some shortcomings of the model. The preference for shorter segments leads to some apparently coherent segments to be oversegmented. See the online article for the color version of this figure.

salience in a more principled way by comparing two grouping hypotheses: the hypothesis in which the part was last present within the computed hierarchy $\boldsymbol{c_1}$, and the hypothesis $\boldsymbol{c_0}$ one step up in the hierarchy in which the part ceases to exist as a distinct entity. The higher this ratio, the stronger the part's role in the inferred explanation of the shape. In the examples that follow, we defined part salience as the log ratio between posterior probabilities of these hypotheses, which captures the weight of evidence in favor of the part. Figure 11 shows some sample results. Note that the log posterior ratio increases with both part length (Figure 11A) and protrusion (Figure 11B), even though neither is an overt factor in its computation. The implication is that the known influence of both factors is in fact a side effect of the unifying Bayesian account.

To demonstrate quantitatively how BHG captures part salience, we compared our model's performance with published human-subject data on part identification. Cohen and Singh (2007) tested the contribution of various geometric factors to part salience. Here, we focus on their experiment concerning part protrusion. In this experiment, subjects were shown a randomly generated shape on each trial, with one of 12 levels of part protrusion (3 [base widths] $\times$ 4 [part lengths]), after which they were shown a test part depicting a part taken from this shape (see Figure 12A). Subjects were asked to indicate in which of four display quadrants this part was present in the original, complete shape. Cohen and Singh found that subjects' accuracy in this task increased monotonically with increasing protrusion of the test part. For each of the 12 levels of part protrusion, subjects were shown 50 different randomly generated shapes. In order for us to compare our model's performance with subjects' accuracy, we ran BHG on 20 randomly selected shapes for each level of part protrusion. We then looked for the presence of the test part in the hierarchy generated by BHG and computed the log posterior ratio between the hypotheses $\boldsymbol{c_1}$ and $\boldsymbol{c_0}$ (Equation 10). Because subject responses were binomial, we computed the log odds ratio. Figure 12 shows that log odds of the subjects' accuracy on the task increases monotonically with the log posterior ratio of the test part. The log posterior ratio of the test part was found to be a good predictor of the log odds of the subject's accuracy at part identification (LRT = 50.594, $df = 1$, $R^2 = 0.4109$; BF = $8.0271e + 14$; see Footnote 7).

Shape Completion

Shape completion refers to the visual interpolation of contour segments that are separated by gaps, often caused in natural images by occlusion. In dealing with partially occluded object boundaries, the visual system needs to solve two problems (Singh & Fulvio, 2007; Takeichi, Nakazawa, Murakami, & Shimojo, 1995). First, it needs to determine if two contour elements should be grouped together into the representation of single, extended contour (the *grouping problem*). Second, it has to determine what the shape of the contour within the gap region should be (the *shape problem*). Here, we focus primarily on the shape problem and show how our model can make perceptually natural predictions about the missing shape of a partly occluded contour.

Local and global influences. Most previous models of completion base their predictions on purely local contour information, namely, the position and orientation of the contour at the point at which it disappears behind the occluder (called the *inducer*; Ben-Yosef & Ben-Shahar, 2012; Fantoni & Gerbino, 2003; Williams & Jacobs, 1997). Such models cannot explain the influence of nonlocal factors such as global symmetry (Sekuler, Palmer, & Flynn, 1994; van Lier et al., 1995) and axial structure (Fulvio & Singh, 2006). Nonlocal aspects of shape are, however, notoriously difficult to capture mathematically. But the BHG framework allows shape to be represented at any and all hierarchical levels, allowing it to make predictions that combine local and global factors in a comprehensive fashion.

In the BHG framework, completion is based on the posterior predictive distribution over missing contour elements, which assigns probability to potential completions conditioned on the estimated model. Assuming that figure-ground segmentation has already been established, we first compute the hierarchical representation of the occluded shape (given the object definitions set up in the context of part decomposition earlier) with the missing boundary segment, then we compute the posterior predictive (Equation 8) based on the best grouping hypothesis, that is, MAP tree slice. This induces a probability distribution over the field of potential positions of contour elements in the occluded area as influenced by the global shape. This field should thus be understood as a global shape prior over possible completions, which could be used by future models in combination with some good-continuation constraint based on the local inducers to infer a particular completion. For example, consider a simple case in which a circle is occluded by a square (Figure 13A). The model predicts that the interpolated contour must lie along a circular path with the same curvature as the rest of the circle. The model makes similarly intuitive predictions even with multipart shapes (Figure 13B). The model can also handle cases in which grouping is ambiguous (the grouping problem mentioned earlier), that is, in which the projection of the partly occluded object is fragmented in the image by the occluder (Figure 13C). The model can even make a prediction in cases in which there is not enough information for a local model to specify a boundary (Figure 13D). All these completions follow directly from the framework, with no fundamentally new mechanisms. The completion problem, in this framework, is just an aspect of the effective quantification of the "goodness" or Prägnanz of grouping hypotheses, operationalized via the Bayesian posterior.

