# BAYESIAN MIXTURE ESTIMATION FOR PERCEPTUAL GROUPING

**by**

**VICKY FROYEN**

**A dissertation submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Doctor of Philosophy**

**Graduate Program in Psychology**

**Written under the direction of**

**Dr. Jacob Feldman**

**and**

**Dr. Manish Singh**

**and approved by**

_____

_____

_____

_____

**New Brunswick, New Jersey**

**January, 2014**

**ABSTRACT OF THE DISSERTATION**

## Bayesian mixture estimation for perceptual grouping

**By VICKY FROYEN**

**Dissertation Director:**
**Dr. Jacob Feldman**
**and**
**Dr. Manish Singh**

Perceptual grouping is the process by which a set of image elements is divided into distinct "objects" or components. In this dissertation I propose a Bayesian framework for understanding perceptual grouping, in which the goal of the computation is to estimate the organization that best explains the observed configuration of image elements. I formalize the problem of perceptual grouping as a mixture estimation problem, where it is assumed that the configuration of elements is generated by a set of distinct components (or "objects"), whose underlying parameters one seeks to estimate. In the first part of this dissertation I will propose a simplified version of the framework and show how it can be used to estimate the number of objects, more specifically clusters of dots, present in the image. Across two experiments I show how the model gives an accurate and quantitatively precise account of subjects' numerosity judgments, while at the same time outperforming more standard accounts for dot clustering. In the second part of the dissertation this simplified framework is expanded to estimate a hierarchical representation of the image elements. This framework can easily be adjusted to different subproblems of perceptual grouping. Here I will show how an instantiation of our framework for contour integration, part decomposition, and shape completion can account for several key perceptual phenomena and previously collected human subject data.

## Acknowledgements

This research was supported by NIH EY021494 to J.F. and M.S., and NSF DGE 0549115 (Rutgers IGERT in Perceptual Science). I'm grateful to Rutgers Graduate students John Wilder and Brian McMahan, and K.U.Leuven post-doc Naoki Kogo for their many helpful comments and discussions. I thank Lorilei Alley for her help at various stages of the studies presented in Chapter 2. I would also like to thank all members from the Visual Cognition Lab that came and went during my presence here. Lastly I would like to thank Gene Roddenberry for creating Star Trek, whose infinite amount of episodes have allowed me to keep going on this project without burning out before the final line was put on paper. All of this dissertation research was conducted under the helpful mentorship of both Jacob Feldman and Manish Singh, to whom I give special thanks.

## Dedication

This dissertation is dedicated to my supportive parents without whom I would never have ended on this side of the Atlantic ocean. I would also like to express my gratitude to my girlfriend, Jewel Lim, my friends back home in Belgium, and new friends found here in the U.S for their support and help along the way.

# Table of Contents

# List of Tables

# List of Figures

# 1. Introduction

Perceptual grouping is the process by which otherwise chaotic visual "stuff" is organized into distinct and coherent objects. Although in the past 100 years of Gestalt research many theories and models have been proposed for several subproblems of perceptual grouping, no mathematically rigorous and coherent overarching framework has achieved wide acceptance. In this dissertation I put forward such a framework for perceptual grouping drawing its inspiration from different fields such as cognitive science, machine learning and pattern recognition. The framework proposed seeks to estimate the organization that best explains the observed configuration of image elements. I formalize the problem of perceptual grouping as a mixture estimation problem, where it is assumed that the configuration of image elements is generated by a set of distinct "objects".

The experiments and theory presented in this dissertation are divided into two chapters. Even though each chapter can stand on its own both are deeply connected by the framework underlying them.

In Chapter 2, I introduce our first steps towards understanding the problem of perceptual grouping as a Bayesian mixture estimation problem. Here I present our framework in a simplified form for a relatively simple grouping problem: grouping dots into clusters. I present two experiments in which I tested subjects' clustering behavior by asking them how many clusters were present. I found that the model gives an accurate and quantitatively precise account of subjects' numerosity judgments. Furthermore I found the model to outperform standard models for dot clustering in the field.

In Chapter 3, I introduce a mathematically rigorous framework for perceptual grouping expanding on the simplified framework proposed in Chapter 2. The framework proposed in this chapter creates a hierarchical representation of the image, decomposing it into distinct objects at each level and assigning beliefs to each of these grouping hypotheses. The generality of this framework lies in the flexibility of the object definition. Here I tested an instantiation of the framework for a general

object class incorporating problems such as part-decomposition, contour integration and shape completion. I show how the framework can account for several instant-psychophysical findings, and previously collected human subject data. Furthermore I show how the model has several indirect side-effects allowing us to make predictions about previously untested stimuli, and shed light on the dichotomy between spatial and structural scale.

The studies here are representative of a long-standing tradition of interdisciplinary work between once strongly interconnected fields: cognitive science and computer science. Inspired by recent mathematical advances in computer science I hope to be able to shed light on the processes underlying perceptual grouping based on one central idea: *an image is a mixture of objects*.

# 2. Counting clusters: Bayesian estimation of the number of perceptual groups

## 2.1 Abstract

Dividing a set of visual elements into groups or clusters is a basic problem of perceptual organization, but the computational mechanisms underlying it are still poorly understood. In this chapter we study how subjects group dots into clusters, and in particular how they decide how many clusters are contained in a given display. In two experiments, we showed subjects configurations of dots that were sampled from either two Gaussian clusters (Experiment 1) or three Gaussian clusters (Experiment 2). In both experiments we manipulated the distances between the clusters, relative to the spread of each cluster, in order to modulate the apparent number of clusters. We model the results in a Bayesian framework in which the observer attempts to estimate the *mixture model*, that is, the locations and parameters of the distinct sources from which the dots were generated. redThe model gives an accurate and quantitatively precise account of subjects' judgments. Thus our Bayesian approach to perceptual grouping, as one side-effect, effectively models the perception of cluster numerosity.

## 2.2 Introduction

Perceptual grouping is the process by which image elements are grouped into distinct units, clusters, or objects. The problem of grouping is an inherently difficult one, in that the visual system needs to select the "best" among an exponentially large number of possible grouping interpretations. Yet the visual system generally converges rapidly on an intuitive division. But despite an enormous literature (see Wagemans et al., 2012a; Wagemans et al., 2012b, for a modern review) the computational mechanisms underlying grouping are still poorly understood.

Perhaps the simplest case of perceptual grouping is the division of a set of isotropic visual elements (e.g. dots) into distinct clusters. Subjects can readily judge, for example, whether a set of dots appears to fall into two clusters or one (Fig. 2.1B).

Perceptual segmentation is best viewed as a graded or probabilistic process, however, rather than a binary one. In other words, the visual system represents degrees of belief concerning various segmentation hypotheses (such as" a single cluster," or "two separate clusters") based on evidence from multiple sources. Such graded representation has been shown to manifest itself in "global" judgments involving dot clusters, such as overall orientation (Cohen, Singh, & Maloney, 2008) and overall location (Juni, Singh, & Maloney, 2010). Specifically, the greater the evidence that a small sub-cluster within the overall cluster is a distinct "object" (i.e., arises from a different source), the less the perceptual estimate of overall orientation of the dot cluster is influenced by its presence (Cohen et al., 2008). Hence the graded or probabilistic nature of perceptual segmentation has important implications for estimating overall perceptual properties of dot clusters (and perceptual objects more generally).

Deciding whether a set of dots contains one or two clusters clearly involves the Gestalt principle of proximity (Wertheimer, 1923), in that nearby dots are more likely to be clustered together (Kubovy & Wagemans, 1995; Kubovy, Holcombe, & Wagemans, 1998). But the process by which dots are assigned to clusters, and the overall number of clusters is determined, is less clear. A number of quantitative models for dot clustering and cluster enumeration have been proposed (Van Oeffelen & Vos, 1982; Compton & Logan, 1993; Allik & Tuulmets, 1991). These models fit human data reasonably well, albeit after fitting a number of *ad hoc* parameters. Moreover these models are very specific to the dot grouping problem, and do not generalize to other perceptual grouping problems (see discussion below). We sought an approach to this problem that is both more principled and more generalizable to other problems of perceptual grouping.

In the study below, we asked subjects to judge the number of clusters in configurations of dots, and show that their responses can be effectively modeled by a Bayesian estimation procedure given a few simple assumptions about the statistical properties of the clusters. The model predicts not only the most likely response for each configuration, but also the degree of belief in each potential number estimate, and hence the relative probability of the various responses. Moreover, unlike existing

models of dot clustering, our Bayesian estimation procedure is a special case of a more general model of perceptual grouping, namely Bayesian estimation of a mixture model (Feldman, Singh, & Froyen, submitted), and thus can readily generalize to other types of perceptual groups such as contours and shapes.

There is a large literature on the estimation of numerosity, both in adults (e.g. Miller & Baker, 1968), pre-verbal infants (e.g. Gelman & Gallistel, 1978; Xu & Spelke, 2000), as well as non-human primates (e.g. Brannon & Terrace, 1998). Most of the literature is concerned with the judgment of the number of *individual* items, rather than the number of clusters as in our study. But the two problems are intertwined because estimates of the number of items are influenced by way the items are grouped (Frith & Frith, 1972; Vos, Van Oeffelen, Tibosch, & Allik, 1988; Franconeri, Bemis, & Alvarez, 2009). More broadly, it can be argued that all numerosity judgments reflect prior perceptual grouping mechanisms which determine the underlying units to count (Feldman, 2003b).

In recent years many problems in perception have been modeled in a Bayesian framework, in which each potential interpretation $Y = \{y_1, \ldots, y_I\}$ of the stimulus $X$ is associated with the posterior probability $p(Y|X)$, which according to Bayes' rule is proportional to the product of the prior $p(Y)$ and the likelihood $p(X|Y)$ (see Kersten, Mammasian, & Yuille, 2004; Feldman et al., 2013). To model the dot grouping problems, we adopt the framework of *mixture models* (for a gentle introduction see Bishop, 2006). Specifically, we assume that the dot configuration $X = \{x_1, x_2, \ldots, x_N\}$ was generated by a mixture of distinct probabilistic sources

$$p(x) = \sum_{k=1}^{K} \pi_k g_k(x), \tag{2.1}$$

in which each of the $g_k$ are distinct generative components ("objects"), each having parameters $\theta_k$ and being chosen with probability $\pi_k$. As a simple assumption appropriate for dot clusters, we assume that each source is Gaussian in form, with parameters consisting of a mean $\mu_k$ and covariance matrix $\Sigma_k$, $\theta_k = (\mu_k, \Sigma_k)$. In other words, we assume that the image contains an unknown number $K$ clusters, each of which is generated

normally about some mean location with some spread. The observer's goal is to estimate which dots were generated from which source, and how many sources there are. This is a simple model of situations in which data is generated in spatial proximity to a localized source (here $\mu_k$), so that closely spaced image elements are more likely to have a common source.

The problem of estimating mixtures from data consists of estimating the $\theta_k$ and $\pi_k$ for each of the sources $g_k$. The difficulty lies in the fact that one does not know which datum belongs to which source ("ownership"). These two problems mutually depend on each other: the assignment of data to sources influences the estimation of the parameters of the sources, and conversely, the parameter values determine the probability that the datum arose from any particular source. These intertwined problems also characterize perceptual grouping, where each visual element can be regarded as belonging to any group, but the nature of the groups depends in part on which items seem to belong to it. Within the Bayesian framework presented above the fit of a particular mixture model, i.e. grouping interpretation $y_i$, is represented by $p(X|y_i)$. (For details of the estimation procedure see Appendix 5.1.)

An important application of this approach, which we test in the experiments below, is the estimation of the number of mixture components (clusters) $\hat{\mathcal{K}}$, where $\mathcal{K} = \{K_1 \ldots K_I\}$. Since each mixture model estimate $p(X|y_i)$ is related to a particular $K_i$, the subjective belief in a certain numerosity estimate $p(K_i|X)$, is directly captured by the belief in a certain grouping interpretation $y_i$. One possible estimate for $\hat{\mathcal{K}}$ is the MAP (maximum a posteriori) estimate, i.e. $\hat{\mathcal{K}} = \arg\max_{\mathcal{K}} p(\mathcal{K}|X)$. Estimating the full posterior is computationally expensive because $\mathcal{K}$ can take on any values in $[1, N]$. Hence, within the scope of the current chapter we will only compute the posterior for values of $\mathcal{K}$ allowed as responses by our subjects.

In the experiments below, we create displays with variable spatial configurations of dots, and ask subjects to judge the number of clusters. In Exp. 1, we constrain the responses to 1 vs. 2, in an attempt to gauge (and then model) the subjective threshold separating these two basic cases. In Exp. 2, we allow numerosities of 1, 2, or 3, which substantially complicates the geometric relations among the clusters. In both

experiments, the critical manipulation is the separation of the source components relative to their spreads (which is more complicated in Exp. 2 because of many variations in relative position among 3 components). For both experiments, we run the Bayesian model on the same configurations seen by subjects and compare its estimates to theirs.

## 2.3 Experiment 1

In this first experiment subjects counted clusters in very simple displays only containing two clusters (similar to Cohen et al., 2008). Subjects were shown dots sampled from two bivariate isotropic Gaussian clusters, for which we manipulated the distance $d$ between their underlying means, in order to manipulate the apparant number of clusters from one to two. In the experiments below we seek the threshold for subjects relative belief between one or two clusters. Furthermore, we tested how this threshold changed depending on the variance of the generating Gaussian clusters, $\sigma^2$. One can easily see that larger generating variances will result in larger thresholds. Lastly we also manipulated the number of points that were sampled, and investigate its influence on the estimated thresholds.

### 2.3.1 Methods

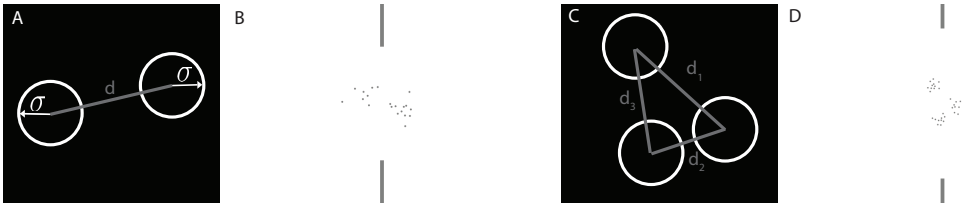


Figure 2.1: Stimulus setup and example stimuli for Exp.1 and Exp.2. A. depicts the setup of stimuli in Exp. 1 where the 6 conditions are defined by 3($\sigma$) and 2($N$), and adaptive staircases are ran over $d$. B. shows an example stimulus for Exp.1 for $\sigma = 0.8$dva, $N = 20$, and $d = 4.0$dva (contrast inverted). C. depicts the setup for part two of Exp 2. where we sampled $[d_1, d_2, d_3]$ at random from a predetermined distribution, while $\sigma$ and $N$ were kept fixed to 0.4dva and 30 respectively. D. shows an example stimulus for part two of Exp.2 (from subject FE's run) where $[d_1, d_2, d_3] = [3.7, 2.2, 3.1]$dva (contrast inverted).

**Participants**

Nine Rutgers University students, naive to the purpose of the experiment, participated for course credit.

**Stimuli**

Each stimulus consisted of a number N of dots sampled from a mixture of two bivariate isotropic Gaussian distributions (Eq. 2.1). Each Gaussian component was governed by a (common) variance $\sigma^2$ and a mean $\mu_k$, whose values were manipulated over the course of the experiment. The values of $\sigma^2$ and $N$ depended on the experimental condition (Fig. 2.1A). The adaptive procedure, which manipulates the distance, $d$, between the two component means, governs the position of the means relative to the center of the screen. More precisely, displays were created by first placing means at $\mu_1 = [-d/2, 0]$ and $\mu_2 = [d/2, 0]$, where $[0, 0]$ is the center of the screen and then applying a random rotation to the entire display. Assuming equal weights for each of the two components the following sampling procedure was repeated $N$ times: first, either of the two components was randomly selected with probability $\pi = .5$; second, the coordinates of dot $x_n$ (with diameter 12 minutes of arc) were randomly sampled from the bivariate isotropic Gaussian $\mathcal{N}(\mu_k, \sigma^2)$ in such a way that $x_n$ did not overlap with any of the other dots. These dots, which were midgray in color, were shown on a black background with two midgray bars on top and bottom of the display to guide the subjects fixation (Fig. 2.1B).

**Design and Procedure**

Subjects sat at 85 cm from a 20″ LCD monitor (60 Hz, 1680pxl x 1050pxl) connected to a Windows 7 PC, on which the displays were presented using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007). Subject ran 6 randomly interleaved adaptive staircases to estimate the distance threshold, $t$, between the two component means, for each of the experimental conditions, i.e. $3(\sigma = [0.4, 0.8, 1.2]\text{dva}) \times 2(N = [10, 20])$ (Fig. 2.1A-B). The adaptive procedure used was the Psi method (Kontsevich, Tyler, et al., 1999)

implemented in the Palamedes toolbox for Matlab (Kingdom & Prins, 2010; Prins & Kingdom, 2009). redThis method estimates both the threshold ($\hat{t}$) and its standard error ($SE_{\hat{t}}$), and the slope and its standard error of the underlying psychometric curve by selecting a value for $d$ at each trial that minimizes the expected entropy of the posterior distribution of the psychometric curve before that trial. Since the primary interest of this experiment was the threshold estimates, the slope estimates are not reported. Furthermore, each staircase ran for 100 trials, a number we found to yield small confidence intervals for the threshold estimates, but unreliable slope estimates.

Before the main experiment subjects ran 16 randomly selected training trials to acquaint them with the procedure. The main experiment then consisted of a total of 600 trials divided over four blocks of 150 trials each. Each block was followed by a mandatory one minute break. Each trial started out with a fixation display consisiting of two midgray bars on top of a black background (e.g. Fig. 2.1A without the dots) for 750 ms followed by the actual stimulus, which was shown for 250 ms. Thereafter the subjects responded if they saw one or two clusters of dots by means of the numbers on a numeric keyboard.

### 2.3.2 Results and Discussion

One out of the nine subjects that ran in this experiment was excluded because the threshold estimates for some conditions had standard errors larger than the standard deviation used to generate the clusters, $SE_{\hat{t}} > \sigma$.[1] Fig. 2.2A shows, for each subject, the estimated threshold, $\hat{t}$, and its $SE_{\hat{t}}$ (standard error) for each condition. As expected, there is a positive trend of threshold depending on $\sigma$, where the larger the $\sigma$ used to generate the clusters, the greater the separation $d$ required for subjects to perceive two clusters. Since the psi-method adaptive procedure only returns an estimate of the threshold and its standard error, we compute the significance of the positive trend for each of the subjects (and N) by means of a t-statistic computed between the $\sigma = 0.4$dva

[1] This criterion was chosen because this would mean that the standard error on the threshold estimate was larger than the standard deviation of the Gaussian process that generated the dots in that particular condition

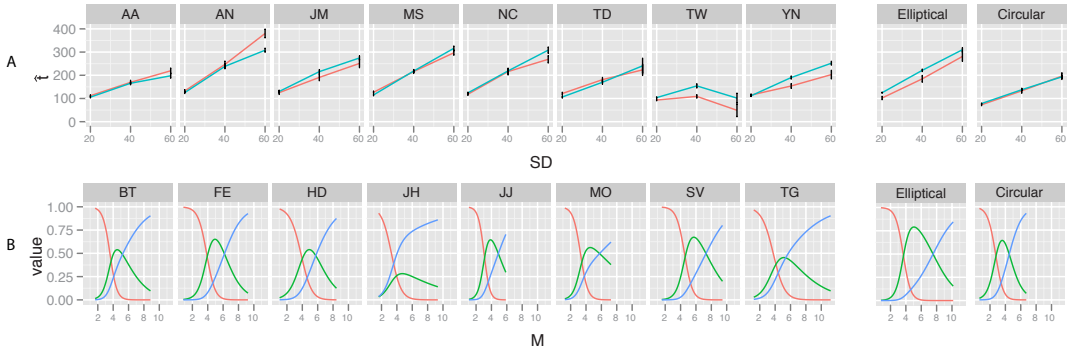


Figure 2.2: A. Results for Exp. 1 showing the threshold ($\hat{t}$) and standard error ($SE_{\hat{t}}$) for all conditions experiment (with red representing $N = 10$, and green $N = 20$). B. Results for Exp 2 showing the dependency between the modality paramete, $M$, and the numerical judgements as presented by the multinomial logistic regression model. Each of the colored line depicts the probability of a different numerocity hypothesis: red, $p(K = 1|M)$; green, $p(K = 2|M)$; blue, $p(K = 3|M)$. The plots on the left depict results from the subjects. The plots on the rights show the results of the two model version.

and $\sigma = 1.2$dva conditions. A two-tailed t-test then indicated that the positive trend was significant for all subjects ($t = [4.12, 20.04]$[2], $p < .001$), except TW. The number of dots (N), on the other hand, did not to have a consistent effect across $\sigma$ conditions and subjects, as can be seen in Figure 2.2A.

A natural question is whether subjects' judgments are scale-invariant, meaning that a scalar model with $\sigma$ as predictor ($\hat{t} = a\sigma$) would be a good explanation for the positive trend seen in the data (Fig. 2.2A). Since the psi-method only returns us with threshold values and their standard errors, we have to take a different route to test the scale invariance hypothesis. By dividing the thresholds by the $\sigma$ of that condition, we essentially compute the residuals of a scalar model containing $\sigma$ as a predictor. This is analogous to the computation of $D'$ (d-prime). If subjects were scale invariant this value, $\hat{t}/\sigma$, should be constant across different values of $\sigma$. By repeating the above t-statistic based analysis on this reparametrization of $\hat{t}$ we found that most subjects were not scale invariant for both levels of N ($t = [-9.13, -2.17]$, $p < .05$), with the exception of AN in the $N = 10$ condition, for which we could not reject the scale invariance

[2]This shows the range of t-statistic values obtained across subjects. We will continue to use this notation in later parts of the dissertation

hypothesis. Moreover, there seems to be a negative relation between $\hat{t}/\sigma$ and $\sigma$.

## 2.4 Experiment 2

In Experiment 2 we increased the number of possible numerosities to 1,2, or 3. This substantially complicates the geometric relations among the clusters. More specifically in Exp. 1, each of the possible displays was generated by only two isotropic Gaussian clusters, and the means governing them could hence be parametrized by the distance between them. However in the current experiment each of the displays was generated by three clusters, increasing the number of distances needed to define the means for the clusters, to three. Since, manipulating these three distances in a balanced design would have required an enormous number of trials, we choose to randomly sample the three distances for each trial, and use these to define the (relative) cluster means. In order to get a good sample around informative values, i.e. threshold values between two numerosity judgements, we first ran every subject on one staircase as in Experiment 1 to find out their distance threshold $\hat{t}$ for the two cluster case. The three distances were then sampled from a Gaussian distribution with $\mu = \hat{t}$, and variance $\sigma_d^2$ (see Methods). Furthermore in order to get a general idea about how these three distance values influence numerosity judgements, we collapsed them into a measure of cluster separability for each trial. This measure, also called the modality parameter (Feldman, 2012) , much like a traditional F-statistic, takes the ratio between the between cluster standard deviation ($SD_\mu$) and the standard deviation of the clusters (referred to as $\sigma$). Formally this measure is defined as $M = (2SD_\mu)/\sigma$. The larger the value of $M$ the higher the likelihood that more than one cluster is present.

### 2.4.1 Methods

#### Participants

Eight Rutgers University students, naive to the purpose of the experiment, participated for course credit.

**Stimuli and Procedure**

The first part of Exp.2 was identical to one staircase in Exp.1 where $\sigma = 0.4$dva and $N = 20$. The threshold for $\hat{t}$ acquired through this procedure was then used to generate stimuli for the second part.

The stimuli in the second part consisted of 30 dots sampled from a mixture of three bivariate isotropic Gaussian distributions (Eq. 2.1). Each of the components was governed by a fixed standard deviation $\sigma = 0.4$dva and a mean $\mu$, which was chosen as follows. The three means $\mu$ are vertices of a triangle with edges of length $[d_1, d_2, d_3]$. Given $\hat{t}$ found in the first part of the experiment these edges are sampled from a Normal distribution, $d_i \sim \mathcal{N}(\hat{t}, \sigma_d)$, bounded by $[0, +\infty[$. We set $\sigma_d = \hat{d}/1.96$ so that $p(d_i > 0) = 0.95$. All three edges were sampled in this way such that the triangle inequality ($d_1 < d_2 + d_3$) holds (Fig. 2.1C-D). The goal of this procedure was to create a wide variety of configurations of cluster geometries. Subjects were tested in the same environment using the same protocol as in Exp. 1. Each subject ran a total of 1000 trials, randomly generated in the fashion outlined above, split into eight blocks.

### 2.4.2 Results and Discussion

For each of the subjects we fitted a multinomial logistic regression model to their numerosity judgements, with as independent variable the modality parameter $M$. By means of a likelihood ratio-test comparing a model containing $M$ and a unconditional means model (containing only an intercept), we found that $M$ was a good predictor for all subjects' numerosity judgements ($LR = [251, 907]$, $df = 2$, $p < 0.001$). That is, as in Exp. 1, subjects responses were substantially driven by the degree to which the clusters were separated relative to their spreads. Figure 2.2B shows the multinomial logistic regression models for each of the subjects. One can easily see that with increasing values of $M$ higher numerosity judgements become more and more likely. Note however, although $M$ is a good predictor for the range of stimuli tested here it is not a universal predictor of number.

## 2.5 A Bayesian Model for these Tasks

In the current experiments only a limited set of numerosities were considered, summarized in the posterior $p(\mathcal{K}|X)$, where $\mathcal{K} = \{1,2\}$ for Exp. 1 and $\mathcal{K} = \{1,2,3\}$ for Exp. 2. The likelihood of each hypothesis $p(X|K_i)$ was assessed by the fit of a mixture model with $\mathcal{K} = K_i$, to a given geometric configuration of dots $X$. The prior belief $p(\mathcal{K})$ over all of the hypotheses was assumed to be uniform. redOne might point out that hypotheses containing fewer clusters should have a more favorable prior, however true, such a trade-off between complexity and fit of the hypotheses is already embodied by the likelihood $p(X|K_i)$ through what is knows as Bayes Occam (see Appendix 5.1). Together the prior and likelihood define the posterior probability for $K_i$:

$$p(K_i|X) = \frac{p(X|K_i)p(K_i)}{\sum_j p(X|K_j)p(K_j)}. \tag{2.2}$$

This posterior was computed for two version of the model. In one version, the "elliptical" model ($M_e$), the model is free to estimate the full covariance matrix $\Sigma$ for each of the component distributions, amounting in a total of 5 parameters ($\mu_x$,$\mu_y$,$\sigma_x$,$\sigma_y$,$\sigma_{xy}$) to be estimated, meaning that each cluster is assumed to be elliptical in shape. In the other version, called the "circular" model ($M_c$), the model is constrained to only circular component distributions (where $\sigma_x = \sigma_y$ and $\sigma_{xy} = 0$), leaving only three parameters to be estmated ($\mu_x$,$\mu_y$,$\sigma$). We included both model versions because it is unclear a priori which model subjects would adopt, and they can lead to substantially diferent numerosity estimates. For example, a subject assuming an elliptical model might judge an elongated cloud of dots to comprise of one cluster, whereas one with a circular model might prefer two clusters. (Appendix 5.2 shows how both of these assumptions can be encompassed by a prior over the covariance matrix $\Sigma$).

We evaluate the equivalence of our model to the human data in three ways. First, we will let the model run the same experiment as the subjects, with double the amount of trials per staircase (i.e. 200) for Exp. 1 and a total of 1000 samples for Exp. 2. For each trial the model computes the posterior as shown above, yielding a numerosity

estimate by sampling from this posterior. redEven though such a decision strategy, also referred to as *probability matching*, is sometimes regarded as suboptimal it appears to be used in some perceptual tasks (Mamassian & Landy, 1998; Wozny, Beierholm, & Shams, 2010). Secondly, we also computed how well each of the two versions explained subject data. This was done by first computing the posterior, $p(K|X_v)$ for each dot-configuration $X_v$ shown to the subjects during their run of the experiments. Subsequently we computed how likely each of the subjects' responses $R = \{r_1...r_v\}$ were generated by either of the two versions on the model. The ratio between both likelihoods yielded the Bayes factor, $BF_{ec} = p(R|M_e)/p(R|M_c)$, indicating how much more likely subjects' responses were generated under the elliptical assumption versus the circular. Finally, we compared how well our model explained the human subjects' data to the standard CODE model (henceforth referred to as $\text{CODE}_1$) by Van Oeffelen and Vos (1982). The CODE model superimposes a Gaussian function $f_n$ (i.e. a Gaussian distribution without the normalization factor) centered on each of the dots $x_n$ with a standard deviation $s_n$. In the standard account $s_n$ is defined as $s_n = d_n/2$, where $d_n$ is the distance of $x_n$ to its nearest neighbor. Clusters are then found by computing the mixture of all these Gaussians $F(x) = \sum_n f_n(x)$ and finding the contour where $F(x) = 1$. The number of closed contours found is then said to be the estimated number of clusters, which could range from $[0, +\text{inf}]$. In order to compare this model to our own, we also ran it on all the dot-configurations shown to the subjects. We then computed the proportion of trials in which the $\text{CODE}_1$ model gave the same response as the subject, $p(correct)$. Early pilot test with the standard $\text{CODE}_1$ model showed its performance to be severly lacking. We therefore also tested an augmented version of the CODE model, just as one of the versions tested by Compton and Logan (1993), in which we ensured a more global influence of the dots upon each other by setting $s_n = d_n$ (further referred to as $\text{CODE}_2$). The percent correct of these two models was then compared to our models percent correct which was computed as $\sum_n p(r_v|M)$.

| Experiment 1 | | Experiment 2 | |
|---|---|---|---|
| ID | $log(BF_{ec})$ | ID | $log(BF_{ec})$ |
| AA | -215 | BT | 142 |
| AN | 827 | FE | 1857 |
| JM | 693 | HD | 1519 |
| MS | 432 | JH | -912 |
| NC | 524 | JJ | -325 |
| TD | 55 | MO | 137 |
| TW | -92 | SV | 3540 |
| YN | 104 | TG | 869 |

Table 2.1: Fitting results for both experiments indicating the Bayes Factor log($BF_{ec}$ of the elliptical versus the circular version of the model. Positive numbers indicate that this particular subjects results were more likely to be generated by the elliptical model, while negative numbers indicate a circular assumption.

### 2.5.1 Model Performance

As shown in Fig 2.2A, the model's performance closely matches that of subjects in Exp. 1. Analyzing both model versions in the same way as the subjects' data (see Exp. 1 for details) yielded a significant positive effect of $\sigma$ (elliptical, $t = \{8.18, 20.34\}$[3], $p < .001$; circular, $t = \{8.54, 18.73\}$, $p < .05$). Furthermore one can clearly see that the positive trend is different for the two versions of the model, with a shallower slope for the circular model, and lower threshold values in general. This is a clear side effect of the shape assumption underlying the model versions. The same scale invariance analysis as in Exp. 1 yielded that both model versions were not scale invariant for $N = 20$ (elliptical, $t = -5.96$, $p < .001$; circular, $t = -5.57$, $p < .001$), on the other hand scale invariance could not be rejected for $N = 10$ (elliptical, $t = -0.86$; circular, $t = -1.26$). One could compare each of the two model versions to the subject data by merely eyeballing the plots shown in Fig. 2.2, and get an idea which subject held which shape assumption. A more objective measurement is given by the Bayes factor $BF_{ec}$ as described above. Table 2.1 shows the logarithm of $BF_{ec}$ for each of the subjects. Most subjects seems to adhere to a elliptical hypothesis.

A similar close correspondence between the model and the subjects' data was found for Exp. 2 (Fig. 2.2B). As was the case for the subjects the modulation parameter

[3] t-statistic for N=10 and N=20.

$M$ was found to be a significant predictor for both models' numerosity judgements, as shown by a likelihood ratio test between a multinomial regression model containing $M$ as a predictor and an unconditional means model (elliptical, $LR = 857$, $p < 0.001$; circular, $LR = 907$, $p < 0.001$). Which subject held which shape assumption is shown in Table 2.1. Again, as in Exp. 1, most of the subjects' data in Exp.2 was best explained by the model making the elliptical assumption.

Our model clearly outperformed the standard $\text{CODE}_1$ model (Fig. 2.3) in both experiments. Specifically, $\text{CODE}_1$ hardly ever gets the numerosity estimate right, especially in Exp. 2. As can be seen in Fig. 2.3 augmenting the CODE model to take into account more global aspects of the configuration clearly benifits the model $\text{CODE}_2$, resulting in performance close to our own models $M_c$ and $M_e$. The difference in overall performance between both versions of our model was marginal.

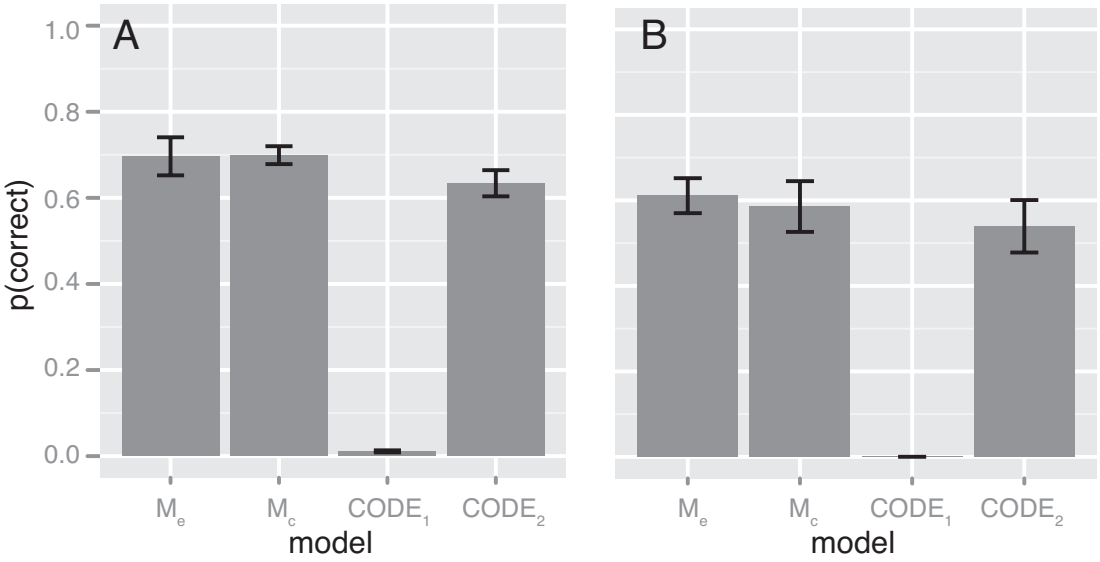


Figure 2.3: Comparison of each of the models performance as operationalized by the proportion of correct numerosity judgements for each of the subjects. Here the average proportion correct across all subjects and 95% confidence interval ($1.96 \times \text{SE}$) is shown for Exp. 1 (A) and Exp. 2 (B).

## 2.6 General Discussion

In this chapter we studied how subjects group dots into clusters, and more particularly how they decide how many clusters are present. Across a set of two experiments we

tested numerosity judgments for clouds of dots sampled from mixtures of bivariate Gaussian clusters. We showed that subjects numerosity judgements were substantially driven by the degree to which clusters were separated relative to their spreads. A similar result has been found using a more indirect measure in case of only two clusters (Cohen et al., 2008). However, at least in Exp. 1 we were able to show that this relationship was not a scalar one, indicating that subjects were not scale invariant. We modelled these results in a Bayesian framework in which different grouping hypotheses, and thus numerosity estimates, are assigned probabilities. In order to model these grouping hypotheses we adopted a mixture model framework, that estimates the posterior probability of each grouping hypothesis. We found our model to give an accurate and quantitatively precise account of subjects' numerosity responses. Furthermore, we tested two versions of the model to accommodate for possible prior assumption about cluster shape that might be held by the subjects. We found that only a few subjects (4/16) tended to assume the clusters were constrained to take on circular shapes, while all the others had fewer constraints on shape, i.e. they assumed elliptical shapes.

Our model outperformed the standard CODE model (Van Oeffelen & Vos, 1982). When augmenting the standard CODE model to take into account more global aspects of the dot configurations its performance came close to our own models. Moreover, the approach proposed here has several benefits over such traditional approaches. First of all the current model is not only applicable to Gaussian dot clusters, but is a framework that is generalizable to other problems in perceptual grouping. The current model assumed that each of the components was Gaussian in form. A byproduct of imposing such a monotonically decreasing density function is that more closely elements are more likely to be generated from a common source, i.e. grouped together. Thus in essence our model is a version of the Gestalt principle of proximity. In other words the Gestalt principle of proximity be can thought of as a heuristic to solve a mixture model with Gaussian components. In a more general sense Gestalt principles can be seen as heuristics to solving mixture models with different classes of component distributions (Feldman et al., submitted). For example above Bayesian framework could be extended

to other problems in perceptual organization such as contour integration, where edges are generated from underlying contours. Secondly, unlike the CODE model, a Bayesian approach has the advantage of being able to assign different degrees of belief to different grouping hypotheses, allowing us to model the often intermediate judgements present in subjects data. redThird, Bayesian inference makes optimal use of the information and assumptions available to the observer (Jaynes, 2003).

The current chapter puts forward an approach showing the interconnection between perceptual grouping and numerosity estimation. We claim that perceptual grouping precedes numerosity estimation, in that accurate numerosity judgments require an image to be segmented into distinct objects. These objects can be clusters as in the current chapter, or the dots themselves as in traditional numerosity literature. The framework we propose shows how a model for perceptual grouping can make numerosity judgements for Gaussian clusters as objects, but as discussed above can easily be extended to different objects by changing the class of component distributions. We hope the framework will pave the road for more principled models of numerosity estimation driven by principles of perceptual grouping.