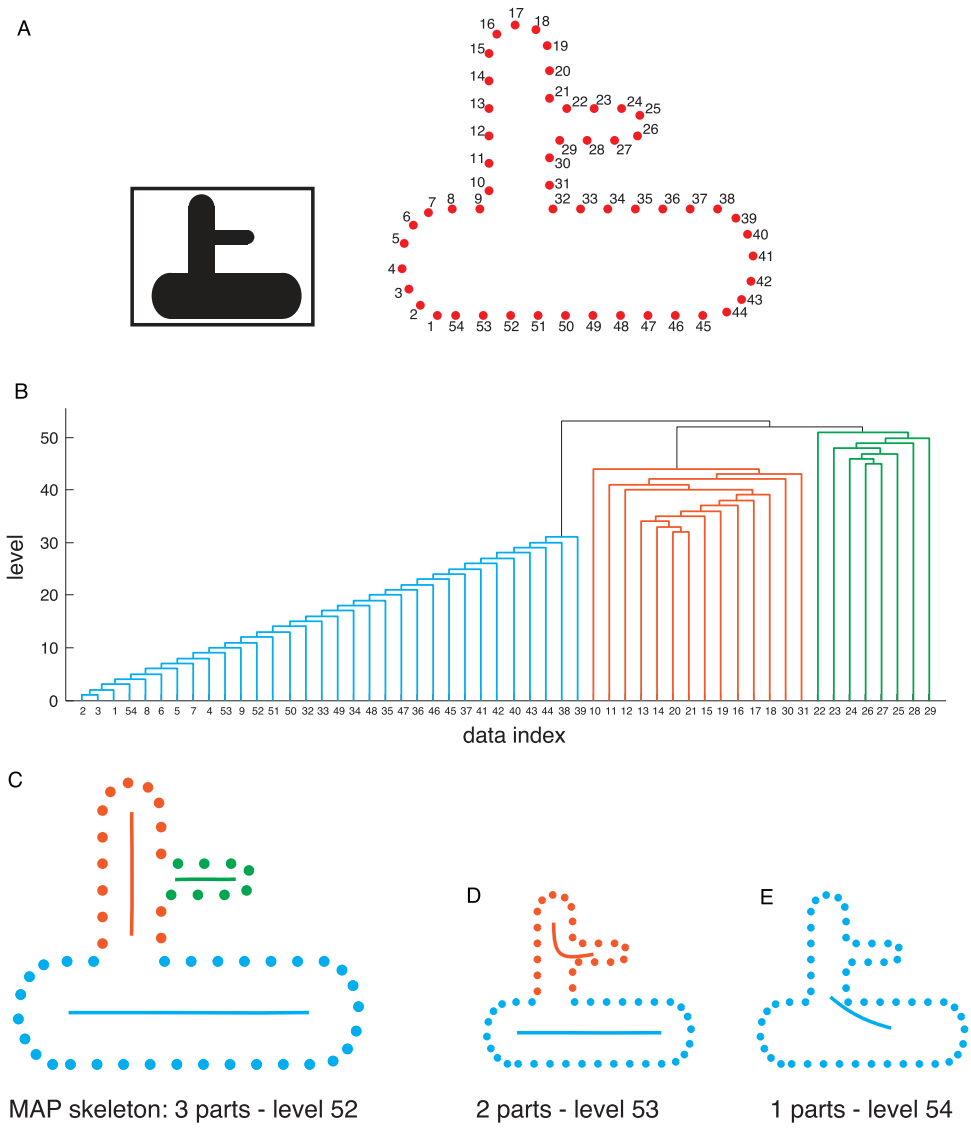

Figure 8. Shape decomposition in Bayesian hierarchical grouping. (A) A multipart shape, and (B) resulting tree structure depicted as a dendrogram. Colors indicate MAP decomposition, corresponding to the boundary labeling shown in (C). D and E show (lower probability) decompositions at other levels of the hierarchy. See the online article for the color version of this figure.

Dissociating global and local predictions. Some of the model predictions mentioned above could also have been made by a purely local model. However, global and local models often predict vastly different shape completions, and the strongest evidence for global influences rests on such cases. For example Sekuler et al. (1994) found that certain symmetries in the completed shape facilitated shape completion. More generally, van Lier et al. (1994, 1995) found that the regularity of the completed shape, as formulated in their regularity-based framework, likewise influences completion. The shapes in Figure 14, containing so-called *fuzzy* regularities, further illustrate the necessity for global accounts (van Lier, 1999). As we keep the inducer orientation and position constant, we can increase the complexity of a tubular shape's contour (Figure 14A and C). A model based merely on local inducers would expect the completed contour to look exactly the same in both cases, but to most observers, it does not. Because BHG takes the global shape into account, it predicts a more uncertain (i.e., noisy) completion when the global shape is noisier (Figure 14B, D, and E). The capacity of the BHG framework to precisely quantify such influences raises the possibility that, in future work, we might be able to fully disentangle local and global influences and determine how they combine to determine visual shape completion.

Discussion

In this article we proposed a principled and mathematically coherent framework for perceptual grouping, based on a central

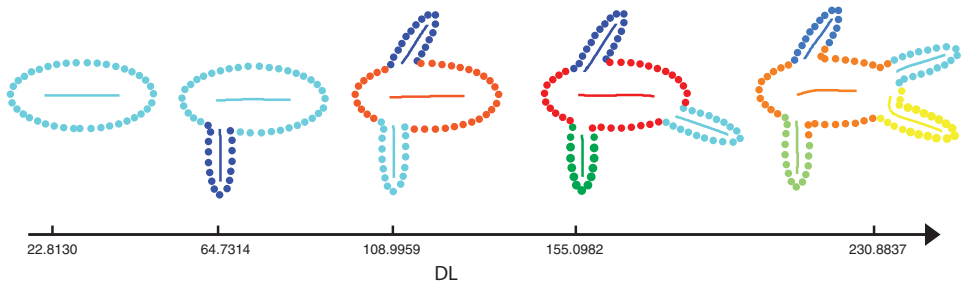

Figure 9. MAP skeleton as computed by the Bayesian hierarchical grouping for shapes of increasing complexity. The axis depicts the expected complexity, DL (Equation 11), of each of the shapes based on the entire tree decomposition computed. See the online article for the color version of this figure.

theoretical construct (mixture estimation) implemented in a unified algorithmic framework (BHC). The main idea, building on proposals from our previous articles (Feldman et al., 2014; Froyen et al., 2015), is to recast the problem of perceptual grouping as a mixture estimation problem. We model the image as a mixture (in the technical sense) of objects, and then introduce effective computational techniques for estimating the mixture—or, in other words, decomposing the image into objects. Above, we illustrated the framework for several key problems in perceptual grouping: contour integration, part decomposition, and shape completion. In this section we elaborate on several properties of the framework, and point out some of its advantages over competing approaches.

Object Classes

The flexibility of the BHG framework lies in the freedom to define object classes. In the literature, the term "object" encompasses a wide variety of image structures, simply referring to whatever units result from grouping operations (Feldman, 2003). In our framework, objects are reconceived *generatively* as stochastic image-generating processes, which produce image elements under a set of probabilistic assumptions (Feldman & Singh, 2006). Formally, object classes are defined by two components: the generative (likelihood) function, $p(D|\theta)$, which defines how image elements are generated for a given class, and the prior over parameters, $p(\theta|\beta)$, which modulates the values the parameters themselves are likely to take on. Taken together, these two components define what objects in the class tend to look like.

In all the applications demonstrated above, we assumed a generic object class appropriate for spatial grouping problems, which generalizes the skeletal generating function proposed in Feldman and Singh (2006). Objects drawn from this class contain elements generated at stochastically chosen distances from an underlying generating curve, whose shape is itself chosen stochastically from a curve class. (The curve may have length zero, in which case the generating curve is a point, and the resulting shape approximately circular.) Depending on the parameters, this object class can generate contour fragments, dot clouds, or shape boundaries. This broad class unifies a number of types of perceptual units, such as contours, dot clusters, and shapes, that are traditionally treated as distinct in the grouping literature, but which we regard as a connected family. For example, in our framework, contours are, in effect, elongated shapes with short ribs, dot clusters are shapes with image elements generated in their interiors, and so forth. Integrating these object classes under a common umbrella makes it possible to treat the corresponding grouping rules, contour integration, dot clustering, and the others mentioned above as special cases of a common mechanism.

One can, of course, extend the current object definition into 3D (El-Gaaly, Froyen, Elgammal, Feldman, & Singh, 2015), or imagine alternative object definitions, such as Gaussian objects (Froyen et al., 2015; Juni, Singh, & Maloney, 2010). Furthermore, objects classes can be defined in nonspatial domains, using such features as color, texture, and contrast (Blaser, Pylyshyn, & Holcombe, 2000). As long as the object classes can be expressed in the appropriate technical form, the Bayesian grouping machinery can be pressed into service.

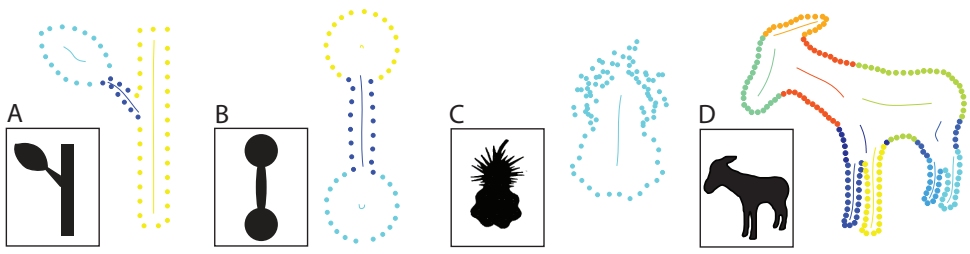

Figure 10. Examples of MAP tree slices for (A) leaf on a branch, (B) dumbbells, (C) "prickly pear" from Richards, Dawson, and Whittington (1986), and (D) Donkey. (Example D has a higher dot density because the original image was larger.) See the online article for the color version of this figure.