# 3. Bayesian hierarchical grouping

## 3.1 Abstract

Perceptual grouping is the process in which image elements are grouped into distinct units or "objects". We propose a Bayesian framework for grouping in which the goal of the computation is to estimate the hierarchical representation that explains the observed configuration of the image elements. We formalize the problem of perceptual grouping as a mixture estimation problem, where it is assumed that the configuration of image elements is generated by a set of distinct components (or "objects"). This framework can easily be adjusted to different subproblems of perceptual grouping by changing the object definitions. In the current chapter we discuss an instantiation of our framework for contour integration, part decomposition, and shape completion. We show how the framework can account for several perceptual phenomena, as well as account for human subject data from previous experiments for these specific grouping problems. In the end the framework proposed here gives us insight in how image elements are grouped together into distinct objects.

## 3.2 Introduction

Perceptual grouping is the process in which image elements are grouped into distinct clusters or objects. The problem of grouping is inherently difficult because the system has to choose the best grouping interpretation among many. Specifically, as the number of image elements $N$ increases, the number of possible grouping interpretations increases exponentially with $N$. Despite the many studies on different aspects of perceptual grouping (for review see Wagemans et al., 2012a; Feldman et al., 2013) the mechanisms underlying them are poorly understood.

Many different models have been proposed for subproblems of perceptual grouping, such as contour integration (e.g. Field, Hayes, & Hess, 1993; Geisler, Perry, Super, & Gallogly, 2001; Ernst et al., 2012), completion (e.g. Van Lier, 1999; Kalar, Garrigan, Wickens, Hilger, & Kellman, 2010) and figure-ground organization (e.g.

Sajda & Finkel, 1995; Craft, Schütze, Niebur, & von der Heydt, 2007; Froyen, Feldman, & Singh, 2010). Even other problems in visual perception such as part decomposition can, as we will show below, be cast as a perceptual grouping problem (e.g. Siddiqi & Kimia, 1995; Singh, Seyranian, & Hoffman, 1999). However, often these models describe the underlying mechanism in terms of somewhat underdetermined and poorly understood Gestalt principles or heuristics. Even though these models can make predictions about human perceptual grouping within their particular focus, they fail to assign subjective beliefs to these grouping hypotheses. Such is important to make quantitative predictions about subject behavior. Furthermore most of these models are tailored to a specific subproblem of perceptual grouping. Hence they are not easily generalized to other problems within perceptual grouping.

redIn this chapter we propose a mathematically rigorous and generalized framework for perceptual grouping resulting in a formal definition of perceptual grouping. More specifically we will discuss a specific instantiation of it for contour integration, shape completion, and part-decomposition.

## 3.3 The computational framework

In recent years Bayesian models have been developed to explain a variety of problems in visual perception. In these models each possible interpretation $\mathbf{C} = \{\mathbf{c_1} \dots \mathbf{c_J}\}$ of an image $D$ is related to a posterior $p(\mathbf{C}|D)$, which according to Bayes rule is proportional to the product of a prior $p(\mathbf{C})$ and likelihood $p(D|\mathbf{C})$ (for review see Kersten et al., 2004; Feldman et al., 2013)[1]. In recent papers we proposed that in order to model grouping problems, in particular dot clustering, a mixture model framework can be adopted to compute the probability of a particular grouping hypothesis $p(\mathbf{c_j}|D)$ (Feldman et al., submitted, and Chapter 2). Such a model has been shown to give quantitative and accurate predictions of human subjects judgments in the case of dot clustering (Chapter 2).

[1]Note that the notation in this chapter is slightly different from Chapter 2, due to the more elaborate mixture model definitions that will be adopted

### 3.3.1 An image is a mixture of objects

Let $D = \{x_1 \dots x_N\}$ denote the image data with each representing a 2-dimensional vector in $\mathbb{R}^2$. A mixture density is a probability distribution that is composed of the weighted sum of K components or objects labeled $\{1 \dots K\}$,

$$p(x_n|\phi) = \sum_{k=1}^{K} p(x_n|\theta_k)p(c_n = k|\mathbf{p}) \tag{3.1}$$

where $c_n \in \mathbf{c} = \{c_1 \dots c_N\}$ are the object assignments, $\mathbf{p}$ is a parameter vector of a multinomial distribution with $p(c_n = k|\mathbf{p}) = p_k$, $\theta_k$ are the parameters of the kth object, and $\phi = \{\theta_1, \dots, \theta_K, \mathbf{p}\}$. Even though often each of these objects takes on simple form, the resulting mixture can be highly complex and irregular in structure. The problem of mixtures is to represent such a complex dataset as the result of a combination of homogeneous objects (McLachlan & Basford, 1988). Depending on the task at hand these objects can take on different forms. In a case as simple as clustering dots, the image data could be said to be generated by a mixture of Gaussian objects with a mean, $\mu_k$ and covariance matrix $\Sigma_k$ (Chapter 2). However for more complex classes of grouping problems such as contour integration or part-decomposition different types of objects will have to be defined. In case of contour integration we will define that an image as a mixture of contours, while in case of part-decomposition we will define a shape as a mixture of parts. Below we will describe a generalized object class that can easily be tailored to generate image data as if it were a part, a contour, or a cluster.

To obtain a full Bayesian formulation of mixture models we define a prior over the object parameters $p(\theta|\beta)$ and over the mixing distribution $p(\mathbf{p}|\alpha)$. The former prior defines, on top of the object definition $p(x|\theta)$, what our prior beliefs are about what the objects look like before even seeing the image $D$. The latter prior is the natural conjugate prior for the mixing distribution, the Dirichlet distribution with parameter $\alpha$. When $\alpha > 1$ there is a bias towards more objects each explaining a small number of image data, while when $0 < \alpha < 1$ there is a bias towards fewer objects each explaining a large number of image data. Using these two priors we can rewrite the mixture model in Eq. 3.1 to compute the probability of a particular grouping hypothesis $\mathbf{c_j}$. The

likelihood of a particular grouping hypothesis is computed by marginalizing over the parameters $\theta_k$, $p(D|\mathbf{c_j},\beta) = \int \prod_{n=1}^{N} p(x_n|\theta_{c_n}) \prod_{k=1}^{K} p(\theta_k|\beta) d\theta$. This results in,

$$p(\mathbf{c_j}|D,\alpha,\beta) \propto p(D|\mathbf{c_j},\beta)p(\mathbf{c_j}|\alpha), \tag{3.2}$$

where $p(\mathbf{c_j}|\alpha) = \int p(\mathbf{c_j}|\mathbf{p})p(\mathbf{p}|\alpha)d\mathbf{p}$ is a Dirichlet integral. redNote that rewriting the mixture model in this way decomposes the right-hand side of the equation into two intuitive factors: a likelihood depicting how well the current grouping hypothesis $\mathbf{c_j}$ fits the data $D$ given object prior $\beta$; and a prior depicting the complexity of this grouping hypothesis $\mathbf{c_j}$, i.e. assigning the image data to the objects in this way, given the mixing prior $\alpha$. Unfortunately the posterior over all possible assignments $\mathbf{c_j}$ is intractable even for a fixed number of components $K$ (Gershman & Blei, 2012). For many clustering problems, such as perceptual grouping we often do not know which grouping hypothesis $\mathbf{c_j}$ to test or the number of objects that are present. For such cases we will need to generalize the above finite mixture model formulation to allow for an infinite number of objects. In other words, the number of objects is now considered a free parameter. One can easily see that in that case estimating the posterior over grouping hypotheses becomes even less tractable. Several approximation methods have been proposed to compute this posterior, such as Markov-Chain Monte Carlo (McLachlan & Peel, 2004) or variational methods (Attias, 2000). In the current chapter we choose a method that conforms with the idea that perceptual organization is hierarchical, the Bayesian Hierarchical Clustering method as proposed by Heller and Ghahramani (2005).

### 3.3.2 Bayesian hierarchical grouping

The idea that perceptual organization tends to be hierarchical is hardly novel (e.g. Pomerantz, Sager, & Stoever, 1977; Palmer, 1977; Baylis & Driver, 1993; Lee & Mumford, 2003; Marr & Nishihara, 1978). Formally a hierarchical structure corresponds to a tree where the root node represents the image data at the most global level, i.e. the grouping hypothesis postulates that all image data is generated by one underlying object. Subtrees then describe finer and more local relations between image data,

all the way down to the leaves, which explain only one image datum $x_n$ (i.e. one object for each datum). Even though this formalism has become popular over the recent years (Amir & Lindenbaum, 1998; Shi & Malik, 2000; Feldman, 1997b, 2003a), no formal framework has been proposed on how to build this tree structure for a particular image. Within the field of machine learning many different methods have been proposed for hierarchically clustering data (for overview see Bishop, 2006). In this chapter a method that integrates the idea of understanding grouping as a mixture model problem with the idea of hierarchical clustering is adapted to the setting of perceptual grouping.

The Bayesian hierarchical clustering algorithm (BHC) is similar to traditional agglomerative clustering methods (Heller & Ghahramani, 2005), with its main difference being in how the algorithm uses a Bayesian hypothesis test to decide when to merge clusters. Here we will, in a general manner, explain the workings of this algorithm. Given the dataset $D = \{x_1 \ldots x_N\}$, the algorithm is initiated with $N$ trees $T_i$ each containing one data point $D_i = \{x_n\}$. At each stage the original BHC algorithm would consider merging all possible pairs of trees. Then, by means of the statistical test explained below it is decided which two trees $T_i$ and $T_j$ to merge, resulting in a new tree $T_k$, with its associated merged dataset $D_k = D_i \cup D_j$ (Fig. 3.1A). However, testing all possible pairs in the context of perceptual grouping is rather intractable, hence to increase performance we propose the following. In our implementation of the BHC we only consider pairs of trees that have data points near each other as defined by Delaunay triangulation, i.e. Delaunay-consistent pairs. Doing so substantially reduces the number of pairs to be tested. redSpecifically while the original approach in its initial phase would have a complexity of $O(N^2)$, our approach reduces this initial complexity to $O(N \log(N))$ (also see App. 5.3).

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 $D_k$ is generated by only one underlying object $p(D_k|\theta)$, with unknown parameters $\theta$. In order to evaluate the probability of the data given this hypothesis $p(X|\mathcal{H}_0)$ we introduce priors over the objects as described above $p(\theta|\beta)$, in order to integrate over the to be estimated

parameters $\theta$,

$$p(D_k|\mathcal{H}_0) = \int_\theta \prod_{x_n \in D_k} p(x_n|\theta)p(\theta|\beta) \tag{3.3}$$

For simple objects such as Gaussian clusters this integral can be computed analytically. In case of more complex objects this integral becomes less tractable (see App. 5.4).

The second hypothesis, $\mathcal{H}_1$, is the sum of all possible partitionings of $D_k$ into two or more objects. However, as already mentioned above such computation is intractable. Therefore the BHC algorithm circumvents this problem by restricting itself to partitionings that are consistent with the tree structure of the two to be merged trees $T_i$ and $T_j$. For example for the tree structure as in Fig. 3.1A the possible tree-consistent partitionings are shown Fig. 3.1B. As you can see many possible partitionings are not considered. So, the probability of the data under $\mathcal{H}_1$ can be computed by taking the product over the subtrees $p(D_k|\mathcal{H}_1) = p(D_i|T_i)p(D_j|T_j)$. Below we will define $p(D_i|T_i)$, and it will become clear how this sum over all partitionings can easily be computed recursively as the tree is built.

In order to get the marginal likelihood of the data under the tree $T_k$ we need to combine $p(D_k|\mathcal{H}_0)$ and $p(D_k|\mathcal{H}_1)$. In this way we get the probability of the data integrated across all possible partitions, including the one object hypothesis $\mathcal{H}_0$. Weighting these hypotheses by a prior on all data $D_k$ being explained by one object $p(\mathcal{H}_0)$, we get our definition for computing $p(D_i|T_i)$ when building the tree,

$$p(D_k|T_k) = p(\mathcal{H}_0)p(D_k|\mathcal{H}_0) + (1 - p(\mathcal{H}_0))p(D_i|T_i)p(D_j|T_j). \tag{3.4}$$

Note that $p(\mathcal{H}_0)$ is also computed bottom-up as the tree is built and is based on a Dirichlet process prior (Eq. 3.6, for details see Heller & Ghahramani, 2005). We will discuss this prior in further depth below. Given the above equation, the probability of the merged hypothesis $p(\mathcal{H}_0|D_k)$ can easily be found by means of,

$$p(\mathcal{H}_0|D_k) = \frac{p(\mathcal{H}_0)p(D_k|\mathcal{H}_0)}{p(D_k|T_k)} \tag{3.5}$$

This probability is then computed for all the Delaunay-consistent pairs. The pair with the highest probability of merging is then merged. In this way the algorithm greedily builds the tree until all data is merged into one cluster. We will refer to our implementation of the BHC for perceptual grouping as Bayesian hierarchical grouping (BHG).

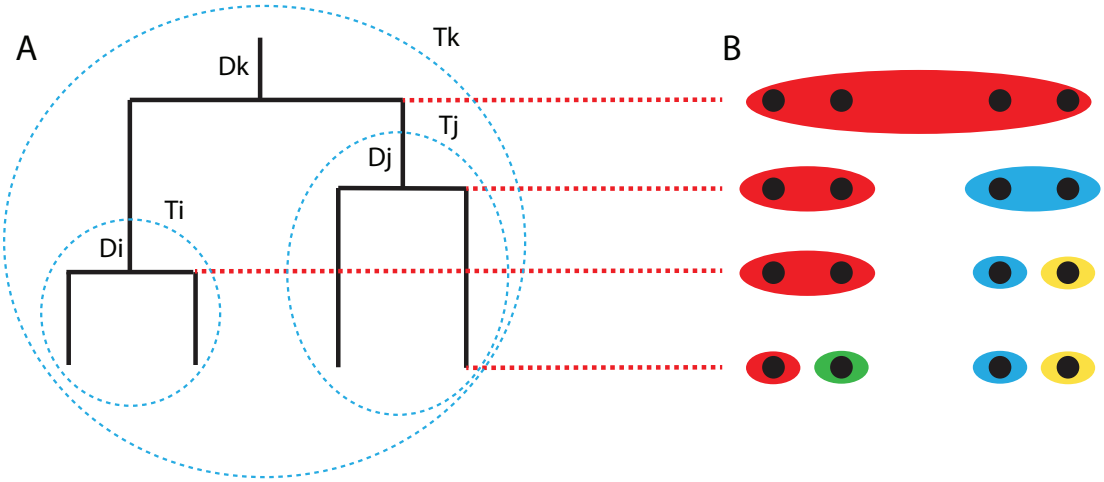


Figure 3.1: Explaining the BHC process. A. Tree decomposition (adapted from Heller & Ghahramani, 2005); B. Tree slices, i.e. different grouping hypotheses.

**Tree-slices**

During the construction of the tree one can simply find the MAP decomposition by splitting the tree once $p(\mathcal{H}_0|D_k) < .5$. However, especially in perceptual grouping, we are more often interested in the distribution over all the possible grouping hypothesis that were considered. In order to do so we need to build the entire tree, and subsequently take what are called tree slices at every level in the tree (Fig. 3.1B), and compute their respective probabilities $p(\mathbf{c_j}|D,\alpha,\beta)$. Since the BHC model employs infinite clustering rather than finite clustering, the Dirichlet prior as explained above is no longer applicable to compute this quantity. We therefore turn to its infinite variant, the Dirichlet Process prior (similar to the Chinese restaurant process). This process was independently discovered by Anderson (1991) in the context of categorization. Formally this prior is defined as

$$p(\mathbf{c_j}|\alpha) = \frac{\Gamma(\alpha)\alpha^K \prod_{k=1}^{K} \Gamma(n_k)}{\Gamma(N+\alpha)}, \tag{3.6}$$

where $n_k$ is the number of datapoints explained by object with index $k$. Inserting Eq. 3.6 into Eq. 3.2 we can easily compute the posterior distribution across all tree-consistent decompositions of data $D$:

$$p(\mathbf{c_j}|D,\alpha,\beta) \propto p(\mathbf{c_j}|\alpha) \prod_{k=1}^{K} p(D_k|\beta) \tag{3.7}$$

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

**Prediction**

For any grouping hypothesis, we can compute the probability of a new point $x^*$ given the data $D$. This distribution is called the posterior predictive $p(x^*|D,\mathbf{c_j})$. As in Eq. 3.1, the new data is generated from a mixture model governed by $K$ components as present in this particular grouping hypothesis. More specifically new data is generated as a weighted sum of predictive distributions $p(x^*|D_k) = \int p(x^*|\theta)p(\theta|D_k,\beta)$ (where $D_k$ is the data associated with object $k$),

$$p(x^*|D,\mathbf{c_j}) = \sum_{k=1}^{K} p(x^*|D_k)\pi_k. \tag{3.8}$$

Here $\pi_k$ is the posterior predictive of the Dirichlet process prior defined as $\pi_k = (\alpha + n_k)/\sum_{i=1}^{K}(\alpha + n_i)$ (see Bishop for a derivation). Using this approach the framework we propose for perceptual grouping is able to make predictions about missing parts of shapes, as for example is the case in amodal completion. Several examples of this will be shown in the results section.

### 3.3.3 The objects

In the general framework presented above the objects can take on any form. For simple dot clustering problems we can assume Gaussian objects, governed by a mean, $\mu_k$ and

a covariance matrix $\Sigma_k$. We have previously shown that such a representation accurately and quantitatively predicts subject cluster enumeration judgments (Chapter 2). However, more complex grouping problems such as part-decomposition and contour integration call for a more elaborate object definition. In the current instantiation of the framework the objects are represented as B-spline curves $G = \{g_1 \dots g_K\}$ (Figure 3.2), each governed by a parameter vector $\mathbf{q_k}$. Given this underlying curve, datapoints $x_n$ are generated perpendicular to the curve as follows,

$$p(x_n|\theta_k) = \mathcal{N}(d_n|\mu_k, \sigma_k), \tag{3.9}$$

where $d_n = \|x_n - g_k(n)\|$ is the distance between the datapoint $x_n$ and its perpendicular projection to the curve $g_k(n)$ (also referred to as the riblength), $\mu_k$ is the mean riblength, and $\sigma_k$ is the variance on the riblength for this component. Put together the parameter vector for each component is defined as $\theta_k = \{\mu_k, \sigma_k, \mathbf{q_k}\}$. One can easily see in Figure 3.2, how by formulating the generative function as such it can be adapted to generate either image data coming from a contour like object (Fig. 3.2A) when $\mu_k = 0$. On the other hand when $\mu_k > 0$ the image data is generated with some distance from the curve, as if an axis was generating a part (Fig. 3.2B). Furthermore given an object with $\mu_k = 0$, and a larger variance $\sigma_k$ the image data generated will look like dots generated from a cluster. redNote that the generative function is symmetric along the underlying curve (Fig. 3.2), which results in the object class indirectly incorporating a symmetry constraint (e.g. Kanizsa & Gerbino, 1976; Machielsen, Pauwels, & Wagemans, 2009) when estimating said underlying curve for a given set of data $D$.

As noted above in Bayesian mixture models we furthermore need to define a prior on these parameters $p(\theta|\beta)$. Given the definition of our objects we introduce two sets of priors for the object parameters $\theta$. A first set is introduced on the shape of the underlying curve $g_k$. Specifically, a bias towards short curves was introduced by means of a prior on the arclength of the curve, as computed by $F_{k1} = \int \|g_k'\|^2$, $F_{k1} \sim exp(\lambda_1)$. Furthermore a bias towards straight curves was introduced by means of the total curvature along the curve as computed by $F_{k2} = \int \|g_k''\|^2$, $F_{k2} \sim exp(\lambda_2)$. Both

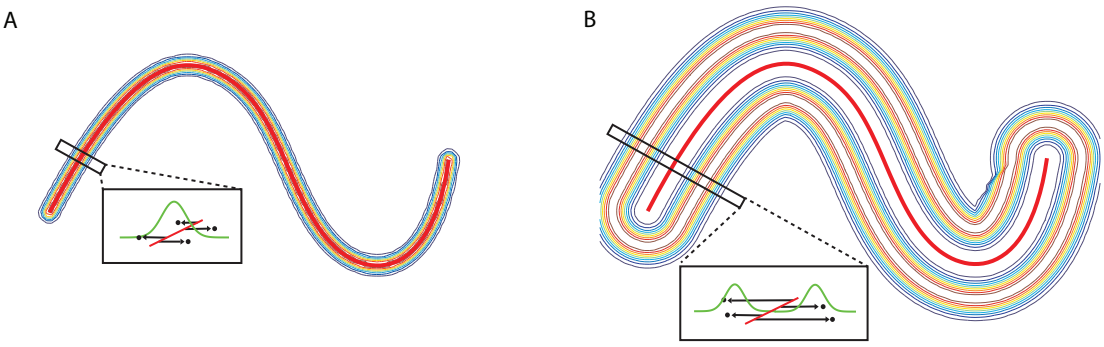


Figure 3.2: The generative function of our model depicted as a field. Ribs sprout perpendicularly from the curve (red) and the length they take on is depicted by the contour plot. A. For snakes ribs are sprouted with a $\mu$ close to zero, resulting in a Gaussian fall-off along the curve. B. For shapes ribs are sprouted with $\mu > 0$ resulting in a band surrounding the curve.

values were computed numerically (see Appendix 5.4). A second set of priors was introduced onto the function that generated the datapoints from this curve (Eq. 3.9). We introduced a Normal-$inv(\chi^2)$ distribution as the conjugate prior for the normal distribution with parameters $\{\mu_0, \kappa_0, \nu_0, \sigma_0\}$. $\mu_0$ is our prior belief of the mean riblength and $\kappa_0$ defines how strongly we believe this; $\sigma_0$ is our prior belief of the variance of the riblength, and $\nu_0$ defines how strongly believe this. Putting all this together the object hyperparameter vector is defined as $\beta = \{\mu_0, \kappa_0, \nu_0, \sigma_0, \lambda_1, \lambda_2\}$. Together the generative function (Eq. 3.9) and the object prior define the object class. The object hyperparameter vector then governs our belief about what we think an object looks like within this particular object class, and is what makes this particular object class flexible enough to account for different spatially defined perceptual grouping problems.

We proposed a mathematically rigorous and coherent framework for understanding perceptual grouping based on one central idea: *an image is a mixture of objects*. In what follows we will show how the specific instantiation of the framework discussed here can account for several perceptual phenomena and human subject data in domains such as contour integration, part decomposition and shape completion.

## 3.4 Results

We will show how our model explains many known findings in contour integration, part decomposition and shape completion by means of comparisons to instant-psychophysics and previously collected human subject data. On top of that the model is able to make novel predictions within these different fields of perceptual grouping, hopefully sparking new experiments and findings in the future.

### 3.4.1 Contours

As already mentioned one of the perceptual grouping problems our computational framework is able to handle is the integration of dots into contours. Before we can run the model on different stimuli within this class, we need to define what we assume a contour looks like. In other words we need to set the hyperparameters $\beta$ that define the objects. In the case of contours we can assume that the mean riblength is very close to zero, that is edges are generated as a Gaussian falloff from the actual curve. This is reflected in the hyperparameters by setting $\mu_0 = 0$ and $\kappa = 1 \times 10^4$. Furthermore we can say that the variance on the riblength ought to be rather small because contours are long and narrow objects ($\sigma_0 = .01$; $\nu_0 = 20$). The remaining two parameters, $\lambda_1$ and $\lambda_2$, are the ones governing the actual shape of the contours. It are these parameters that indirectly reflect the Gestalt principles of proximity ($\lambda_1$) and good continuation ($\lambda_2$). These parameters were set to values that gave intuitive results for the examples below ($\lambda_1 = 0.16$; $\lambda_2 = 0.05$). Finally we set the parameter of the dirichlet prior to $\alpha = 0.1$. Naturally, alternative setting of these parameters may be appropriate for other contexts.

In what follows we will illustrate our framework on contours. We will first introduce dot-lattices in which only proximity is often said to be important, and show how they can be understood as being part of the same class as other contour integration problems. Afterwards we will show how the model can explain the formation of the classic association field, even though its definition is not explicitly included in the model. With these sanity checks in hand we than show how the BHG approach is able

to segment scenes of simple and complex edge configurations into intuitive contours, without any prior knowledge about what grouping hypotheses to test. Furthermore we will compare our frameworks contour grouping judgments to human subject data previously collected by Feldman (2001).

**Dot-lattices**

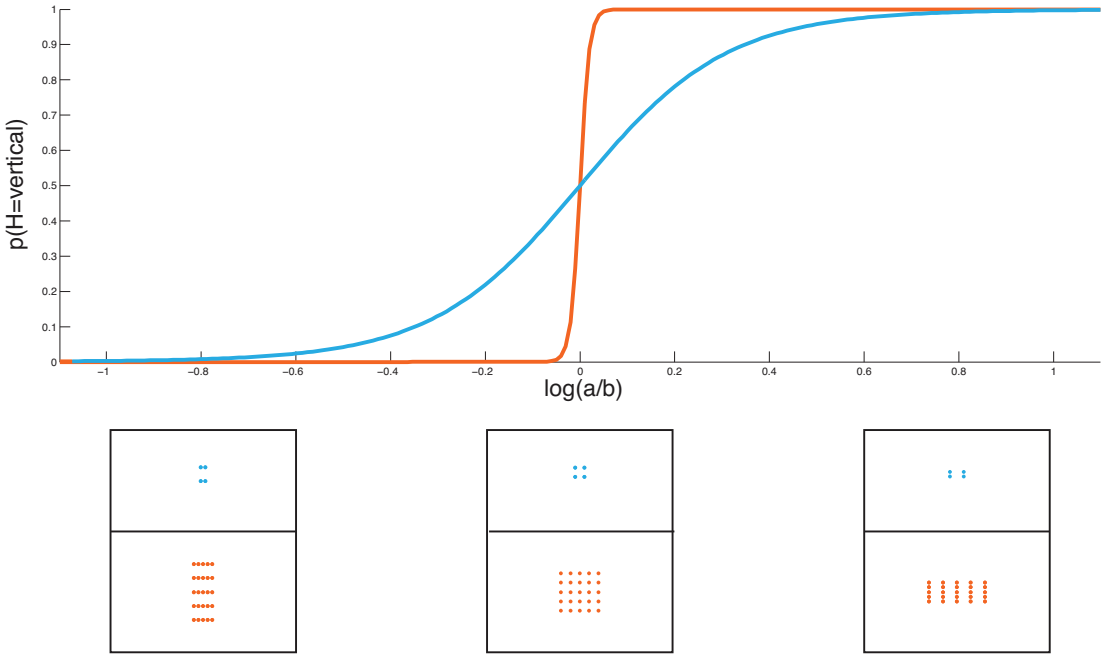


Figure 3.3: Framework predictions for simple dot lattices (Kubovy & Wagemans, 1995; Kubovy et al., 1998). As input the model received the location of the dots as seen in the bottom two rows, where the ratio of the vertical ($a$) over the horizontal ($b$) dot distance was manipulated. The graph on top shows how the probability of seeing vertical lines versus horizontal lines progressed as the ratio $a/b$ increased. The blue line shows this for small dot lattices of 2x2 elements, while the red line makes predictions for a larger 5x5 dot lattice.

Ever since Wertheimer (1923), researchers have used dot lattices to study grouping, and more specifically the Gestalt principle of proximity (e.g. Zucker, Stevens, & Sander, 1983; Kubovy & Wagemans, 1995). A dot lattice is a collection of dots arranged on a grid-like structure (e.g. Fig. 3.3), with for example vertical spacing $|a|$ and horizontal spacing $|b|$. These two lengths and the angle between them defines the dot lattice. For the example discussed here the angle was kept to $\pi/2$. Mathematical formulations

of how grouping is established within these lattices have been proposed before, of which the most notable is the pure-distance law (Kubovy & Wagemans, 1995; Kubovy et al., 1998). This formulation however only characterizes the local grouping strength as present between nearby dots. This works well within the context of equally spaced lattices because the computation of this grouping strength is the same for each dot. However, it neglects possible context influences from the overall lattice structure, such as the size of the lattice.

In order for our model to make predictions about lattice interpretations, we first need to realize that these simple structures are a simplification of the contour integration problem. Given this, we can set up the hypothesis space of possible grouping interpretations. In order to keep it simple for the example in Fig. 3.3, we only consider two hypotheses: rows (horizontal contours) versus columns (vertical contours). The posterior distribution over these hypotheses can easily be computed using Eq. 3.7. The results for a 2 by 2 and a 5 by 5 lattice can be seen in Fig. 3.3, where the larger the ratio of $|a|/|b|$ becomes the more probable the vertical contours hypothesis, consistent with empirical findings (Kubovy & Wagemans, 1995). Since the model does not base this posterior solely on the spacing between nearby dots, we can make more sophisticated predictions about the influence of the entire lattice structure. Specifically, we predict that the size of the grid matters. As can be seen by studying the psychometric curves in Fig. 3.3, the larger the grid, the higher we predict the sensitivity of the subjects to be. There are two reasons why our model makes this prediction. First of all with larger grids there are more dots per contour, essentially increasing the certainty of that particular contour interpretation. Secondly larger grids also have more contours increasing the evidence for a particular percept (horizontal versus vertical). This prediction, however, even though not yet tested directly, is in line with a similar finding in figure-ground perception. For classic figure-ground stimuli with alternating convex and concave regions Peterson and Salvagio (2008) showed that the more regions were present the stronger the bias became towards seeing the convex regions as figural.

**Collinearity**

Apart from the proximity cue, good continuation is another important Gestalt principle involved in contour integration. The visual system's tendency to group dots into approximately collinear segments has been studied in some detail (Smits, Vos, & Van Oeffelen, 1985; Smits & Vos, 1987; Feldman, 1997b, 2001). Here we will show how the model is also sensitive to collinearity of dot patterns by comparing its performance to human subject data. Feldman (2001) conducted an experiment in which subjects saw dot patterns consisting of six dots (e.g. Fig. 3.4A and B) and were asked to indicate if they saw one or two contours. To model the data from this experiment we simply ran our model on all the stimuli and computed the probability for the two alternative grouping hypotheses: ($\boldsymbol{c_0}$) all dots are generated by one underlying contour, or ($\boldsymbol{c_1}$) the dots were generated by two contours. The latter hypothesis encompasses several hypotheses, that is all 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 order of the dots. In other words we only considered 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 $\{(1,2,3,4,5),(6)\}$. Summing the probability for these hypotheses together yields the probability for $\boldsymbol{c_1}$. We then compare the posterior probability $p(\boldsymbol{c_1}|D)$ to the pooled subject responses for all 343 stimuli shown in the experiment. We found a monotonic relationship between the model and the subject responses. Because this relation was sigmoidal, we used an inverse cumulative Gaussian to linearize this relationship (Fig. 3.4). The framework responses and subject responses were highly correlated ($LRT = 375.06, df = 1, p < .001, R^2 = 0.6650$; $BF = 1.1764e+79$) [2].

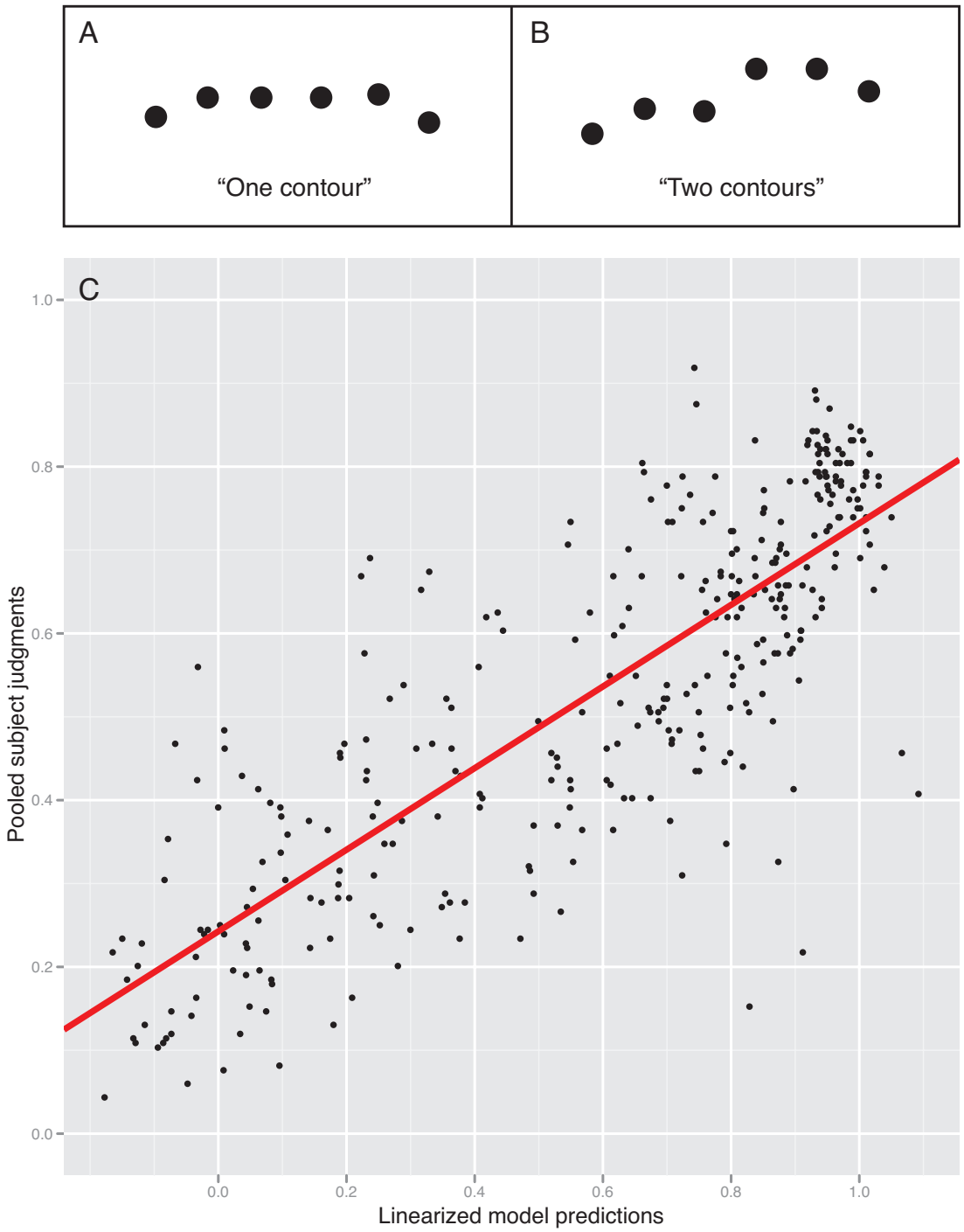


Figure 3.4: Model's performance on data from Feldman (2001). A/B. Sample stimuli with likely responses (stimuli not drawn to scale). C. Pooled subject responses plotted as a function of the model responses, where each point depicts one of the 343 stimuli shown in the experiment. Both indicate the probability of seeing two contours $p(\boldsymbol{c_1}|D)$. Note that the model responses are linearized using an inverse cumulative Gaussian.

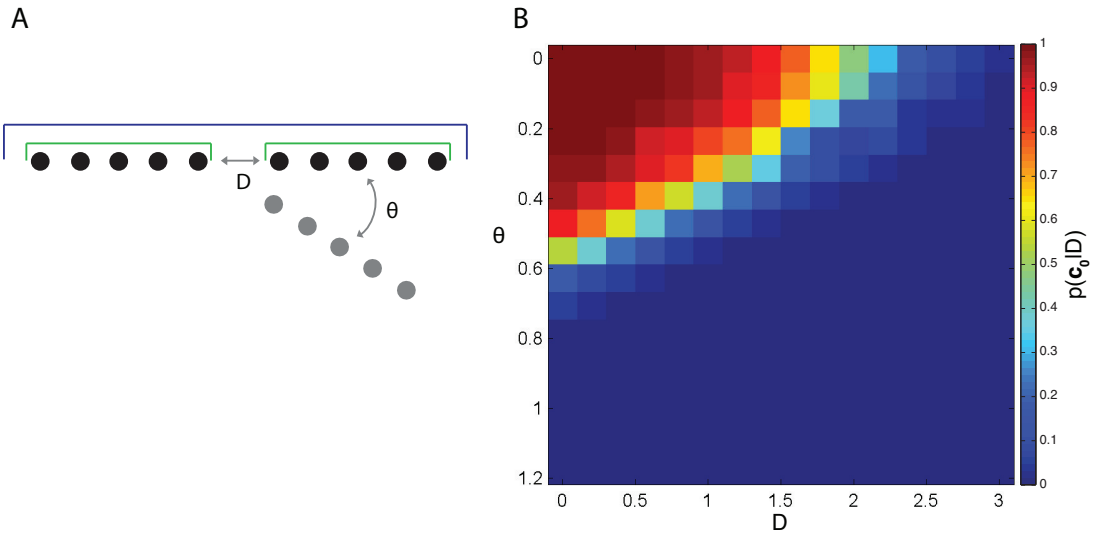


Figure 3.5: Association field between to line segments each containing 5 dots. A. shows our manipulation of the distance and angle between these two line segments. The blue line depicts the one object hypothesis while the two green lines depict the two objects hypothesis. B. depicts the association field for the posterior probability of $p(\boldsymbol{c_0}|D)$ when put into competition with $p(\boldsymbol{c_1}|D)$, where the gradient from blue to red depicts $p(\boldsymbol{c_0}|D)$

**Association field**

The interaction between the principles of good continuation and proximity has been well studied by Field et al. (1993), who propose an association field linking nearby elements for oriented edge segments. A closely related idea, co-circular support neighborhoods, was proposed earlier in the context of contour refinement in computer vision (Zucker, 1985; Parent & Zucker, 1989). Contrary to the classic association field our model does not concern itself with oriented edges. Nevertheless, similar interactions are present in case of dot patterns and an association field like pattern can be seen to emerge in our model (Fig. 3.5). Given two segments each consisting of five dots we manipulate the distance, $D$, and the angle, $\theta$, between the two segments (Fig. 3.5A). We then pit two grouping hypotheses against each other: ($\boldsymbol{c_0}$) all dots are generated by only single underlying contour, or ($\boldsymbol{c_1}$) both segments are generated by two separate

[2]Both the Bayes factor (*BF*) and the likelihood ratio (LRT) were computed by comparing a regression model in which the linearized framework responses were taken as a predictor versus an unconditional means model only containing an intercept.

contours. The posterior probability $p(\boldsymbol{c_0}|D)$ is shown in Fig. 3.5B. As one can see the bigger the angle between the two segments, and/or the further they are apart from each other the less probable the grouping hypothesis $\boldsymbol{c_0}$ becomes, where the gradient from blue to red depicts $p(\boldsymbol{c_0}|D)$.