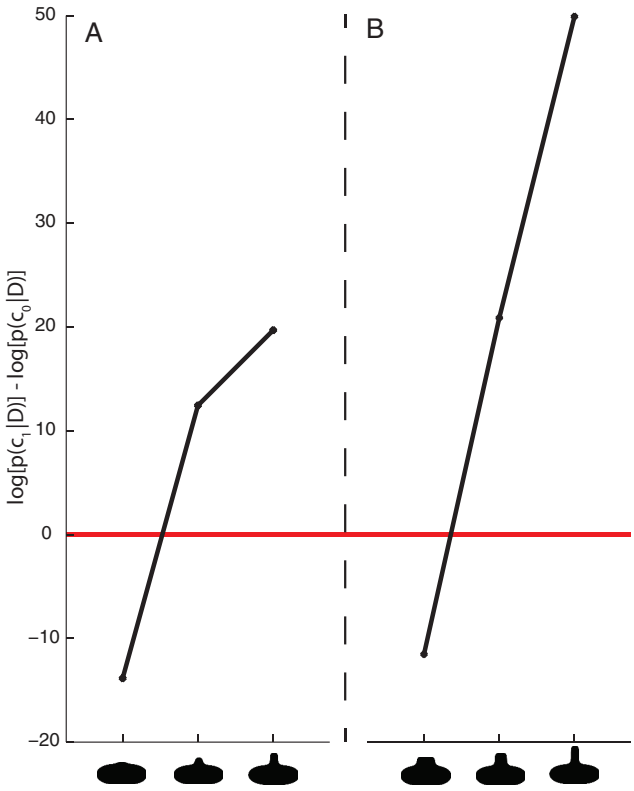

Figure 11. Log posterior ratio between tree consistent one- and two-component hypotheses, as a function of (A) part length, and (B) part protrusion. See the online article for the color version of this figure.

Hierarchical Structure

BHG constructs a hierarchical representation of the image elements based on the object definitions and assumption set. Hierarchical approaches to perceptual grouping have been very influential (e.g., Baylis & Driver, 1993; Lee, 2003; Palmer, 1977; Pomerantz et al., 1977). At the coarsest level of representation, our approach represents all the image elements as one object. For shapes, this is similar to the notion of a *model axis* by Marr and Nishihara (1978), providing only coarse information such as size and orientation, of the entire shape (Figure 15A). Lower down in the hierarchy, individual axes correspond to something more like classical shape primitives such as geons (Biederman, 1987) or generalized cones (Binford, 1971; Marr, 1982; Figure 15A to C). Note, however, that our axial representation not only describes the shape of each part but also entails a statistical characterization of the image data that supports the description. This is what allows descriptions of missing parts (see Figure 13). Even further down the hierarchy, more and more detailed aspects of image structure are represented (Figure 15B and C).

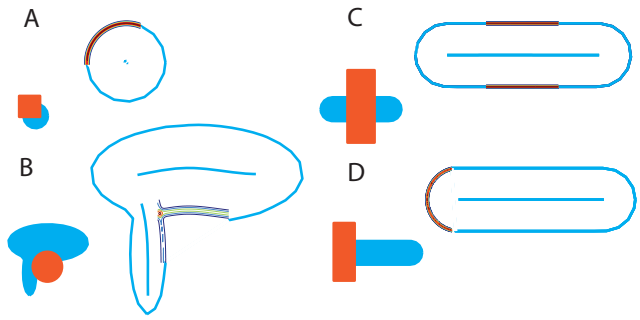

Figure 13. Completion predictions based on the posterior predictive distribution based on the MAP skeleton (as computed by Bayesian hierarchical grouping). See the online article for the color version of this figure.

Structural versus spatial scale. Two different notions of *scale* can be distinguished in the literature. Most commonly, scale is defined in terms of spatial frequency, invoking a hierarchy of

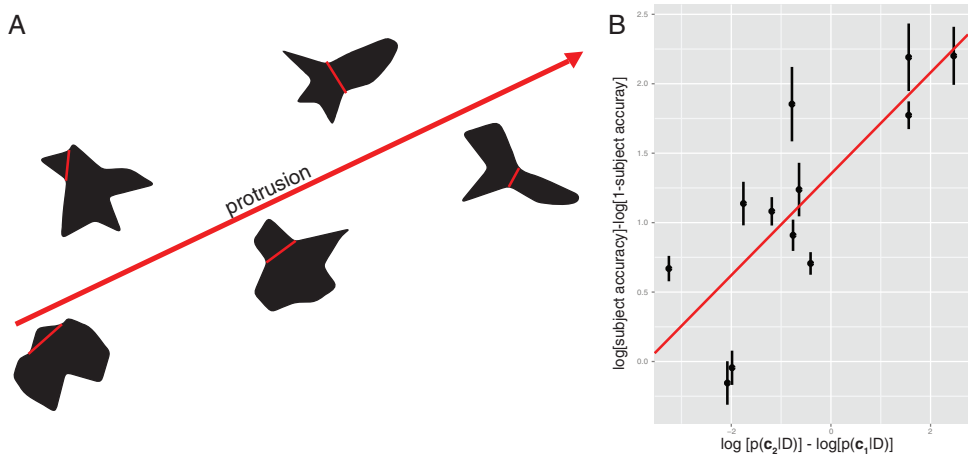

Figure 12. (A) Representative stimuli used in the Cohen and Singh (2007) experiment relating part protrusion to part saliency. As part protrusion increases, so does subjective part saliency. (Parts are indicated by a red part cut.) (B) Log odds of subject accuracy as a function of log posterior ratio log $p(\boldsymbol{c_1}|D)$ − log $p(\boldsymbol{c_0}|D)$ as computed by the model. (Error bars depict the 95% confidence interval across subjects. The red curve depicts the linear regression.) See the online article for the color version of this figure.

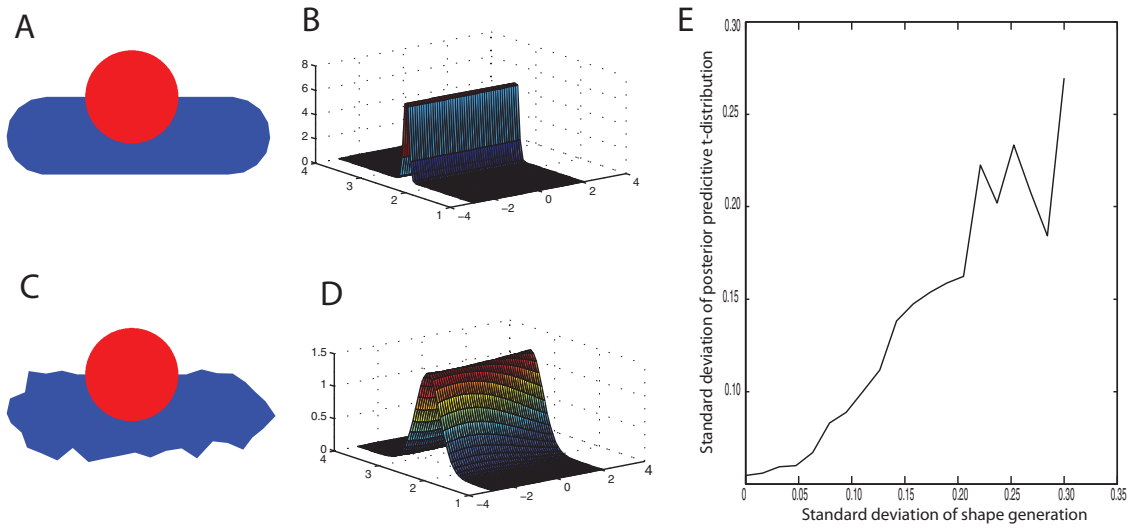

Figure 14. A simple tubular shape was generated with different magnitudes (*SD*) of contour noise. Note that the local inducers are identical in both input images (A and C). For noiseless contours (A), the posterior predictive distribution over completions has a narrow noise distribution (B), whereas for noisy contours (C), the distribution has more variance (D). Panel E shows the relationship between the noise on the contour and the completion uncertainty as reflected by the posterior predictive distribution. See the online article for the color version of this figure.