**Contour integration with BHG**

In all three of the examples above we were able to set the grouping hypotheses beforehand. However in many cases we do not know what the alternative hypotheses are. It is for those cases that we will use the BHG to come up with a posterior distribution over possible grouping hypotheses. We first ran our framework on a set of simple edge configurations (Fig 3.6). One can see that the framework decomposes these into intuitive segments at each step in the hierarchy. Fig. 3.6A gives an illustration of how the MAP (maximum-a-posteriori) hypothesis is that all edges are generated by the same underlying contour, while the hypothesis one step down segments it into two intuitive segments. The latter hypothesis, however, is less likely under the current generative function and the assumptions of what we set to be a contour (i.e. the hyperparameter settings). Another example of an intuitive hierarchy can be seen in Fig. 3.6D, where the MAP estimate consists of three segments. The decomposition one level up (the 2 contour hypothesis) then groups the two segments that are abutting together. Again this hypothesis was found to be less likely given our assumptions. These simple cases show that the model can build up an intuitive grouping hypothesis space. To show that our frameowrk generalizes to more complex edge configuration we ran it on the stimuli shown in Fig. 3.7. The MAP grouping hypotheses are what one would expect in these cases. In Fig. 3.7B the longer segment was broken into two parts. Although this seems rather non-intuitive, it is what follows from the current features present in the model and the assumptions we set up above. Specifically, we included a penalty for the length of the curve in the form of $\lambda_1$. This penalty was given to the global configuration, i.e. the entire curve. Previous models, on the other hand have modelled the effect of the distances between two nearby dots (e.g. Geisler et al., 2001). As shown before (Feldman, 1997a), if follows from our simulation that in order to fully capture

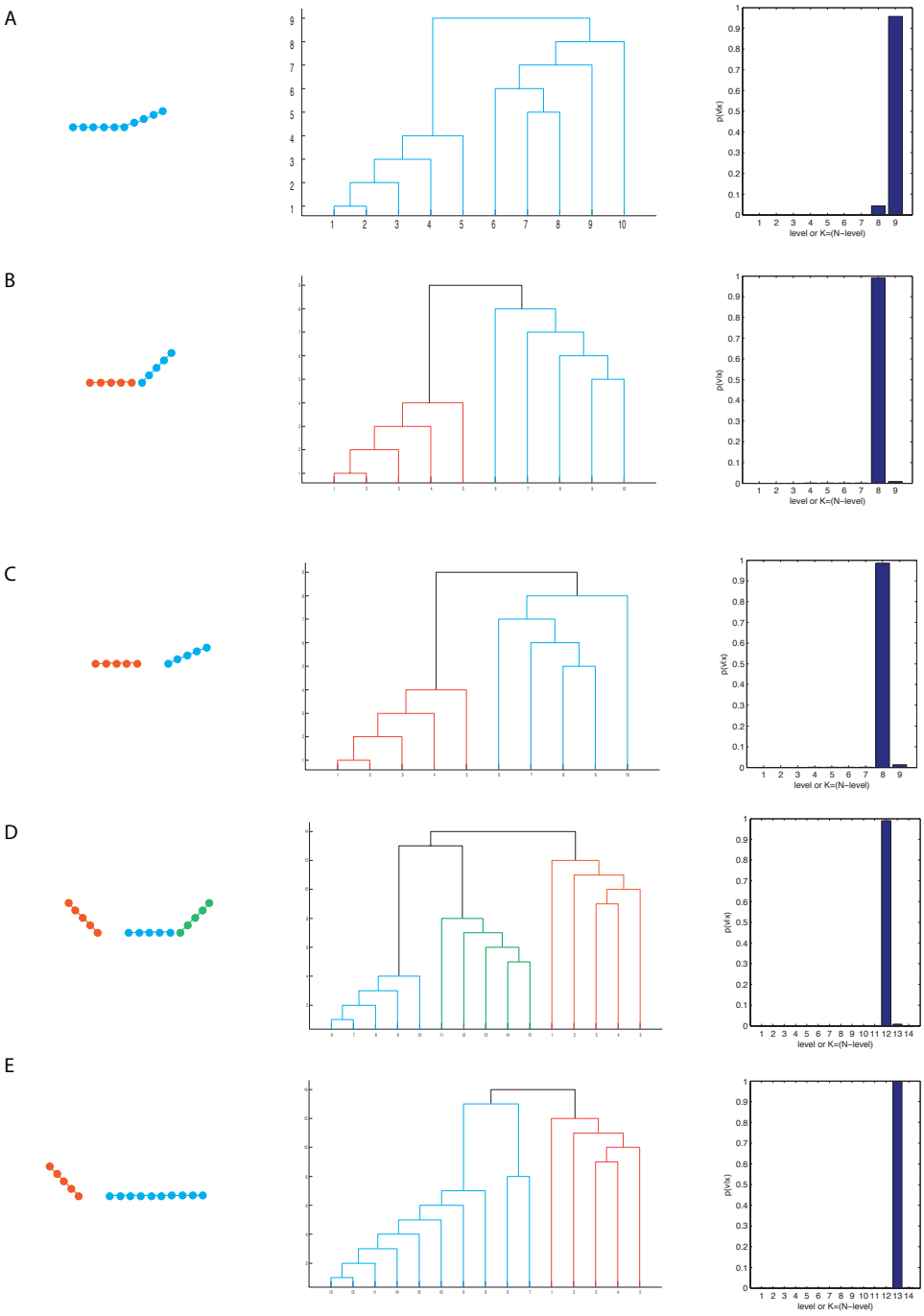


Figure 3.6: BHG results for simple dot contours. The first column shows the input images and their MAP segmentation. Here, the input tokens are numbered from left to right. The second column shows the tree decomposition as computed by the BHG algorithm. The third column depicts the posterior probability distribution over all tree-consistent decompositions (i.e. grouping hypotheses).

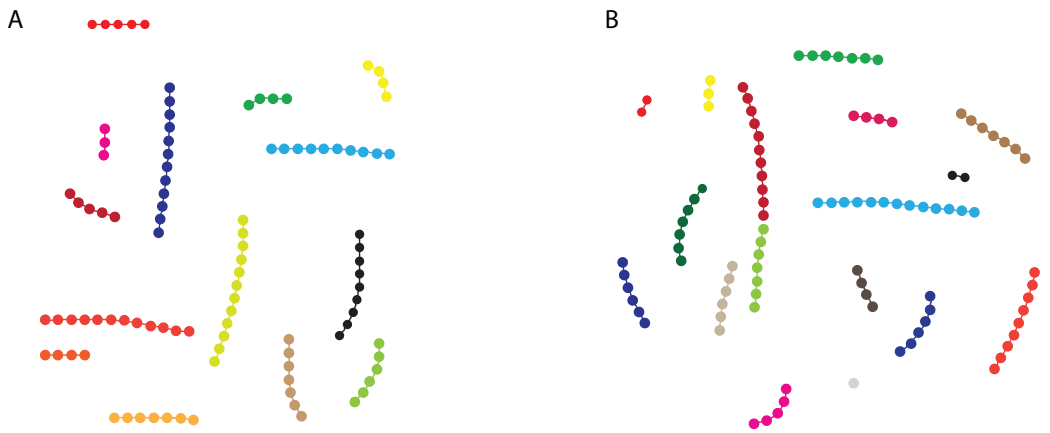


Figure 3.7: MAP grouping hypothesis for more complex dot configurations indicated by the color code (each group is assigned a unique color). B. Shows where the model has shortcomings, in that the length constraint prefers shorter segments. This results in longer contours to be split up. Introducing spacing as a constraint into the model might potentially solve this issue.

the complexity of contour integration the spacing of the dots along the contours has to be taken into account in some way.

### 3.4.2 Parts of objects

One way of representing shapes is by means of their parts. Many heuristics have been proposed for decomposing shapes into parts based on the geometry of the shape: the minima rule (Hoffman & Richards, 1984), the short-cut rule (Singh et al., 1999), and limbs and necks (Siddiqi & Kimia, 1995). Unfortunately these heuristics all have their exceptions and complex interactions with each other. Recently, Jiang, Dong, Ma, and Wang (2013), combined all these rules together to get around these limitations. Nevertheless one unifying theory of part-decomposition is lacking, and the mechanisms underlying it are not well understood. Like many before us (Blum, 1973; Singh & Feldman, 2008; Singh, Feldman, & Froyen, in preparation) we propose that skeletal computation could potentially be such a theory. Here we say that a skeleton is an ensemble of axes, each representing a part of the shape. The traditional approach to computing skeletons is the Medial Axis Transform (MAT) by Blum (1973). However,

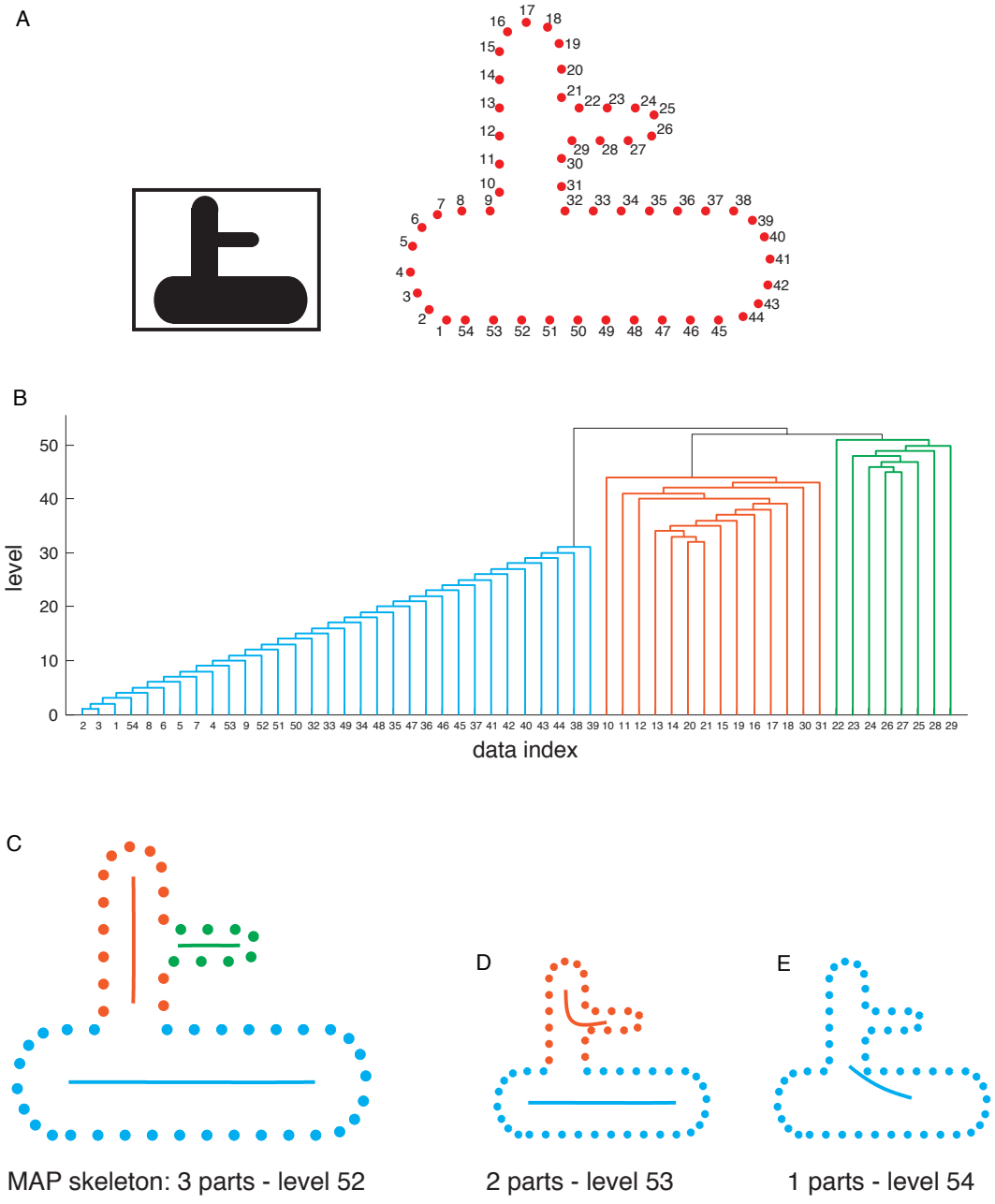


Figure 3.8: A. The top shape is decomposed by BHG. B. The tree structure that results from it is shown as a dendrogram. The MAP partitioning is given by the coloured parts in the dendrogram. This corresponds to the figure in C. Higher levels (D and E) show intuitive partitioning, depicting the hierarchical structure of the shape in A.

the computation of the MAT suffers from several problems such as spurious axes stemming from its sensitivity to boundary noise. Therefore a more probabilistic approach of skeletal computation is needed, similar to the MAP skeleton by Feldman and Singh (2006). Our framework supplies such a probabilistic theory by recasting the problem of part-decomposition as a grouping problem. Within our framework a shape is said to be generated from a mixtures of axes, where each axis represents a part and the ensemble of axes is called the skeleton of the shape.

In order for our model to analyse shapes we begin by preprocessing shapes in the following way. We assume that the edges of the shape are detected and that figure-ground is established, i.e. the shape needs to be explained from the inside. We create a discrete approximation of the shape $D = \{x_1 \dots x_N\}$ by subsampling the outline of the shape. Next, we need to set up the hyperparameters so they reflect our assumption about what a part looks like. The main difference with the hyperparameters for the contours is that we now do not want the mean riblength to be zero. That is in case of parts we assume that the mean riblength can be assigned freely with a slight bias towards shorter mean riblengths 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 that parts should preferably have smooth non-noisy 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 $\alpha = .001$.

**Shapes and their parts**

We ran the full BHG model on a simple multipart shape shown in Fig. 3.8A. The model finds the most probable part decomposition (Fig. 3.8B), and the entire structural hierarchy of this particular shape (Fig. 3.8C). In other words the BHG finds the entire description of the shape at different levels of the structural hierarchy, and does so intuitively (Fig. 3.8D and E). The MAP part decomposition for several shapes of increasing number of parts is shown in Fig. 3.10. Fig. 3.9C shows the algorithm's robustness to

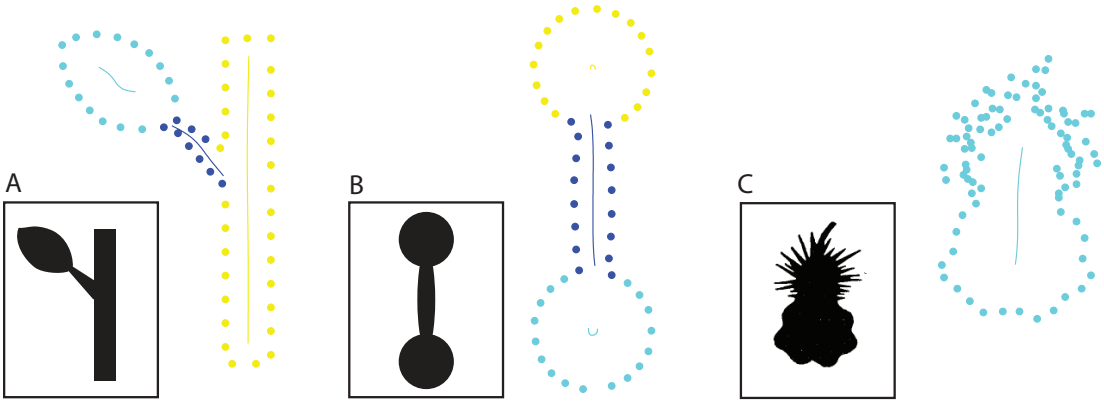


Figure 3.9: Examples of MAP tree-slices for: A. leaf on a branch, B. dumbbells, and C. "prickly pear" from Richards et al. (1986)

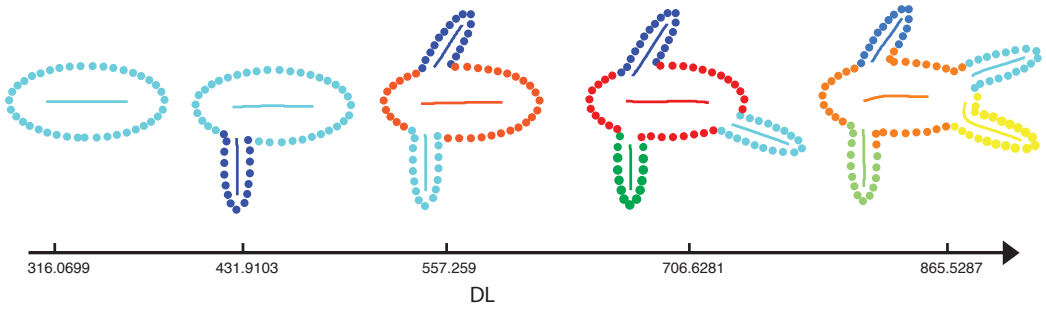


Figure 3.10: MAP skeleton as computed by the BHG for shapes of increasing complexity. The axis depicts the expected complexity, *DL* of each of the shapes based on the entire tree decomposition computed.

contour noise. This shape, the "prickly pear" taken from Richards et al. (1986) is especially interesting because different noise is added to different parts of the shape, which cannot be correctly handled by smoothing techniques. Note that each axis represents a part. As a consequence a particular grouping hypothesis generates a mixture of parts and not the shape per se (see Fig. 3.15). Hence, some additional post-processing in the form of smoothing and pruning of internal structures might be required to recover the shape.

As a side-effect of how we recast the problem of part-decomposition, ligature regions are not present, and the framework thus correctly handles difficult cases such as a leaf on a stem, and dumbbells (Fig. 3.9A-B), while still maintaining the hierarchical structure of the shapes (such as Siddiqi, Shokoufandeh, Dickinson, & Zucker, 1999). A ligature is often referred to as the "glue" that binds two axes together (e.g. connecting the leaf to its stem in Fig. 3.9A). Such regions have been identified before (August, Siddiqi, & Zucker, 1999b) to cause internal instability in the MAT (Blum, 1967), diminishing their usefulness for object recognition. In contrast to our approach past models had to cope with this problem by explicitly identifying and deleting such regions (e.g. August et al., 1999b).

Apart from the part decomposition we would like to say something about the complexity of the shape which is known to influence shape detectability (Wilder, 2013). In other words we would like to compute the description length (DL) of the shape. This value reflects the complexity of expressing the hypothesis in an optimal code (Rissanen, 1989). In order to compute the DL of the shape we first need to integrate over the entire grouping hypothesis space $\boldsymbol{C} = \{\mathbf{c_1} \dots \mathbf{c_J}\}$:

$$p(D|\alpha,\beta) = \frac{1}{J}\sum_{j=1}^{J} p(D|\beta,\boldsymbol{c}_j)p(\boldsymbol{c}_j|\alpha) \tag{3.10}$$

The DL is then defined as $\mathrm{DL} = -\log(p(D|\alpha,\beta))$. Fig. 3.10 shows how with increasing perceptual shape complexity this value also increases. This metric is universal to our framework and can be used to express the complexity of any image given any object definition, such as depicting the complexity of an image consisting of contours. In

other words the DL expresses the complexity (or description length) of a stimulus given our assumptions of what an object in it ought to look like.

**Part salience**

Hoffman and Singh (1997) proposed that the representation of a part is graded, and the visual saliency of a part is modulated by the sharpness of its boundaries and several geometric factors such as: its size relative to the entire object, and its degree of protrusion (defined as the ratio of its perimiter and the width at the base of the part). Within our framework we define saliency of a part by comparing two grouping hypotheses: the grouping hypothesis where the part was last present within the computed hierarchy $\boldsymbol{c_1}$, and the hypothesis one step up in the hierarchy where the part ceases to exist $\boldsymbol{c_0}$. In the examples that follow we defined part saliency as the log ratio between posterior probabilities of these hypotheses, indicating how much more likely it is for the part to be present versus absent. Naturally, depending on the task at hand part saliency can be operationalized differently by means of both hypotheses. Fig. 3.11 shows how our model captures part-saliency. Our model found that both for increasing part-length (Fig. 3.11A) and part-protrusion (Fig. 3.11B), the log posterior ratio increases monotonically.

To more systematically show how our model captures part saliency, we compare our model's performance to human subject data. Cohen and Singh (2007) showed empirically that several geometric factors contribute to part-saliency. Here we will focus on their experiment concerning part-protrusion. In this experiment subjects where shown a randomly generated shape, from one of 12 levels of part-protrusion (3[base widths]x4[part lengths]), after which they were shown a test part depicting a part of this shape (see Fig. 3.12A). They were then asked to indicate in which of four display quadrants this part was present in the shape. The authors found that subject's accuracy for this task increased with increasing part-protrusion of the test part. For each of the 12 levels of part-protrusion subjects were shown 50 randomly generated shapes. In order for us to compare our models performance to the subject accuracy we ran our model on 20 shapes for each level of part-protrusion. We then looked for the

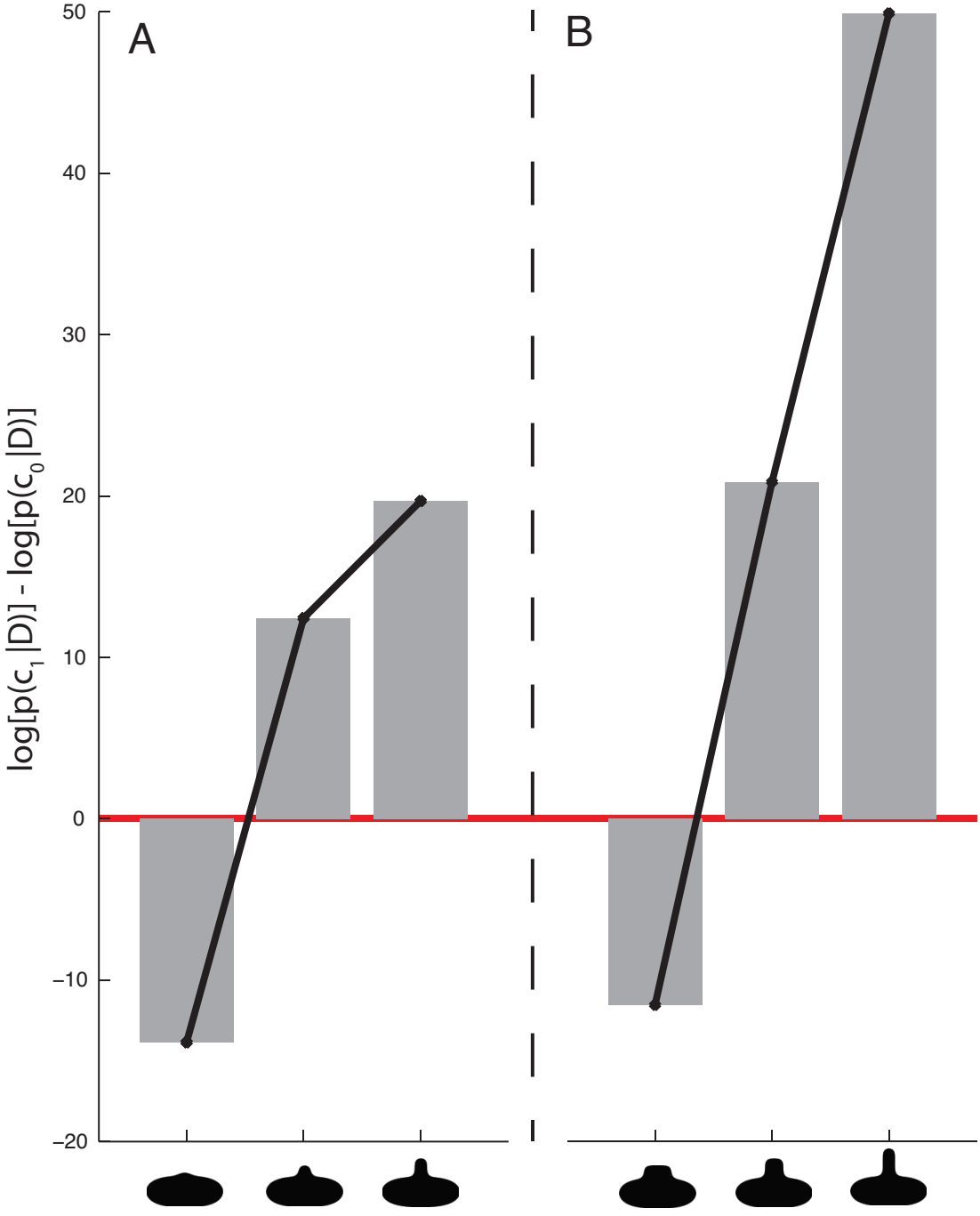


Figure 3.11: Log posterior ratio as computed from the BHG between the tree consistent 1 and 2 component hypotheses. A. Part protrusion, B. Part length.

presence of the test part in the hierarchy generated by our model. Given this we can then compute the posterior probability of this test part as follows:

$$p(\boldsymbol{c_1}|D) = \frac{p(D|\beta, \boldsymbol{c_1})p(\boldsymbol{c_1}|\alpha)}{p(D|\beta, \boldsymbol{c_0})p(\boldsymbol{c_0}|\alpha) + p(D|\beta, \boldsymbol{c_1})p(\boldsymbol{c_1}|\alpha)} \tag{3.11}$$

Fig. 3.12 shows a monotonically increasing relation between the subject's accuracy and the probability of the test part as computed by our model. Because this relationship was sigmoidal we used an inverse cumulative gaussian to linearize it. Our model's computed probability of the test part was found to be a good predictor of subject's accuracy( $LRT = 74.75$, $df = 1$, $R^2 = 0.4050$; $BF = 4.2376e+14$)[3].

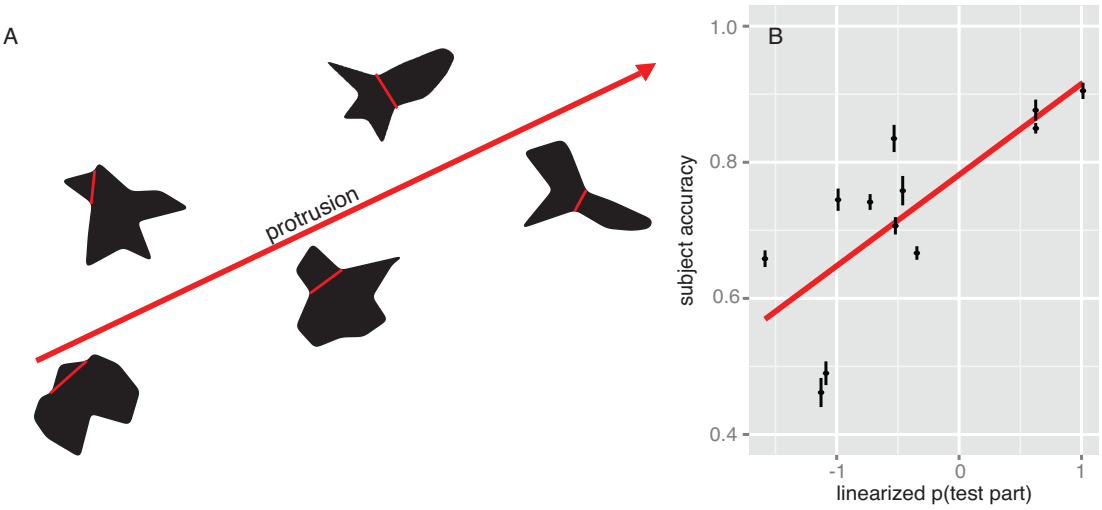


Figure 3.12: A. Representative stimuli used in Cohen and Singh (2007) experiment relating part-protrusion to part saliency. As part protrusion increases, so does subjects perceived saliency of that part. The test part here is indicated by the red part cut. B. Representing the relationship between subject accuracy for several levels of part-protrusion and the models computed probability of the test part $p(\boldsymbol{c_1}|D)$ (error bars depict the 95% confidence interval across subjects. The red curve depicts the linear regression.

[3]Both the Bayes factor (*BF*) and the likelihood ratio (LRT) were computed by comparing a regression model in which the linearized framework responses were taken as a predictor versus an unconditional means model only containing an intercept.

### 3.4.3 Shape completion

Completion refers to the integration of contour elements that are separated by gaps, caused by occlusion. For contour completion the visual system needs to solve two problems (Takeichi, Nakazawa, Murakami, & Shimojo, 1995). First it needs to derter-mine if these contour elements need to be grouped together (*grouping problem*), and subsequently what the shape of contour is inside this gap (*shape problem*). Here we will focus on the latter problem and show how our model can make specific predic-tions about the shape of the contour. Nevertheless as we shall touch on it briefly, the framework also contains properties that make it possible to address the grouping problem.

#### Global predictions

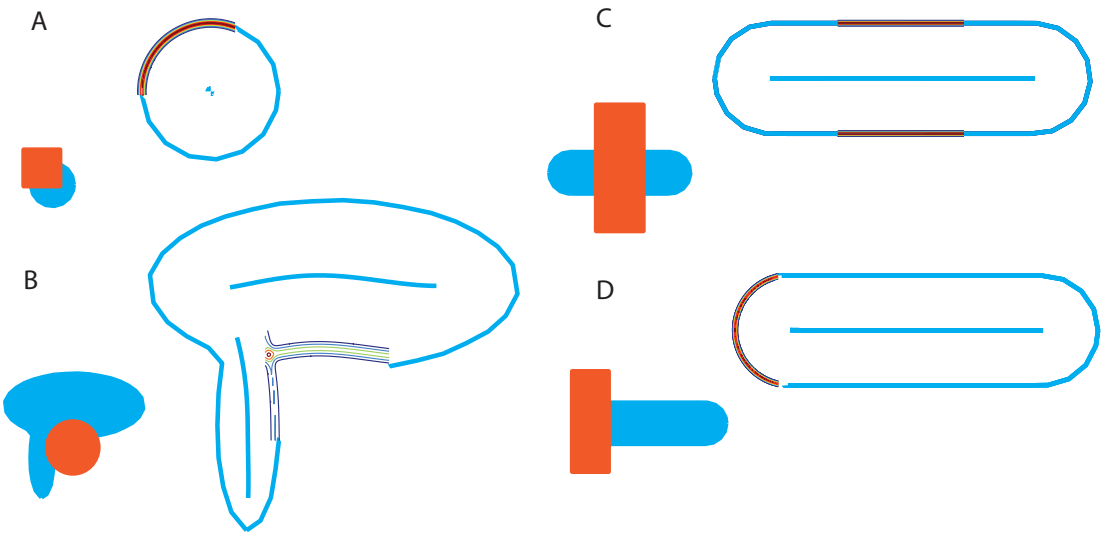


Figure 3.13: Posterior predictive based on the MAP skeleton (as computed by BHG) for the occluded shape with a part of the boundary missing.

Most past models that make predictions about the shape of the occluded por-tion of the contour have based their predictions solely on local contour information (Williams & Jacobs, 1997; Fantoni & Gerbino, 2003; Ben-Yosef & Ben-Shahar, 2012). That is, they only used the position and the orientation of the contour at the point where it disappears behind the occluder (called the inducers). Such models, however, can not

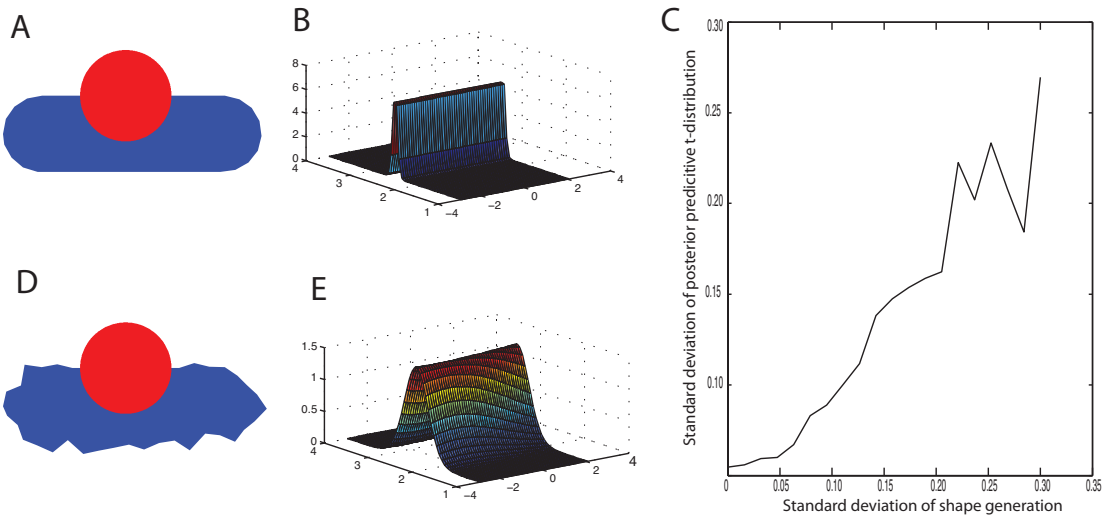


Figure 3.14: A simple tubular shape was generated with different standard deviations of noise on its contour. Note that for each image (A and D), the local first and second order information at the T-junction is kept equal. For noiseless contours the posterior predictive for the occluded part is rather narrow (A and B), while for noisy contours the posterior predictive takes on a wider form (E), depicting the uncertainty of the position of the boundary based on the shape alone. C. Shows the relationship between the noise on the contour and the completion uncertainty as reflected by the posterior predictive.

explain the non-local influences on shape completion found in some studies (Fulvio & Singh, 2006; Sekuler, Palmer, & Flynn, 1994; Van Lier, Van der Helm, & Leeuwenberg, 1995). Our model on the other hand, much in the same spirit as August, Siddiqi, and Zucker (1999a), can make prediction based on the entire shape of which a part is missing due to occlusion. By first computing the hierarchical representation of the shape (given the object definitions setup for part-decomposition) with the missing boundary segment, we then compute the posterior predictive (Eq. 3.8) based on the best grouping hypothesis, i.e. MAP tree-slice. The predictions our model makes however are probabilistic. Therefore when plotting these predictions we choose not to plot one particular completion interpretation, but rather a field within which a infinite number of possible contours lie. Note that the predictions made here are only based on the global shape and do not take into account the location nor orientation of the local inducers. That is, when inferring a particular contour between those inducers, the prediction field here should be understood as a global shape prior and should be combined with a

good-continuation constraint on that contour. For example for a simple case in which a circle is occluded by a square the model makes a prediction that a contour must lie along a circular path with the same arc as the rest of the circle (Fig. 3.13A). When more parts are present, the model can still make predictions (Fig. 3.13B). Furthermore, even in cases where there is not even enough information present for local models to create a boundary, our model can readily make a prediction (Fig. 3.13D). Finally the model can also handle cases in which the grouping problem also needs to be solved (Fig. 3.13C). This essentially follows directly from our framework. The grouping problem in our case reduces to a problem of grouping axial fragments on either side of an occluder. The rules for grouping them are present in the hyperparameters defining our object assumptions. Specifically, as was the case for contour integration $\lambda_1$ and $\lambda_2$ indirectly encode proximity and good continuation, in this case of an axis, and $\alpha$ encodes our prior belief of grouping two parts together into one objects.

**Dissociating global and local predictions**

Many of the predictions made by our model in the above paragraph could also easily be accounted for by a model using local fragments only. However, often global and local predictions will give vastly different shape completions. In the past researchers have found cases in which the shape of the completed shape influences our perceived completion. For example Sekuler et al. (1994) found that the presence of certain symmetries in the completed shape facilitates that specific shape completion. More general Van Lier et al. (1995, 1994) found the regularity of the completed shape as formulated in their regularity-based framework to influence perceived shape completion. The shapes in Fig. 3.14, 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 (Fig. 3.14A and D). It is clear that a model based merely on local inducers would predict the contour to look exactly the same in both cases. In other words the complexity of the shape's contour does not add to the uncertainty of the estimation of the shape completion. On the other hand in our framework the global shape is taken into account, resulting in

the uncertainty of the shape completion to go up with the complexity of the contour (Fig. 3.14B, E, and C). This prediction is interesting in that it could potentially dissociate which of the two kinds of information are more important, or more particularly how both types of information are combined to form our percept.

## 3.5 Discussion

In the current chapter we put forward a mathematically rigorous and principled framework for understanding perceptual grouping. We tested an instantiation of this framework for several key problems in perceptual grouping: contour integration, part decomposition and shape completion. The framework has several properties that makes it stand out when compared to other models. First of all we will discuss how the framework is able to deal with perceptual grouping in general by means of defining the appropriate objects. Secondly we will discuss the hierarchical nature of the model and how it relates to structural scale and the notion of selective organization.

### 3.5.1 A framework for grouping

The framework presented in the current chapter is one for understanding perceptual grouping. Within this framework we follow ideas we have proposed in other papers (Feldman et al., submitted, and Chapter 2), and recast the problem of perceptual grouping as a mixture estimation problem, such that an image is said to be a mixture of objects. redIn our model we define objects based on what organization decomposes the image into distinct components. The question then remains what kind of underlying mechanisms structure otherwise complex and heterogeneous image data into an organization consisting of coherent objects (Feldman, 2003b). No one unified answer arises from the perceptual grouping literature. Many different grouping cues (e.g. proximity, closure, convexity, ...) have been presented resulting in several different object classes (e.g. contours, shapes, ...). Several of these cues are often needed to define certain classes of objects, raising the need for one all overarching organizing principle. One such famous principle is *Pragnanz* (Wertheimer, 1923), also referred to as "goodness of

form". redAs some researchers before us (e.g. Ommer & Buhmann, 2003; Song & Hall, 2008) the current framework proposes a formal definition of this idea. The framework itself gives us the machinery to understand how grouping is established independent of what objects we are creating, establishing a universal (Bayesian) "language" for grouping principles to "speak" to each other. Within our framework we compute the posterior probability of several grouping interpretations, defining the "goodness" of a certain decomposition of the image into object-like components.