operators or receptive fields of different sizes, with broader receptive fields integrating image data from larger regions, and finer receptive fields integrating smaller regions. This notion of spatial scale has been incorporated into a number of models of perceptual grouping, including figure-ground (e.g., Jehee, Lamme, & Roelfsema, 2007; Klymenko & Weisstein, 1986) and shape representation (e.g., Burbeck & Pizer, 1994). Pure spatial scale effects have, however, been shown not to account for essential aspects of grouping (Field et al., 1993; Jáñez, 1984). In contrast, any hierarchical representation of image structure implicitly defines a notion of *structural scale*, referring to the position of a particular structural unit along the hierarchy (Feldman, 2003; Marr & Nishihara, 1978; Palmer, 1977; Siddiqi et al., 1999). Typically, structural scale assumes a tree structure, with nodes at various levels describing various aggregations of image units, the top node incorporating the entire image, and leaf nodes referring to individual image elements.

These two notions of scale have not generally been explicitly distinguished in the literature, but we would argue that they are distinct. Elements at a common level of structural scale need not, in fact, have similar spatial frequency content, whereas elements with similar spatial scales may well occupy distinct levels of the organizational hierarchy. The BHG framework incorporates both notions of scale in a manner that keeps their distinct contributions clear (Figure 16A). In BHG, spatial scale is modulated in effect by the prior over the variance of rib length, $p(\sigma|\sigma_0, \nu_0)$ (also see Appendix B). Large values of σ_0 makes object hypotheses more tolerant to variance in the spatial location of image elements.

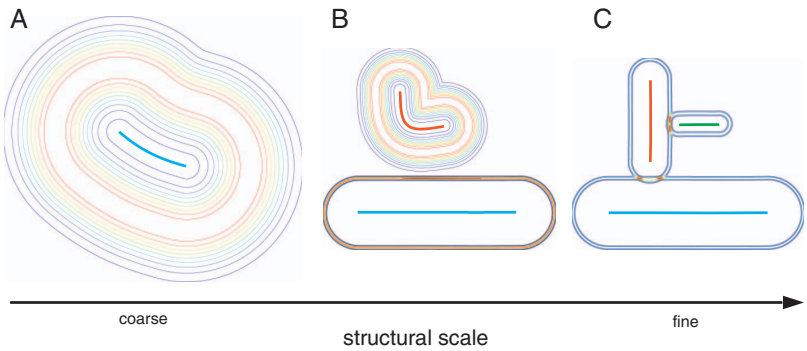

Figure 15. Prediction fields for the shape in Figure 8 for three different levels of the hierarchy. In order to illustrate how underlying objects also represent the statistical information about the image elements, they explain the prediction/completion field was computed for each object separately without normalization so that the highest point for each object is equalized. See the online article for the color version of this figure.

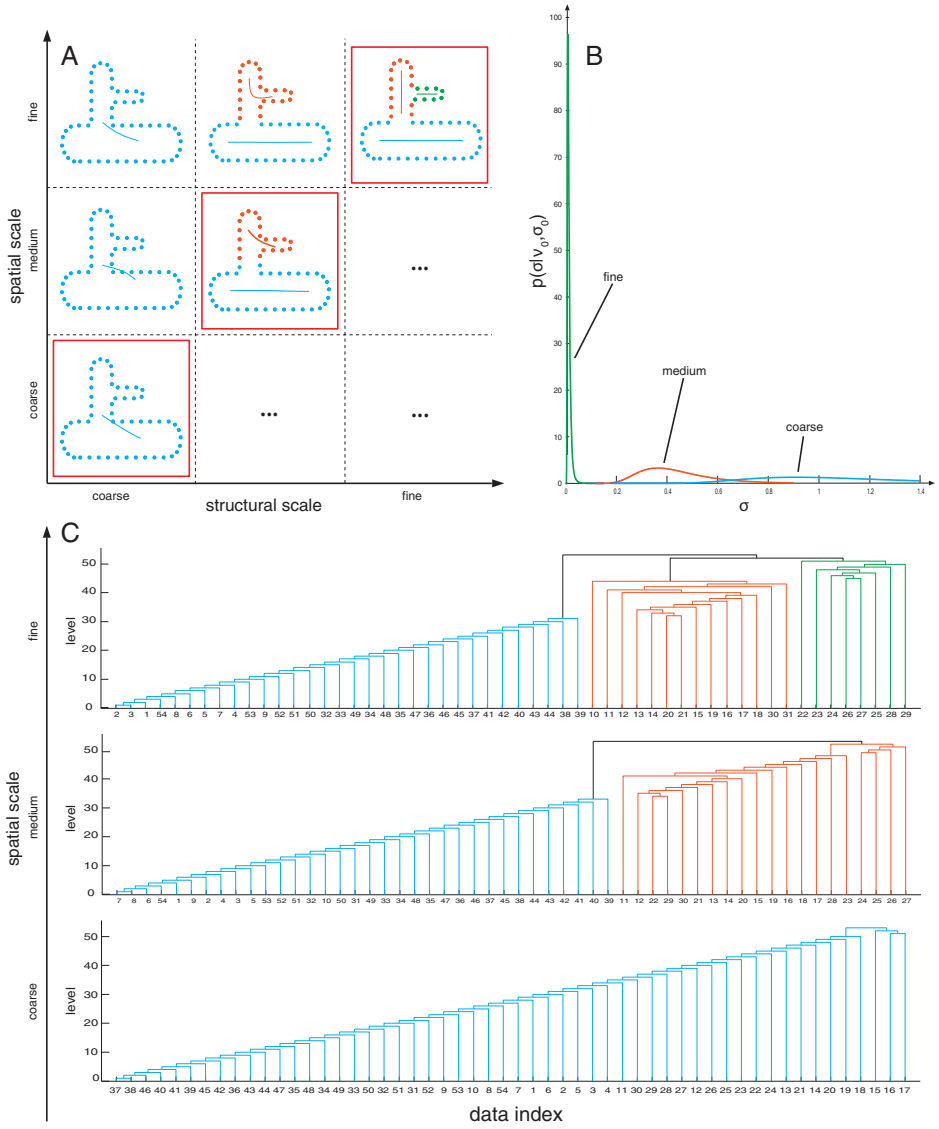

Figure 16. Relating structural and spatial scale in our model by means of the shape in Figure 8. (A) Relationship between structural and spatial scale depicting their orthogonality. The red squares depict the most probable structural grouping hypothesis for each spatial scale. (B) Showing the priors over the variance of the rib length, σ for each spatial scale. (C) Hierarchical structure as computed by our framework depicted as a dendrogram for each spatial scale. The most probable hypothesis is shown in color. See the online article for the color version of this figure.

Figure 16B illustrates the effect of changing this prior on the resulting interpretation. Structural scale, in contrast, is defined by the level in the estimated hierarchy. Figure 16C illustrates how the hierarchy changes as spatial scale is modified. At fine spatial scales, the inferred hierarchy includes three hypotheses at distinct structural scales, with the three-part hypothesis being the most probable (Figure 16A). But at a coarser spatial scale, a different hierarchical interpretation is drawn, with the two-part hypothesis the most probable, and the three-part hypothesis, as found in the finer spatial scale, completely absent. This example illustrates how structural scale and spatial scale are related but distinct, describing different aspects of the grouping mechanism.

Selective organization. *Selective organization* refers to the fact that the visual system overtly represents some subsets of image elements but not others (Palmer, 1977). The way BHG considers grouping hypotheses and builds hierarchical representations realizes the notion of selective organization in a coherent way. In our model, only N grouping hypotheses are considered, whereas the total number of possible grouping hypotheses $\boldsymbol{c}$ is exponential in N. Grouping hypotheses selected at one level of the hierarchy depend directly on those chosen at lower level (grouping hypotheses are "tree-consistent"), leading to a clean and consistent hierarchical structure. Inevitably, this means that numerous grouping hypotheses are not represented at all in the hierarchy. Such

selective organization is empirically supported by the observation that object parts not represented in the hierarchy are more difficult to retrieve (Cohen & Singh, 2007; Palmer, 1977).

Advantages of the Bayesian Framework

The Bayesian approach used here has several advantages over traditional approaches to perceptual grouping. First, it allows us to assign different degrees of belief, that is, probabilities, to different grouping hypotheses, capturing the often graded responses found in subject data. Previous nonprobabilistic models often only converge on one particular grouping hypothesis (e.g., Williams & Jacobs, 1997), or are unable to assign graded degrees of belief to different grouping hypotheses at all (e.g., Compton & Logan, 1993). Second, Bayesian inference makes optimal use of available information modulo the assumptions adopted by the observer (Jaynes, 2003). In the context of perceptual grouping, this means that a Bayesian framework provides the optimal way of grouping the image, given both the image data and the particular set of assumptions about object classes adopted by the observer. The posterior over grouping hypotheses, in this sense, represents the most "reasonable" way of grouping the image, or more accurately, the most reasonable way of assigning degrees of belief to distinct ways of grouping the image.

Conclusion

In this article we have presented a novel, principled, and mathematically coherent framework for understanding perceptual grouping. Bayesian Hierarchical Grouping (BHG) defines the image as a mixture of objects, and thus reformulates the problem of perceptual grouping as a mixture estimation problem. In the BHG framework, perceptual grouping means estimating how the image is most plausibly decomposed into distinct stochastic image sources—objects—and deciding which image elements belong to which sources, that is, which elements were generated by which object. The generality of the framework stems from the freedom it allows in how objects are defined. In the formulation discussed above, we used a single simple but flexible object definition, allowing us to apply BHG to such diverse problems as contour integration, dot clustering, part decomposition, and shape completion. But the framework can easily be extended to other problems and contexts simply by employing alternative object class definitions.

BHG has a number of advantages over conventional approaches to grouping. First, its generality allows a wide range of grouping problems to be handled in a unified way. Second, as illustrated above, with only a small number of parameters, it can explain a wide array of human grouping data. Third, in contrast to classical approaches, it allows grouping interpretations to be assigned degrees of belief, helping to explain a range of graded percepts and ambiguities, some of which were exhibited above. Fourth, it provides hierarchically structured interpretations, helping to explain human grouping percepts that arise at a variety of spatial and structural scales. Finally, in keeping with its Bayesian roots, the framework assigns degrees of belief to grouping interpretations in a rational and principled way.

Naturally, the BHG model framework has some limitations. One broad limitation is the general assumption that image data are independent and identically distributed (i.i.d.) conditioned on the mixture model, meaning essentially that image elements are randomly drawn from the objects present in the scene. Such an assumption is nearly universal in statistical inference, but certainly limits the generality of the approach, and may be especially inappropriate in dynamic situations. We note a few additional limitations more specific to our framework and algorithm.

First, in its current implementation, the framework cannot easily be applied to natural images. Standard front-ends for natural images (e.g., edge detectors or V1-like operator banks) yield extremely noisy outputs, and natural scenes contain object classes for which we have not yet developed suitable generative models. (Algorithms for perceptual grouping are routinely applied to real images in the computer vision literature, but such schemes are notoriously limited compared to human observers, and are generally unable to handle the subtle Gestalt grouping phenomena discussed here.) Extending our framework so that it can be applied to natural images is a primary goal of our future research.

Second, broad though our framework may be, it is limited to perceptual grouping and does not solve other fundamental interrelated problems such as 3D inference. Extending our framework to 3D is actually conceptually simple, albeit computationally complex, in that it simply requires the generative model be generalized to 3D, whereas estimation and hypothesis comparison mechanisms would remain essentially the same. Indeed in recent work (El-Gaaly et al., 2015), we have taken steps toward such an extension, allowing us to decompose 3D objects into parts.

Finally, it is unclear how exactly the framework could be carried out by neural hardware. However, Beck, Heller, and Pouget (2012) have recently derived neural plausible implementations for variational inference (a technique for providing Bayeisan estimates of mixtures; Attias, 2000) within the context of probabilistic population codes. Because such techniques allow mixture decomposition by neural networks, they suggest a promising route toward a plausible neural implementation of the theoretical model we have laid out here.

We end with a comment on the "big picture" contribution of this article. The perceptual grouping literature contains a wealth of narrow grouping principles covering specific grouping situations. Some of these, such as proximity (Kubovy et al., 1998; Kubovy & Wagemans, 1995) and good continuation (e.g., Feldman, 2001; Field et al., 1993; Parent & Zucker, 1989; Singh & Fulvio, 2005), have been given mathematically concrete and empirically persuasive accounts. But (with a few intriguing exceptions such as Zhu, 1999; Geisler & Super, 2000; Ommer & Buhmann, 2003; Song & Hall, 2008) similarly satisfying accounts of the broader overarching principles of grouping—if indeed any exist—remain comparatively vague. Terms such as *Prägnanz*, *coherence*, *simplicity*, *configural goodness*, and *meaningfulness* are repeatedly invoked without concrete definitions, often accompanied by ritual protestations about the difficulty in defining them. In this article we have taken a step toward solving this problem by introducing an approach to perceptual grouping that is both principled, general, and mathematically concrete. In BHG, the posterior distribution quantifies the degree to which each grouping interpretation "makes sense"—the degree to which it is both plausible a priori and fits the image data. The data reviewed above suggests that this formulation effectively quantifies the degree to which grouping interpretations

make sense to the human visual system. This represents an important advance toward translating the insights of the Gestalt psychologists—still routinely cited a century after they were first introduced (Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012)—into rigorous modern terms.

References

Amir, A., & Lindenbaum, M. (1998). A generic grouping algorithm and its quantitative analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 168–185.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Attias, H. (2000). A variational Bayesian framework for graphical models. In S. A. Solla, T. K. Leen, & K. Müller (Eds.), *Advances in neural information processing systems 12* (pp. 209–215). Cambridge, MA: MIT Press.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61,* 183–193.

August, J., Siddiqi, K., & Zucker, S. W. (1999). Contour fragment grouping and shared, simple occluders. *Computer Vision and Image Understanding, 76,* 146–162.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 451–470.

Beck, J., Heller, K., & Pouget, A. (2012). Complex inference in neural-circuits with probabilistic population codes and topic models. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 25* (pp. 3059–3067). Mahwah, NJ: Curran Associates, Inc.

Ben-Yosef, G., & Ben-Shahar, O. (2012). A tangent bundle theory for visual curve completion. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34,* 1263–1280.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Binford, T. O. (1971). Visual perception by a computer. In *IEEE Conf. on Systems and Controls, Miami*.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.