The flexibility of our framework lies in the object definition. The definition of objects, and indirectly the grouping cues underlying them, consists of two components. A first component is the object class, which is defined by the generative (likelihood) function, $p(D|\theta)$. This component defines how image elements were generated given the underlying object. In our instantation of the framework we proposed an object class for spatial grouping problems. The class proposed is a rather general one in that the definition makes a statement about the fact that image elements need to be generated perpendicular at a certain distance from the underlying objects (curved) shape governed by a Gaussian fall-off centered around some mean distance. As discussed above this object class can be made to generate contour fragments, dot clouds, and edges of a shape. A second component of our object definition are the priors on the parameters governing this generative function, $p(\theta|\beta)$. These priors define our assumption about what we think objects within this class look like, and are what makes an object class able generate several different objects. Specifically, it unifies several formerly distinct object classes (such as contours, dot clusters, and shapes) under a common object class. That is, within our proposed object class contours are merely elongated shape, and dot clusters are shapes with image elements present in their interior. Together both components define all objects in our framework. How both components are defined can be task dependent, i.e. depending on the object(ive) of the task. redOne can imagine different object definitions, such as the Gaussian objects used in Chapter 2, or other spatial features that can be introduced such as dot-spacing as proposed in the contour integration section. Furthermore, objects can be defined that go beyond the spatial domain, including features such as color, texture, and contrast. As long as one

can define both components, the framework's machinery can be put to work to group image elements based on these definitions.

### 3.5.2 A hierarchical framework

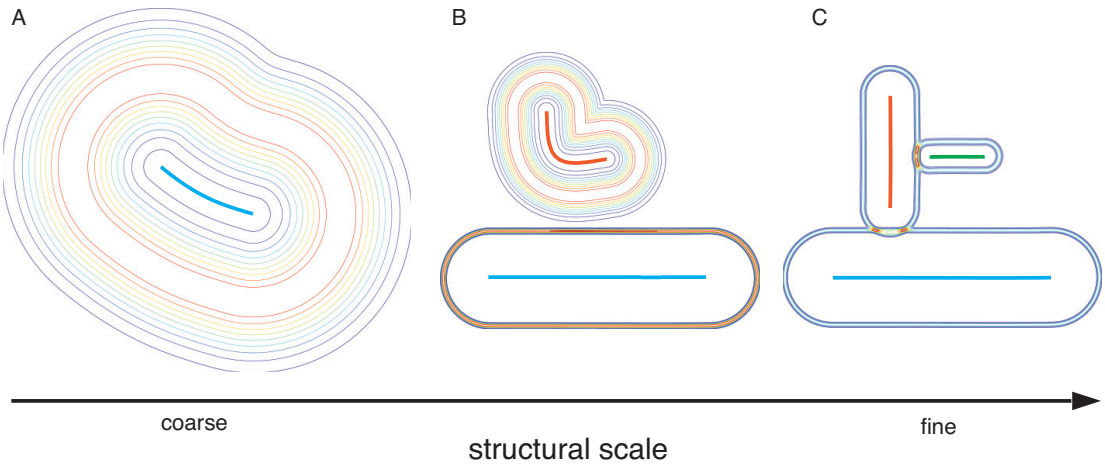


Figure 3.15: Prediction fields for the shape in Fig. 3.8 for three different levels of the hierarchy. In order to illustrate how underlying objects also represent the statistical information about the image elements they explain the prediction/completion field was computed for each object separately without normalization so that the highest point for each object is equalized.

Our framework constructs a hierarchical representation of the image elements based on the object definitions and assumption set. Hierarchical approaches to perceptual grouping have been plentiful (e.g. Pomerantz et al., 1977; Palmer, 1977; Baylis & Driver, 1993; Lee & Mumford, 2003). 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, about the shape (Fig. 3.15A). An axis in our account, or object in general, can be seen as the more classical *shape primitive*. However, our objects are highly flexible with not only "memory" about the shape of the primitive (i.e. object), but also the statistical information about the image elements it explains (Fig. 3.15). That is given the object we can generate the image elements and make

predictions about missing parts of the image (Fig. 3.13). The further down the hierarchy we go, the more detail of the image will become visible, that is more and more objects will be come visible. Furthermore note that the prediction fields become more and more narrow depicting that the objects are *fitting* more and more details about the shape (Fig. 3.15).

**Structural versus spatial scale**

In the literature two different types of scales can be distinguished. On the one hand *structural scale* describes the structural organization of the image based on perceptual grouping rules (Palmer, 1977; Feldman, 2003b). In essence it describes a tree structure of how the image can be parsed from the root node representing image as one object to its leave nodes where each image element is represented by itself. How this tree is constructed then depends on the grouping rules that were instated. On the other hand, more popularly, images are analyzed using different *spatial scales*. Spatial scales are often defined as a hierarchy of receptive fields of increasing size, taking in more and more global image information while climbing up the hierarchy. This type of scale has been used to explain several problems in perceptual grouping such as figure-ground (e.g. Jehee, Lamme, & Roelfsema, 2007), and shape representation (e.g. Burbeck & Pizer, 1994)

redIn the past these different notions of scale have been equated with each other because grouping more image elements together into one object is sometimes regarded as to analyzing the image at a larger spatial scale. However, the two notions of scale approach the problem of perceptual grouping from rather orthogonal points of view. Specifically, when moving through different spatial scales we in essense change the way we look at the image elements, i.e. we change the way we analyze them. This on the other hand is not true when we build a structural hierarchy. In that case we look at the image in exactly the same way, i.e. keeping the spatial scale fixed, and apply the same grouping principles as we climb up the hierarchy. This distinction becomes more apparent in our framework than in any of the hierarchical approaches proposed before.

In our framework spatial scale can be incorporated into the object definitions. More specifically in for our current object class we can manipulate spatial scale by changing the prior of the variance over the riblength, $\sigma$. Making this prior's mode shift to larger values of $\sigma$ ensures that objects will be more tolerant to noise in the image elements. Fig. 3.16B shows three different priors depicting three different spatial scales. Structural scale then is defined by the hierarchical structure built up by our framework. Fig. 3.16C show how for each different spatial scale the hierarchical structure built up changes considerably. For the finest spatial scale used here, we find a structural hierarchy that builds up including the three to one intuitive object hypotheses, with the three part hypothesis being the most probable (Fig. 3.16A). On the other hand at coarser spatial scales the parts found in the structural hierarchy are different, and the three parts as found in the finer spatial scale are not even present anymore. Furthermore the most probable hypothesis now is the two part hypothesis. This example clearly shows that structural and spatial scale are related to each other, and describe different (and orthogonal) components of the mechanism underlying perceptual grouping. Spatial scale refers to how the image is analyzed, or our assumptions about the objects, while structural scale refers to the structural hierarchy build up given a particular spatial scale.

**Selective organization**

The way our framework builds hierarchical representations of the image and considers grouping hypotheses is also consistent with the notion of selective organization (Palmer, 1977). Selective organization refers to the fact that some subsets, or objects, are represented while other are not. In our model only $N$ grouping hypotheses are considered while the total amount of possible grouping hypotheses $\boldsymbol{c}$, is exponential in $N$. A grouping hypotheses picked at a particular level is directly dependend on the grouping hypotheses choosen at lower levels. In other words all grouping hypotheses are tree-consistent hypotheses. In this way a clean and unambiguous hierarchical structure is built up by our model. Furthermore this results in some grouping hypotheses not being represented by the hierarchy. Such selective organization has been

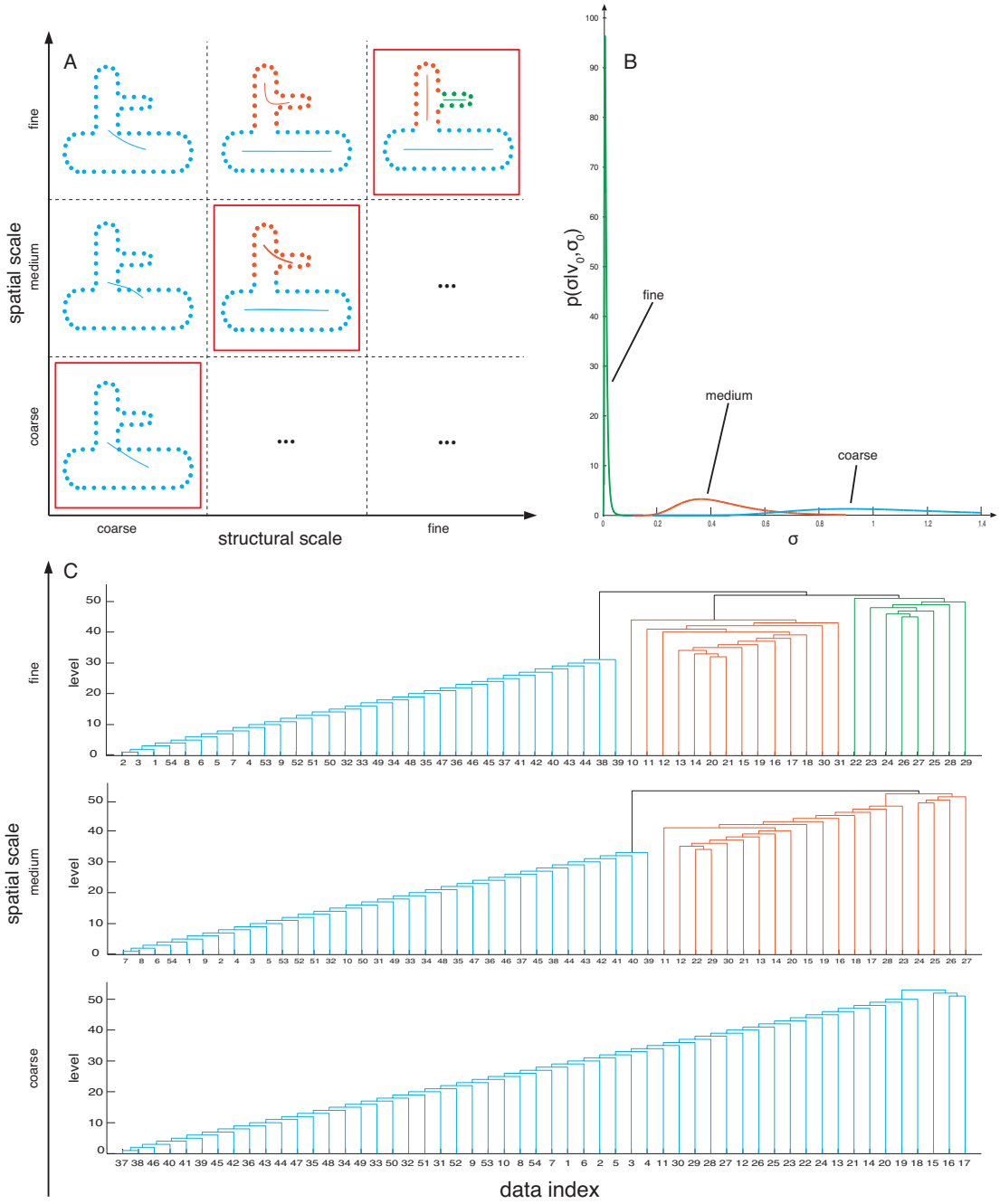


Figure 3.16: Relating structural and spatial scale in our model by means of the shape in Fig 3.8. A. relationship between structural and spatial scale depicting their orthogonality. The red squares depict the most probable structural grouping hypothesis for each spatial scale. B. Showing the priors over the variance of the riblength, $\sigma$ 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.

given some empirical support in the past (Palmer, 1977; Cohen & Singh, 2007), where objects that were not represented in the hierarchy were harder to retrieve.

### 3.5.3 A Bayesian framework

The Bayesian approach used here has several substantial advantages over traditional approaches toward perceptual grouping. First of all a Bayesian approach allows us to assign different degrees of belief (or probabilities) to different grouping hypotheses, in effect capturing the often intermediate responses present in subject data. Specifically, previous non-probabilistic models often only converge on one particular grouping hypothesis (e.g. Williams & Jacobs, 1997), or are unable to assign beliefs to different grouping hypotheses (e.g. Compton & Logan, 1993). Secondly, the rationale behind Bayesian inference includes it making optimal use of the available information and the assumption held by the observer (Jaynes, 2003). In other words the framework makes optimal use of the image elements present and assumptions as defined by the object priors to construct the posterior distribution over the grouping hypotheses. Note that we did not make any claim about how decisions are drawn from this posterior distribution. However one can easily add a loss function on top of this distribution to come up with a Bayesian optimal decision (Maloney, Mamassian, et al., 2009).. On the other hand subjects might also be sampling from this posterior rather than selecting their response based on minimizing their loss. Even though such a decision strategy, also called *probability matching*, is sometimes regarded as suboptimal it appears to be used in some perception tasks (Wozny et al., 2010). Although our framework does not commit to either, it is can easily be made consistent.

## 3.6 Conclusion

In this chapter we presented a novel mathematically rigorous and coherent framework for understanding perceptual grouping. Within our framework we defined an image as a mixture of objects, and thus reformulated the problem of perceptual grouping as a mixture estimation problem. Our framework's generality stems from the freedom

that is given to the object definitions. In its current instantiation we defined an object class that unites problems such as part-decomposition, contour integration, dot clustering and shape completion. Other object classes can be defined depending on the task at hand and easily plugged into our framework. Apart from its generality our framework stands out in that it generates a hierarchical representation of the image. We tested our framework's workings for several key perceptual phenomena in the fields of part-decomposition, contour integration and shape completion. Furthermore we showed that the framework accounts for human subject data from previously conducted experiments in contour integration and part decomposition.

# 4. Conclusions

Dividing a set of visual elements into distinct coherent objects is a basic problem of perceptual grouping. The problem of perceptual grouping is inherently difficult because the visual system has to find the optimal grouping interpretation among a large set of possible interpretations. Many models have been proposed for several subproblems of perceptual grouping such as contour integration, figure-ground, and many others. Furthermore other problems in visual perception such as part-decomposition can as I showed be understood as a grouping problem. However, these models often describe the mechanisms underlying perceptual grouping by means of poorly understood and unprincipled Gestalt principles or heuristics. Though these models can make valid predictions for the subproblem of their focus, they often fail to assign degrees of belief to these grouping interpretation. Hence, (1) they fail to capture the often probabilistic nature of human behavior and (2) they fail to generalize beyond the subproblem they were tailored towards.

In this dissertation I proposed a mathematically rigorous and coherent framework for understanding perceptual grouping. Within this framework I formulate the problem of perceptual grouping as a mixture estimation problem, where it is assumed that the image elements are generated by a set of distinct objects. Intrinsic to the problem of mixture models is the simultaneous estimation of the parameters of the objects ("what do the objects look like?") and assigning each element to an object ("ownership"). In this dissertation I proposed two different operationalizations of this central idea. In the second chapter I implemented a simplified framework instantiated for the rather simple problem of dot clustering. Here the objects were assumed to be Gaussian dot clusters. The third chapter expanded on this to create a more elaborate framework estimating a hierarchical representation of the image data. The advantage of this approach to the simplified approach in Chapter 2 is that the alternative grouping hypotheses are not set by hand, but rather are generated by the framework itself. Furthermore we proposed a general object class which enabled the current instantiation of the framework to handle problems such as contour integration, part decomposition,

shape completion and dot clustering.

We showed that each of the framework's operationalizations quantitatively predicted subjects behavior and was able to account for several perceptual phenomena. In Chapter 2 I conducted two experiment in which I showed subjects dots generated from two Gaussians (experiment 1) or three Gaussians (experiment 2). In both experiments I manipulated the distances between the clusters in order to modulate the apparent number of clusters. Subjects were then asked to indicate how many clusters were present. I found the framework in Chapter 2 to give accurate and quantitative precise account of subjects' numerosity judgments. In Chapter 3 I found the expanded framework to account for several perceptual phenomena specific to contour integration, part decomposition and shape completion. Furthermore the framework was found again to give an accurate and quantitative precise account for subjects' contour integration and part decomposition behavior recorded in previously conducted experiments.

The framework proposed here can be instantiated for several different grouping problems given the appropriate object definitions. All this generality is based on only one central idea: *an image is a mixture of objects*. I hope that this generality will sprout further application of the framework across different subproblems of perceptual grouping. Furthermore I hope that the strong predictive power of the framework will sprout many new experiments and insights about the mechanisms underlying perceptual grouping.

# 5. Appendix

## 5.1 Mixture Estimation

In this section we outline how we computed $p(X|y_i)$, where $y_i$ is related to a mixture with $K_i$ components. First of all let us reformulate Eq. 2.1 as follows:

$$p(x_n|\phi, y_i) = \sum_{k=1}^{K} \pi_k p(x_n|\theta_k), \tag{5.1}$$

where $\phi = \{\pi_1, ..., \pi_K, \theta_1, ..., \theta_K\}$. We assumed the image elements were sampled from a bivariate Gaussian, $\theta_k = \{\mu_k, \Sigma_k\}$. To evaluate the data under the above mixture model we need to specify some prior over the parameters of the model, $p(\phi, y_i)$. We now have the ingredients to compute the probability of the data $X$ under the grouping hypothesis $y_i$:

$$p(X|y_i) = \int_{\phi} p(X|\phi, y_i) p(\phi|y_i) d\phi. \tag{5.2}$$

This computes gives the probability that the data was generated from a mixture model with $K_i$ clusters. Because the computation of this marginal likelihood for each trial would take considerable computation combined with the large number of trials we ran in these experiments, we opted to approximate it as follows. We first estimated the parameter combination $\hat{\phi}$ that best explained the data $X$ by means of the Expectation-Maximization procedure Dempster, Laird, and Rubin, 1977. These parameter estimates then yield us, $\hat{L}(y_i|X)$, the likelihood of the data given these parameters. This in turn can be used to approximate the actual marginal likelihood $p(X|y_i)$ by means of the Schwartz criterion:

$$log(p(X|y_i)) \approx log(\hat{L}(y_i|X)) - \frac{p_i}{2} log(N), \tag{5.3}$$

where $p_i$ is the number of parameters estimated for the mixture model $y_i$ and $N$ is the number of data-points. A similar approach was used by Pelleg and Moore (2000) to compute the relative strength of different clustering interpretations as estimated by the

K-means algorithm. The number of parameters $p_i$ is simply the sum of $K-1$ cluster probabilities ($\pi$), and $K \times P$. $P$ is the number of parameters estimated for each cluster, depending on the flavor of the model we tested. The elliptical version consisted of a two-dimensional mean, and a full covariance matrix, amounting to a total of $P = 5$ parameters. The circular version also had a two-dimensional mean, but only a circular covariance matrix (i.e. only one variance estimate), amounting to a total of $P = 3$ parameters.