Blaser, E., Pylyshyn, Z., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408,* 196–199.

Blum, H. (1967). A transformation for extracting new descriptors of shape. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 153–171). Cambridge, MA: MIT Press.

Blum, H. (1973). Biological shape and visual science. *Journal of Theoretical Biology, 38,* 205–287.

Boselie, F., & Wouterlood, D. (1989). The minimum principle and visual pattern completion. *Psychological Research, 51,* 93–101.

Brady, M., & Asada, H. (1984). Smoothed local symmetries and their implementation. *The International Journal of Robotics Research, 3,* 36–61.

Burbeck, C., & Pizer, S. (1994). Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research, 35,* 1917–1930.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103,* 566–581.

Claessens, P., & Wagemans, J. (2008). A Bayesian framework for cue integration in multistable grouping: Proximity, collinearity, and orientation priors in zigzag lattices. *Journal of Vision, 8,* 1–23.

Cohen, E. H., & Singh, M. (2006). Perceived orientation of complex shape reflects graded part decomposition. *Journal of Vision, 6,* 805–821.

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research, 47,* 2825–2840.

Compton, B., & Logan, G. (1993). Evaluating a computational model of perceptual grouping by proximity. *Attention, Perception, & Psychophysics, 53,* 403–421.

de Berg, M., Cheong, O., van Kreveld, M., & Overmars, M. (2008). *Computational geometry: Algorithms and applications*. Berlin/Heidelberg, Germany: Springer-Verlag.

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition, 99,* 275–325.

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

El-Gaaly, T., Froyen, V., Elgammal, A., Feldman, J., & Singh, M. (2015). A Bayesian approach to perceptual 3D object-part decomposition using skeleton-based representations. In *AAAI Conference on Artificial Intelligence*. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9563

Ernst, U., Mandon, S., Schinkel-Bielefeld, N., Neitzel, S., Kreiter, A., & Pawelzik, K. (2012). Optimality of human contour integration. *PLoS Computational Biology, 8*(5), e1002520.

Fantoni, C., & Gerbino, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3*(4), 281–303.

Farouki, R. T., & Hinds, J. K. (1985). A hierarchy of geometric forms. *IEEE Computer Graphics and Applications, 5,* 51–78.

Feldman, J. (1997a). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research, 37,* 2835–2848.

Feldman, J. (1997b). Regularity-based perceptual grouping. *Computational Intelligence, 13,* 582–623.

Feldman, J. (2001). Bayesian contour integration. *Perception and Psychophysics, 63,* 1171–1182.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences, 7,* 252–256.

Feldman, J. (2009). Bayes and the simplicity principle in perception. *Psychological Review, 116,* 875–887.

Feldman, J. (2014). Bayesian models of perceptual organization. In J. Wagemans (Ed.), *Handbook of perceptual organization*. New York, NY: Oxford University Press.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 103,* 18014–18019.

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision: An interdisciplinary perspective* (pp. 55–70). London: Springer-Verlag.

Feldman, J., Singh, M., & Froyen, V. (2014). Bayesian perceptual grouping. In S. Gepshtein, L. T. Maloney, & M. Singh (Eds.), *Handbook of computational perceptual organization*. Oxford, UK: Oxford University Press.

Field, D., Hayes, A., & Hess, R. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research, 33,* 173–193.

Flöry, S. (2005). *Fitting B-spline curves to point clouds in the presence of obstacles* (Unpublished doctoral dissertation, master's thesis), TU Wien, Vienna, Austria.

Froyen, V., Feldman, J., & Singh, M. (2010). A Bayesian framework for figure-ground interpretation. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems, 23* (pp. 631–639). Mahwah, NJ: Curran Associates, Inc.

Froyen, V., Feldman, J., & Singh, M. (2015). *Counting clusters: Bayesian estimation of the number of perceptual groups*. Manuscript submitted for publication.

Fulvio, J. M., & Singh, M. (2006). Surface geometry influences the shape of illusory contours. *Acta Psychologica, 123,* 20–40.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711–724.

Geisler, W. S., & Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review, 107,* 677–708.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology, 56,* 1–12.

Gershman, S. J., Jäkel, F. J., & Tenenbaum, J. B. (2013). Bayesian vector analysis and the perception of hierarchical motion. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*.

Heller, K., & Ghahramani, Z. (2005). Bayesian hierarchical clustering. In *Proceedings of the 22nd international conference on Machine learning* (pp. 297–304). ACM.

Helson, H. (1933). The fundamental propositions of gestalt psychology. *Psychological Review, 40,* 13–32.

Hochberg, J., & McAlister, E. (1953). A quantitative approach, to figural "goodness." *Journal of Experimental Psychology, 46,* 361.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18,* 65–96.

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition, 63,* 29–78.

Jáñez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research, 24,* 271–274.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. New York, NY: Cambridge University Press.

Jehee, J. F. M., Lamme, V. A. F., & Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Research, 47,* 1153–1165.

Juni, M. Z., Singh, M., & Maloney, L. T. (2010). Robust visual estimation as source separation. *Journal of Vision, 10,* 1–20.

Kanizsa, G. (1979). *Organization in vision*. New York, NY: Praeger.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55,* 271–304.

Kikuchi, M., & Fukushima, K. (2003). Assignment of figural side to contours based on symmetry, parallelism, and convexity. In V. Palade, R. Howlett, & L. Jain (Eds.), *Knowledge-based intelligent information and engineering systems* (Vol. 2774, pp. 123–130). Berlin/Heidelberg, Germany: Springer.

Klymenko, V., & Weisstein, N. (1986). Spatial frequency differences can determine figure-ground organization. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 324–330.

Köhler, W. (1950). Psychology and evolution. *Acta Psychologica, 7,* 288–297.

Kubovy, M., Holcombe, A., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35,* 71–98.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science, 6,* 225–234.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America, 20,* 1434–1448.

Leeuwenberg, E. L. J. (1969). Quantitative specification of information in sequential patterns. *Psychological Review, 76,* 216–220.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. New York, NY: Cambridge University Press.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience, 26,* 147–155.

Mamassian, P., & Landy, M. S. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Marr, D. (1982). *A computational investigation into the human representation and processing of visual information*. San Francisco, CA: Freeman.

Marr, D., & Nishihara, K. H. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *200,* 269–294.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering (Statistics: Textbooks and Monographs)*. New York, NY: Dekker.

McLachlan, G. J., & Peel, D. (2004). *Finite mixture models*. New York, NY: Wiley.

Mumford, D. (1994). Elastica and computer vision. In C. L. Bajaj (Ed.), *Algebraic geometry and its applications* (pp. 491–506). New York, NY: Springer.

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. Cambridge, MA: MIT Press.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics, 9,* 249–265.

Ommer, B., & Buhmann, J. M. (2003). A compositionality architecture forperceptual feature grouping. In A. Rangarajan, M. Figueiredo, & J. Zerubia (Eds.), *Energy minimization methods in computer vision and pattern Recognition* (pp. 275–290). Berlin/Heidelberg, Germany: Springer.

Ommer, B., & Buhmann, J. M. (2005). Object categorization by compositional graphical models. In A. Rangarajan, B. Vemuri, & A. L. Yuille (Eds.) *Energy minimization methods in computer vision and pattern recognition* (pp. 235–250). Berlin/Heidelberg, Germany: Springer.

Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review, 120,* 297–328.

Oyama, T. (1961). Perceptual grouping as a function of proximity. *Perceptual and Motor Skills, 13,* 305–306.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology, 9,* 441–474.

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11,* 823–839.

Pomerantz, J. R. (1986). Visual form perception: An overview. In E. C. Schwab & H. C. Nussbaum (Eds.), *Pattern recognition by humans and machines. Volume 2: Visual perception*. New York, NY: Academic Press.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422.

Rasmussen, C. E. (2000). The infinite Gaussian mixture model. In S. A. Solla, T. K. Leen, & K. R. Müller (Eds.), *Advances in neural information processing systems, 12* (pp. 554–560). Cambridge, MA: MIT Press.

Richards, W. A., Dawson, B., & Whittington, D. (1986). Encoding contour shape by curvature extrema. *Journal of the Optical Society of America, 3,* 1483–1491.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry theory*. River Edge, NJ: World Scientific Publishing Co.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology, 46,* 178–210.

Sajda, P., & Finkel, L. H. (1995). Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience, 7,* 267–291.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review, 117,* 1144–1167.

Sekuler, A. B., Palmer, S. E., & Flynn, C. (1994). Local and global processes in visual completion. *Psychological Science, 5,* 260–267.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 888–905.

Siddiqi, K., & Kimia, B. B. (1995). Parts of visual form: Computational aspects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17,* 239–251.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal of Computer Vision, 35,* 13–32.

Singh, M., & Fulvio, J. M. (2005). Visual extrapolation of contour geometry. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 102,* 939–944.

Singh, M., & Fulvio, J. M. (2007). Bayesian contour extrapolation: Geometric determinants of good continuation. *Vision Research, 47,* 783–798.

Singh, M., & Hoffman, D. D. (2001). Part-based representation of visual shape and implications for visual cognition. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Grouping and segmentation in vision* (Vol. 130, pp. 401–459). New York, NY: Elsevier Science.

Singh, M., Seyranian, D. G., & Hoffman, D. D. (1999). Parsing silhouettes: The short-cut rule. *Perception & Psychophysics, 61,* 636–660.

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception, 16,* 121–131.

Smits, J. T., Vos, P. G., & Van Oeffelen, M. P. (1985). The perception of a dotted line in noise: A model of good continuation and some experimental results. *Spatial Vision, 1,* 163–177.

Song, Y.-Z., & Hall, P. M. (2008). Stable image descriptions using Gestalt principles. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. Chun, T.-M. Rhyne, & L. Monroe (Eds.), *Advances in visual computing* (pp. 318–327). Berlin/Heidelberg, Germany: Springer.

Takeichi, H., Nakazawa, H., Murakami, I., & Shimojo, S. (1995). The theory of the curvature-constraint line for amodal completion. *Perception, 24,* 373–389.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1991). Accessibility: A criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology, 35,* 151–213.

van der Helm, P. A., & Leeuwenberg, E. L. J. (1996). Goodness of visual regularities: A nontransformational approach. *Psychological Review, 103,* 429–456.

van Leeuwen, C. (1990a). Indeterminacy of the isomorphism heuristic. *Psychological Research, 52,* 1–4.

van Leeuwen, C. (1990b). Perceptual learning systems as conservative structures: Is economy an attractor. *Psychological Research, 52,* 145–152.

van Lier, R. J. (1999). Investigating global effects in visual occlusion: From a partly occluded square to the back of a tree-trunk. *Acta Psychologica, 102,* 203–220.

van Lier, R. J., van der Helm, P. A., & Leeuwenberg, E. L. J. (1994). Integrating global and local aspects of visual occlusion. *Perception, 23,* 883–903.

van Lier, R. J., van der Helm, P. A., & Leeuwenberg, E. L. J. (1995). Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 571.

Wagemans, J. (1999). Toward a better approach to goodness: Comments on van der Helm & Leeuwenberg (1996). *Psychological Review, 106,* 610–621.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychological Bulletin, 138,* 1172–1217.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & Van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138,* 1218–1252.

Watt, R., Ledgeway, T., & Dakin, S. C. (2008). Families of models for gabor paths demonstrate the importance of spatial adjacency. *Journal of Vision, 8*(7), 1–19.

Weiss, Y. (1997). Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 520–527). New York, NY: IEEE.

Wertheimer, M. (1923). Untersuchungen zur lehre von der Gestalt [Investigations in Gestalt theory]. *Psychologische Forschung, 4,* 301–350.

Wilder, J., Singh, M., & Feldman, J. (2015). Contour complexity and contour detection. *Journal of Vision, 15,* 6.

Williams, L. R., & Jacobs, D. W. (1997). Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation, 9,* 837–858.

Zhu, S. C. (1999). Embedding gestalt laws in Markov random fields – a theory for shape modeling and perceptual organization. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21,* 1170–1187.

Zucker, S. W. (1985). Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision, Graphics, and Image Processing, 32,* 74–103.

Zucker, S. W., Stevens, K. A., & Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics, 34,* 513–522.

(*Appendices follow*)

Appendix A

Delaunay-Consistent Pairs

Bayesian hierarchical clustering (BHC) is a pairwise clustering method for which, at each iteration, merges between all possible pairs of trees T_i and T_j are considered. Given a data set $D = \{x_1 \ldots x_N\}$, the algorithm is initiated with N trees T_i each containing one data point $D_i = \{x_n\}$. As N increases, the number of pairs to be checked during this first iteration increases quadratically with N or, more specifically, as follows from combinatorics $\#pairs = (N^2 - N)/2$ (Figure A1), resulting in a complexity of $\mathbb{O}(N^2)$. In each of the following iterations, the hypothesis for merging only needs to be computed for pairs between existing trees and the newly merged tree from iteration $t - 1$. However, computing the hypothesis for merging $p(D_k|\mathcal{H}_0)$ for each possible pair is computationally expensive. Therefore, in our implementation of the BHC, we propose limiting the pairs checked to a local neighborhood as defined by the Delaunay triangulation. In other words, a data point x_n is only considered to be merged with data point x_m if it is a neighbor of that point. To initialize the BHC algorithm, we compute the Delaunay triangulation over the data set D. Given this, we can then compute a binary neighborhood vector $\boldsymbol{b}_n$ of length N for each data point x_n indicating which other data points x_n shares a Delaunay edge with. Together these vectors form a sparse symmetric neighborhood matrix. In contrast, when all pairs were considered this matrix would consist of all ones except for zeros along the diagonal. Using this neighborhood matrix, we can then define which pairs are to be checked at the first iteration. The amount of pairs checked at this initial stage is considerably lower than when all pairs are to be considered. Specifically, when simulating the amount of Delaunay-consistent pairs checked at this first iteration on a randomly scattered data set, the amount of pairs increased linearly with N (Figure A1). This results, when combined with the complexity of Delaunay triangulation $\mathbb{O}(N\log(N))$, in a computational complexity of $\mathbb{O}(N\log(N))$. In all of the following iterations, the neighborhood matrix is updated to reflect how merging trees also causes neighborhoods to merge. In order to implement this, we created a second matrix, D, called the token-to-cluster matrix of size $N \times ([N - 1] + N)$. The rows indicate the data points, and the columns, the possible clusters they can belong to. Given this matrix and the neighborhood matrix, we can then define which pairs to test in each of the iterations following the initial one. Note, when all Delaunay-consistent pairs have been exhausted, our implementation will revert to test all pair-wise comparisons.