## 5.2 Prior on Cluster Shape

Note that in the estimation procedure depicted in Appendix 5.1 we have ignored the prior on the parameters $p(\phi|\beta)$ introduced above. In our current approach we essentially assumed an uninformative prior on all the parameters. One might think of the two model versions proposed in the paper as two different priors over the covariance matrix $\Sigma_k$. More specifically both versions make different assumptions about the shape of the clusters. We can easily show how such assumptions can be incorporated as priors on $\Sigma_k$ estimates by rewriting the covariance matrix as,

$$\Sigma_k = \begin{bmatrix} \sigma_{x,k} & \sigma_{xy,k} \\ \sigma_{xy,k} & \sigma_{y,k} \end{bmatrix} \tag{5.4}$$

can also be written as,

$$\begin{aligned} \sigma_{x,k} &= \frac{\cos^2 \alpha_k}{2M_k} + \frac{M_k \sin^2 \alpha_k}{2R_k} \\ \sigma_{y,k} &= \frac{\sin^2 \alpha_k}{2M_k} + \frac{M_k \cos^2 \alpha_k}{2R_k} \\ \sigma_{xy,k} &= \frac{\sin 2\alpha_k}{4M_k} + \frac{M_k \sin 2\alpha_k}{4R_k}. \end{aligned} \tag{5.5}$$

In this set of equations $\alpha_k$ defines the orientation of the cluster relative to the x-axis. The size of the cluster is defined by $M_k$, which essentially is the variance $\sigma_x$ the cluster would have if it had orientation $\alpha_k = 0$. Finally the shape of the cluster is defined by $R_k$ as the ratio $\sigma_x/\sigma_y$ the cluster would have if it had orientation $\alpha_k = 0$. In this way one can see that assumption of the shape could be defined as a prior over $R_k$. In the current paper we only tested two shape assumptions. The elliptical version

essentially assumes $R_k \sim unif()$ as its prior, while the circular variant assumes a dirac delta $R_k \sim \delta(1)$.

## 5.3 Delaunay-consistent pairs

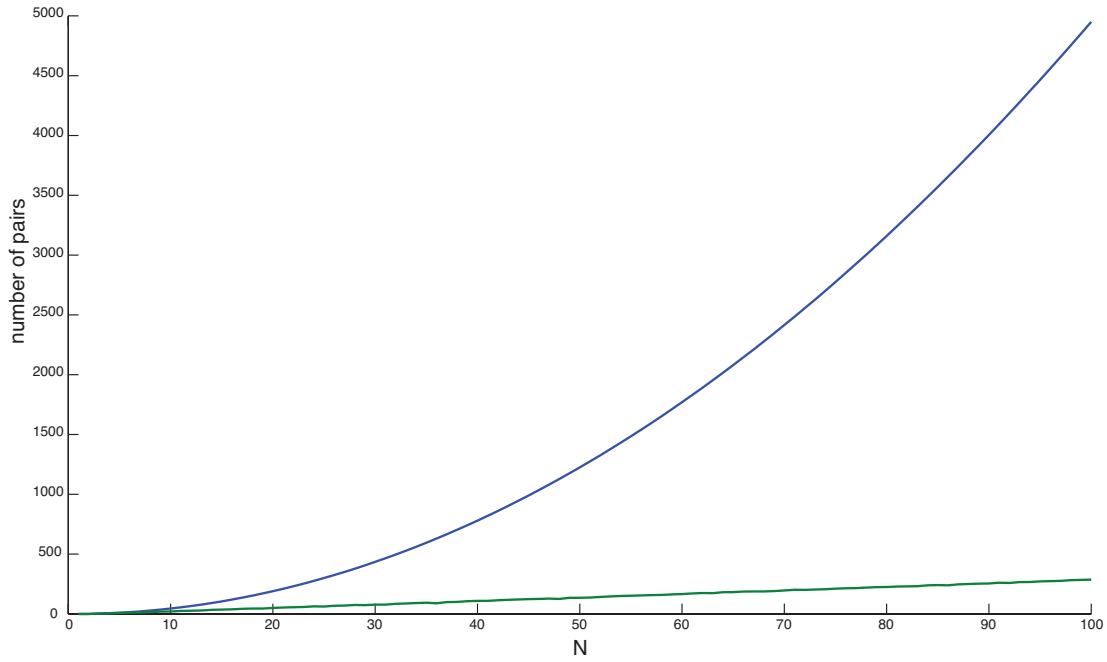


Figure 5.1: Difference between checking all pairs and only Delaunay-consistent pairs at the first initial iteration of the BHG. As the amount of data points, $N$, increases the number of pairs increases differently for the Delaunay-consistent (green), or all pairs (blue).

Bayesian Hierarchical Clustering is a pairwise clustering method, where at each iteration merges between all possible pairs of trees $T_i$ and $T_j$ are considered. Given a dataset $D = \{x_1 \ldots x_N\}$ the algorithm is initiated with $N$ trees $T_i$ each containing one data point $D_i = \{x_n\}$. redAs $N$ increases the number of pairs to be checked during this first iteration increases quadratricaly with N, or more specifically as follows from combinatorics $\#pairs = (N^2 - N)/2$ (Fig. 5.1), resulting in a complexity of $O(N^2)$. In each of the following iterations the hypothesis for merging only needs to be computed for pairs between existing trees and newly merged trees 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 to limit the pairs checked to a local neighbourhood 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 neighbour of that point. To initialize the BHC algorithm we compute the Delaunay Triangulation over the dataset $D$. Given this we can then compute a binary neighbourhood vector $\boldsymbol{b}_n$ of length $N$ for each datapoint $x_n$ indicating which other datapoints $x_n$ shares a Delaunay edge with. Together these vectors form a sparse symmetric neighbourhood matrix. In contrast when all pairs were considered this matrix would consist of all ones except for zeros along the diagonal. Using this neighbourhood 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 considerable 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 dataset, the amount of pairs increased linearly with $N$ (Fig 5.1). redThis results, when combined with the complexity of Delaunay triangulation $O(N\log(N))$, in a complexity of $O(N\log(N))$. In all of the following iterations the neighbourhood matrix is updated to reflect how merging trees, also causes neighbourhoods 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 datapoints, and the columns the possible clusters they can belong to. $N$ for belonging to themselves, and $N-1$ for each of the to be merged clusters throughout the iteration. Given this matrix and the neighbourhood matrix we can then define which pairs to test in each of the iterations following the initial one. redNote, when all Delaunay consistent paris have been exhausted, our implementation will return test all pair-wise comparisons.

## 5.4 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 $\theta = \{\boldsymbol{q}, \mu, \sigma\}$. Again integrating over the Gaussian part of the parameter space ($\mu$ and $\sigma$) is again 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 Eq. 5.9, while at the same time integrating over the Gaussian components. In what follows we will describe how we estimate the B-spline curve for a given dataset $D$.

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

$$g(t) = \sum_{m=1}^{M} B_m(t) q_m \tag{5.6}$$

, where $B_m$ are the basisfunctions and $q_m$ are the weigths assigned to these (also called the control-points). The order of the B-spline curve is defined by the order of the basisfunctions, here cubic splines were used. In the simulations above the number of basisfunctions and control points was set to $M = 6$. This number was choosen because it was a good compromise between the amount of parameters that govern the B-spline and the flexibility to take a wide range of shapes. From this curve we state that datapoint are generated perpendicular according a Gaussian likelihood function over the distance between a point on the curve $g(t_n)$ and the projected datapoint $x_n$ (see Eq. 3.9).

Given a dataset $D = \{x_1 \dots x_n\}$ 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:

$$p(D|\theta) = \prod_{n=1}^{N} \mathcal{N}(\|g(t_n) - x_n\| | \mu, \sigma), \tag{5.7}$$

$$p(\theta|\beta) = \exp(F_1|\lambda_1)\exp(F_2|\lambda_2)\mathcal{N}inv(\chi^2)(\mu, \sigma|\mu_0, \kappa_0, \sigma_0, \nu_0). \tag{5.8}$$

The likelihood function is the same as the generative function defined in Eq. 3.9. The last factor in the prior is the conjugate prior to the Gaussian distribution in the likelihood function, the Normal-inv($\chi^2$), allowing for analytical computation of the marginal over parameters $\mu$ and $\sigma$. The first two factors define the penalties on the first and second derivative of the curve respectively (we will show below how these are computed). 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 $\theta$ and select $\boldsymbol{q}$ as to maximize,

$$p(D|\beta,\boldsymbol{q}) = \int_{\mu,\sigma} \prod_{n=1}^{N} p(x_n|\mu,\sigma,\boldsymbol{q}) p(\theta|\beta) d\mu d\sigma \tag{5.9}$$

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 consists of 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 Eq. 5.9 given these $[x_n, t_n]$ pairs using the derivative-free optimization function *fminsearch* as implemented in MATLAB. Computing the value for the above 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,

$$F_i = \int_t \|D^i g(t)\|^2 dt, \tag{5.10}$$

where $i$ stands for the $i$th derivative. This integral was computed numerically by computing the $i$th derivative of $g(t)$ at 1000 equally sampled points along the curve. The integral over the Gaussian components of Eq. 5.9 could easily be computed analytically.

Specifically with $d_n = \|g(t_n) - x_n\|$ is a proposal parameter vector, this integral can be computed as follows

$$p(D|\beta, \boldsymbol{q}) = \frac{\Gamma(\nu_n/2)}{\Gamma(\nu_0/2)} \sqrt{\frac{\kappa_0}{\kappa_n}} \frac{(v_0\sigma_0)^{v_0/2}}{(v_n\sigma_n)^{v_n/2}} \frac{1}{\pi^{n/2}} \exp(F_1|\lambda_1)\exp(F_2|\lambda_2) \tag{5.11}$$

with,

$$\begin{aligned} \mu_n &= \tfrac{\kappa_0\mu_0 + N\bar{d}}{\kappa_n}, \\ \sigma_n &= \kappa_0 + N, \\ \nu_n &= \nu_0 + N, \\ \sigma_n &= \tfrac{1}{\nu_n}[\nu_0\sigma_0 + \textstyle\sum_n (d_n - \bar{d})^2 + \tfrac{N\kappa_0}{\kappa_0+N}(\mu_0 - \bar{x})^2], \end{aligned} \tag{5.12}$$

where $\bar{d} = \frac{1}{n}\sum_n d_n$. The two stages just described are then repeated until convergence of the function in Eq. 5.9.

# Bibliography


Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*(4), 303–314.

Amir, A., & Lindenbaum, M. (1998). A generic grouping algorithm and its quantitative analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(2), 168–185.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Attias, H. (2000). A variational bayesian framework for graphical models. *Advances in neural information processing systems*, *12*(1-2), 209–215.

August, J., Siddiqi, K., & Zucker, S. W. (1999a). Contour fragment grouping and shared, simple occluders. *Computer Vision and Image Understanding*, *76*(2), 146–162.

August, J., Siddiqi, K., & Zucker, S. W. (1999b). Ligature instabilities in the perceptual organization of shape. In *Computer vision and pattern recognition, 1999. ieee computer society conference on.* (Vol. 2). IEEE.

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*(3), 451.

Ben-Yosef, G., & Ben-Shahar, O. (2012). A tangent bundle theory for visual curve completion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *34*(7), 1263–1280.

Bishop, C. M. (2006). Pattern recognition and machine learning. (Chap. Mixture models and EM). New York, NY: Springer.

Blum, H. (1967). Models for the perception of speech and visual form. In N. Wathen-Dunn (Ed.). (Chap. A transformation for extraction new descriptors of shape). MIT Press, Cambridge, Massachusetts.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science*, *282*(5389), 746–749.

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

Cohen, E. H., & Singh, M. (2007). Geometric determinants of shape segmentation: tests using segment identification. *Vision Research*, *47*(22), 2825–2840.

Cohen, E. H., Singh, M., & Maloney, L. T. (2008). Perceptual segmentation and the perceived orientation of dot clusters: the role of robust statistics. *Journal of Vision*, *8*(7).

Compton, B., & Logan, G. (1993). Evaluating a computational model of perceptual grouping by proximity. *Attention, Perception, & Psychophysics*, *53*(4), 403–421.

Craft, E., Schütze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, *97*, 4310–4326.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

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).

Feldman, J. (2001). Bayesian contour integration. *Perception and Psychophysics*, *63(7)*, 1171–1182.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *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. D. . Z. Pizlo (Ed.). *Shape perception in human and computer vision: an interdisciplinary perspective*. Springer Verlag.

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

Feldman, J. (1997a). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835–2848.

Feldman, J. (1997b). Regularity-based perceptual grouping. *Computational Intelligence*, *13*(4), 582–623.

Feldman, J. (2003a). Perceptual grouping by selection of a logically minimal model. *International Journal of Computer Vision*, *55*(1), 5–25.

Feldman, J. (2003b). What is a visual object? *Trends in Cognitive Sciences*, *7*(6), 252–256.

Feldman, J. (2012). Symbolic representation of probabilistic worlds. *Cognition*, *123*(1), 61.

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*. (PhD thesis, Master Thesis, TU Wien).

Franconeri, S., Bemis, D., & Alvarez, G. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*(1), 1–13.

Frith, C. D., & Frith, U. (1972). The solitaire illusion: an illusion of numerosity. *Perception & Psychophysics*, *11*(6), 409–410.

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).

Fulvio, J. M., & Singh, M. (2006). Surface geometry influences the shape of illusory contours. *Acta Psychologica*, *123*(1), 20–40.

Geisler, W., Perry, J., Super, B., & Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711–724.

Gelman, R, & Gallistel, C. (1978). Young children's understanding of numbers. Cambridge: Harvard University Press.

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

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

Hoffman, D. D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*(1), 29–78.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*(1), 65–96.
Jaynes, E. T. (2003). *Probability theory: the logic of science*. 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.
Jiang, T., Dong, Z., Ma, C., & Wang, Y. (2013). Toward perception-based shape decomposition. In *Computer vision–accv 2012* (pp. 188–201). Springer.
Juni, M. Z., Singh, M., & Maloney, L. T. (2010). Robust visual estimation as source separation. *Journal of vision*, *10*(14).
Kalar, D. J., Garrigan, P., Wickens, T. D., Hilger, J. D., & Kellman, P. J. (2010). A unified model of illusory and occluded contour interpolation. *Vision Research*, *50*(3), 284 –299. doi:DOI:10.1016/j.visres.2009.10.011
Kanizsa, G., & Gerbino, W. (1976). Vision and artifact. In M. Henle (Ed.). (Chap. Convexity and symmetry in figure-ground organisation, pp. 25–32). New York, NY: Springer.
Kersten, D., Mammasian, P., & Yuille, A. (2004). Object perception as bayesian inference. *Annual Review of Psychology*, *55*, 271–304.
Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: a practical introduction*. Elsevier.
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What is new in psychtoolbox 3. *Perception*, *36*(14), ECVP Abstract Supplement.
Kontsevich, L., Tyler, C. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision research*, *39*(16), 2729–2737.
Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: a quantitative gestalt theory. *Psychological Science*, *6*(4), 225–234.
Kubovy, M., Holcombe, A., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive psychology*, *35*(1), 71–98.
Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *Journal of Optical Society of America*, *20*, 1434–1448.
Machielsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror-symmetry in visual shape detection. *Journal of Vision*, *9*(12), 1–11.
Maloney, L. T., Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: testing bayesian transfer. *Visual neuroscience*, *26*(01), 147–155.
Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3d interpretation of line drawings. *Vision research*, *38*(18), 2817–2832.
Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294.
McLachlan, G., & Peel, D. (2004). *Finite mixture models*. Wiley. com.
McLachlan, G. J., & Basford, K. E. (1988). Mixture models. inference and applications to clustering. *Statistics: Textbooks and Monographs, New York: Dekker, 1988*, *1*.
Miller, A. L., & Baker, R. A. (1968). The effects of shape, size, heterogeneity, and instructional set on the judgment of visual number. *The American Journal of Psychology*, *81*(1), 83–91.
Ommer, B., & Buhmann, J. M. (2003). A compositionality architecture for perceptual feature grouping. In *Energy minimization methods in computer vision and pattern recognition* (pp. 275–290). Springer.
Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive psychology*, *9*(4), 441–474.

Parent, P., & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *11*(8), 823–839.

Pelleg, D., & Moore, A. (2000). X-means: extending k-means with efficient estimation of the number of clusters. In *Proceedings of the seventeenth international conference on machine learning* (pp. 727–734). San Francisco.

Peterson, M. A., & Salvagio, E. (2008). Inhibitory competition in figure-ground perception: context and convexity. *Journal of Vision*, *8*(16), 1–13.

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*(3), 422.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: matlab routines for analyzing psychophysical data. *http://www.palamedestoolbox.org*.

Richards, W., Dawson, B., & Whittington, D. (1986). Encoding contour shape by curvature extrema. *JOSA A*, *3*(9), 1483–1491.

Rissanen, J. (1989). *Stochastic complexity in statistical inquiry theory*. World Scientific Publishing Co., Inc.

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

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

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(8), 888–905.

Siddiqi, K., & Kimia, B. B. (1995). Parts of visual form: computational aspects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *17*(3), 239–251.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal of Computer Vision*, *35*(1), 13–32.

Singh, M., & Feldman, J. (2008). Skeleton based decomposition of shapes into parts.

Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing silhouettes: the short-cut rule. *Perception & Psychophysics*, *61*(4), 636–660.

Singh, M., Feldman, J., & Froyen, V. (in preparation). Invited manuscript for: handbook of computational perceptual organization. In S. Gepshtein & L. T. Maloney (Eds.). (Chap. Unifying parts and skeletons: a Bayesian approach to part decomposition). Oxford University Press.

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, *16*(1), 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*(2), 163–177.

Song, Y.-Z., & Hall, P. M. (2008). Stable image descriptions using gestalt principles. In *Advances in visual computing* (pp. 318–327). 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 Lier, R., Van der Helm, P., & Leeuwenberg, E. (1995). Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 571.

Van Lier, R. (1999). Investigating global effects in visual occlusion: from a partly occluded square to the back of a tree-trunk. *Acta Psychologica*, *102*(2), 203–220.

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

Van Oeffelen, M., & Vos, P. (1982). Configurational effects on the enumeration of dots: counting by groups. *Memory & Cognition*, *10*(4), 396–404.

Vos, P. G., Van Oeffelen, M. P., Tibosch, H. J., & Allik, J. (1988). Interactions between area and numerosity. *Psychological Research*, *50*(3), 148–154.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012a). A century of gestalt psychology in visual perception: i. perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*(6), 1172–1217.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012b). A century of gestalt psychology in visual perception: ii. conceptual and theoretical foundations. *Psychological Bulletin*, *138*(6), 1218–1252.

Wertheimer, M. (1923). Untersuchungen zur lehre von der gestalt, ii. *Psychologische Forschung*, *4*, 301–350.

Wilder, J. (2013). *The influence of complexity on the detection of contours*. (PhD thesis, Rutgers University).

Williams, L. R., & Jacobs, D. W. (1997). Stochastic completion fields: a neural model of illusory contour shape and salience. *Neural Computation*, *9*(4), 837–858.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS computational biology*, *6*(8), e1000871.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.

Zucker, S. W. (1985). Early orientation selection: tangent fields and the dimensionality of their support. *Computer Vision, Graphics, and Image Processing*, *32*(1), 74–103.

Zucker, S. W., Stevens, K. A., & Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics*, *34*(6), 513–522.