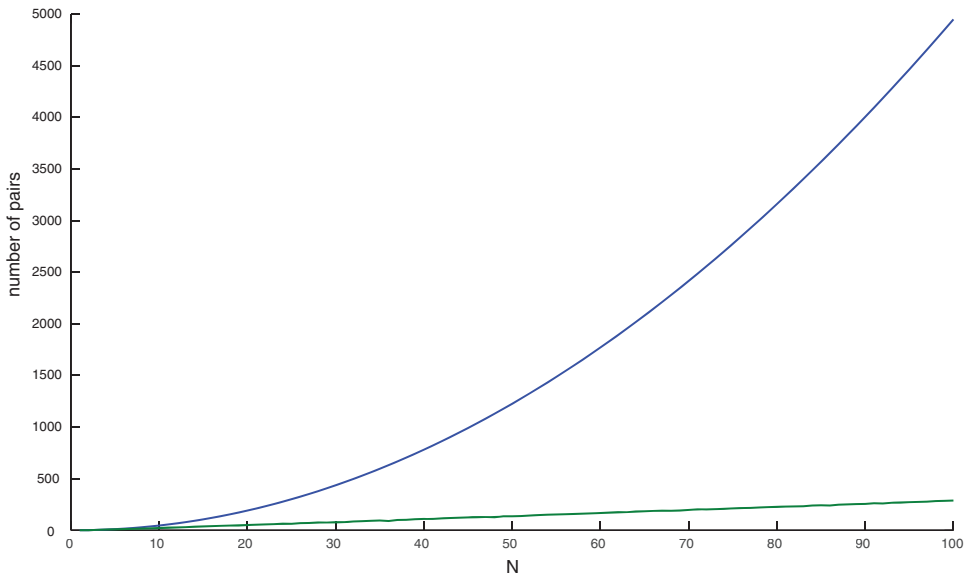

Figure A1. Difference between checking all pairs and only Delaunay-consistent pairs at the first initial iteration of the Bayesian hierarchical grouping. As the number of data points, N, increases, the number of pairs increases differently for the Delaunay-consistent (green) or all pairs (blue). See the online article for the color version of this figure.

(*Appendices continue*)

Appendix B

B-Spline Curve Estimation

Within our approach it is necessary to compute the marginal $p(D|\beta) = \int_\theta p(D|\theta)p(\theta|\beta)$. For simple objects such as Gaussian clusters, this can be solved analytically. However, for the more complex objects discussed here, integrating over the entire parameter space becomes rather intractable. The parameter vector for our objects looks as follows: $\theta = \{\boldsymbol{q}, \mu, \sigma\}$. Integrating over the Gaussian part of the parameter space (μ and σ) is straightforward and can be computed analytically. On the other hand, integrating over all possible B-spline curves as defined by the parameter vector $\boldsymbol{q}$ is intractable for our purposes. We therefore choose to pick the parameter vector $\boldsymbol{q}$ that maximizes Equation 15, while integrating over the Gaussian components. In what follows, we will describe how we estimate the B-spline curve for a given data set D.

B-spline curves were chosen for their versatility in taking many possible shapes by only defining a few parameters. Formally, a parametric B-spline curve is defined as

$$g(t) = \sum_{m=1}^{M} B_m(t) q_m, \tag{12}$$

where B_ms are the basis-functions and q_ms are the weights assigned to these (also called the *control points*). The order of the B-spline curve is defined by the order of the basis functions; we used cubic splines. In the simulations above, the number of basis functions and control points was set to $M = 6$. This number is a good compromise between the number of parameters that govern the B-spline and the flexibility to take a wide range of shapes. From this curve, we state that data points are generated perpendicular according a Gaussian likelihood function over the distance between a point on the curve $g(t_n)$ and the projected data point x_n (see Equation 9), also referred to as the riblength.

Given a data set $D = \{x_1 \ldots x_n\}$, we would like to compute the marginal $p(D|\beta)$. In order to do so, we first need to define the prior $p(\theta|\beta)$ and likelihood function $p(D|\theta)$ inside the integral:

$$p(D|\theta) = \prod_{n=1}^{N} \mathcal{N}(d_n|\mu, \sigma), \tag{13}$$

$$p(\theta|\beta) = \exp(F_1|\lambda_1)\exp(F_2|\lambda_2)\mathcal{N}\chi^{-2}(\mu, \sigma|\mu_0, \kappa_0, \sigma_0, \nu_0), \tag{14}$$

where $d_n = \|g(t_n) - x_n\|$. The likelihood function is the same as the generative function defined in Equation 9. The last factor in the prior is the conjugate prior to the Gaussian distribution in the likelihood function, the Normal-inv(χ^2), allowing for analytical computation of the marginal over parameters μ and σ. The first two factors define the penalties on the first and second derivative of the curve, respectively. Unfortunately, these are not conjugate priors to the distribution over different curves. Hence, integrating over all possible curves would have to be done numerically and is computationally intractable. Therefore, when computing the marginal, we choose to only integrate over the Gaussian components of the parameter vector θ and select $\boldsymbol{q}$ as to maximize

$$p(D|\beta, \boldsymbol{q}) = \int_{\mu,\sigma} \prod_{n=1}^{N} p(x_n|\mu, \sigma, \boldsymbol{q}) p(\theta|\beta) d\mu d\sigma. \tag{15}$$

In order to maximize this function, we followed a simple expectation-maximization (E–M) like algorithm traditional to parametric B-spline estimation (for a review, see Flöry, 2005). This algorithm has two stages. In the first stage (similar to expectation stage in E–M), each data point x_n is assigned a parameter value t_n such that $g(t_n)$ is the closest point on the B-spline curve to x_n, that is, x_n's perpendicular projection to the curve g. Finding these parameter values t_n is also called footpoint computation (the algorithm for this stage is described in Flöry, 2005). In the second, maximization, stage we maximize the function in Equation 15 given these $[x_n, t_n]$ pairs using unconstrained nonlinear optimization (as implemented through the function *fminsearch* in MATLAB). Computing the value if this function, given a specific value of $\boldsymbol{q}$, first of all involved computing the values for F_1 and F_2 in order for us to compute the prior on the curve shape. Both values are formally defined as

$$F_i = \int_t \left\| D^i g(t) \right\|^2 dt, \tag{16}$$

where i stands for the ith derivative. This integral was computed numerically by computing the ith derivative of $g(t)$ at 1,000 equally sampled points along the curve. With these values in hand, the marginal in Equation 15 can now be computed analytically by integrating over the Gaussian components:

$$p(D|\beta, \boldsymbol{q}) = \frac{\Gamma(\nu_n/2)}{\Gamma(\nu_0/2)} \sqrt{\frac{\kappa_0}{\kappa_n}} \frac{(\nu_0\sigma_0)^{\nu_0/2}}{(\nu_n\sigma_n)^{\nu_n/2}} \frac{1}{\pi^{n/2}} \exp(F_1|\lambda_1)\exp(F_2|\lambda_2), \tag{17}$$

with,

$$\begin{aligned} \mu_n &= \frac{\kappa_0\mu_0 + N\bar{d}}{\kappa_n}, \\ \sigma_n &= \kappa_0 + N, \\ \nu_n &= \nu_0 + N, \\ \sigma_n &= \frac{1}{\nu_n}\left[\nu_0\sigma_0 + \sum_n (d_n - \bar{d})^2 + \frac{N\kappa_0}{\kappa_0 + N}(\mu_0 - \bar{x})^2\right], \end{aligned} \tag{18}$$

where $\bar{d} = \frac{1}{n}\sum_n d_n$. The two stages just described are then repeated until convergence of the function in Equation 15.

Received July 8, 2014
Revision received March 16, 2015
Accepted May 18, 2015 ■