

# Perceptual Models of Small Dot Clusters

Jacob Feldman  
Rutgers University Center for Cognitive Science  
New Brunswick, New Jersey

Running head: *Models of dot clusters*

Send correspondence to:

Jacob Feldman  
Center for Cognitive Science  
Department of Psychology, Busch Campus,  
New Brunswick, New Jersey, 08903  
e-mail: [jacob@ruccs.rutgers.edu](mailto:jacob@ruccs.rutgers.edu)

### Abstract

This paper investigates how human observers draw perceptually coherent clusters out of fields of dots, a process analogous to but different from statistical clustering. We focus on the way perceptually minimal clusters—dot triplets—are categorized by human observers as having arisen from either a curvilinear generating process (a chain) or an unordered random process (a cluster). In order to draw perceptual decompositions from larger fields of dots, observers must attach class-conditional probability densities to these minimal (“molecular”) clusters; that is, they must have a model of just how straight is straight “enough” for a dot triplet to be reasonably classified as a chain rather than a cluster. We first consider the normative statistical facts concerning how three completely random dots can be expected to be distributed. Human observers lack this knowledge, however, and must construct prior distributions for the shape of both random and “ordered” triplets from internal models of each generating process. Here, we work out a theory of such models that enables the observer to construct usable (though not actually “correct”) class densities. In accordance with the theory, human subjects interpret a configuration of three equidistant dots (counter-normatively) as the most typical (“generic”) configuration of three randomly positioned dots—apparently because this case is maximally distant in the dot triplet space from the non-generic case of three collinear dots. Subjects’ category judgments precisely resemble the orthodox Bayesian likelihood inference they would draw if their implicit probabilistic models took the regular categorical form described by the theory. In this sense, subjects’ prior probabilistic beliefs, and hence their qualitative interpretations of observed dot patterns, are built around *regularities* that they tacitly believe exist in the dots world.

## 1 Perceptual clustering of dots

The parallel between clustering a set of points in a data space by statistical techniques (on the one hand) and decomposing a field of dots in the visual field by direct inspection (on the other) is an appealing one—and yet the two processes are peculiarly different. Consider, for example, Fig. 1 (below). To a human observer, it is obvious that there are two natural groups or clusters of dots. This inference comes with no conscious effort or thought, and seems both stable and compelling. Notoriously, however, statistical techniques find this kind of inference a difficult one. Commonly, the number and general form of clusters must be supplied to the algorithm (see Duda and Hart, 1973, or Späth, 1980, for introductory surveys). Though techniques exist to cope with clusters both of unknown number (e.g. hierarchical techniques) and of unknown form (e.g. nonparametric techniques such as that of Matthews and Hearne, 1992), their behavior differs from that of human observers in intriguing and systematic ways. Most techniques have particular trouble, for instance, distinguishing among clusters of different apparent dimensionality; or, putting it slightly differently, in determining the exact dimensionality at which a given subset of the data is best modeled (a problem so pervasive it is often termed “the curse of dimensionality”). A normative definition of the “correct” dimension for a given set of data is elusive, if not impossible—and consequently, so is a definition of the “correct,” or even the “most reasonable” decomposition of a given data set. Human observers, however, have strong intuitions on the question. This paper attempts to characterize these intuitions formally, focusing on a critical sub-problem: the classification of minimal groups consisting of three dots.

The three dots case can be seen as the minimal one in the sense that three dots constitute the smallest dot group that has *shape* (cf Kendall and Kendall, 1980): that is, structure after the similarity groups (translation, orientation, and scale) have been removed. Hence three dots form the smallest cluster that can exhibit any tendency to cohere due to shape—in particular, as we shall focus on in this paper, due to alignment or *collinearity*. This perceptual tendency, which will be investigated empirically as well as theoretically below,

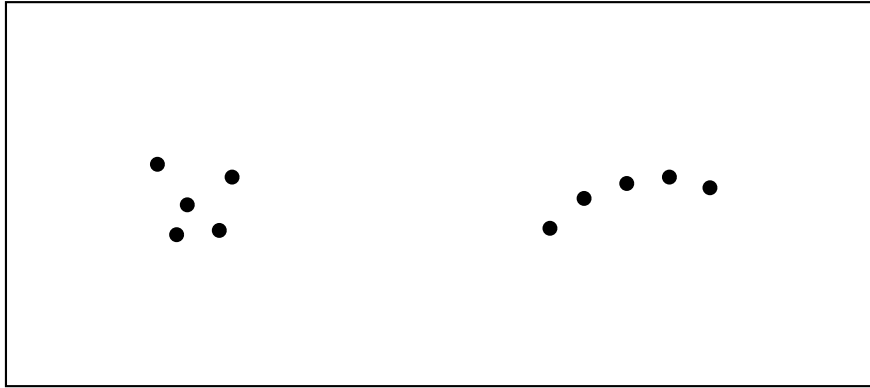


Figure 1: Two clusters, of different perceptual “types”.

can be seen as a minimal kind of categorical inference, in that three nearly-collinear dots appear to be classified as having been produced by a qualitatively different kind of generating process than three far-from-collinear dots (that is, a curvilinear one rather than an unordered random one). In a more abstract sense, we will propose, three collinear dots, but not three unordered dots, are sensibly interpreted as having been generated by a *causally coherent* process: what Jepson and Richards (1993b; cf Feldman, 1992b) call a “regularity”.

Describing the human inference in this domain turns critically on the exact nature of human perceivers’ *prior probabilities* (see Ashby and Gott, 1988; Nosofsky, 1991)—that is, their primitive probabilistic assumptions about exactly how likely various configurations of dots are to occur in the dots world. (Here by “priors” we mean both the prior probability of each class of dot configurations occurring in the world, as well as the class-conditional densities that dictate how likely each configuration is within each class.) We start by assuming that subjects assign different prior probabilities to different dot triplets, the minimal case; and then draw clustering inferences by combining decisions about triplets. The bulk of this paper is occupied with asking what the human intuitive priors for dot triplets are, and why. At the end we return only briefly to the question of how the decisions about triplets are combined to form overall grouping interpretations; a more general model of which is presented in Feldman, 1995. For the three dots case, it will turn out that the human subjects draw inferences in a recognizable (though unorthodox) Bayesian way. It will be argued

that the exact assumptions prior they make in order to do so reflect an attempt to recover regularities—the observable evidence of underlying causal forces—in the dots world.

The central issue of the paper, then, concerns human observers’ tacit beliefs about the probability of various configurations of dots in the plane, in the absence of any explicit domain-specific knowledge. To provide some kind of a standard against which these mental constructs can be compared, we present some quasi-normative facts about dot densities due to the statisticians Kendall and Kendall (1980). These densities, of which subjects presumably have no conscious knowledge (they are in fact counterintuitive in several respects) serve as kind of neutral backdrop for observers’ naive hypotheses. Section 2.2.1 lays out these basic statistical “facts.” In grouping dots, as will be discussed below, observers need to know how typical the observed configuration is of each of several hypotheses, in order to judge whether a given grouping hypothesis explains the observation to a substantial degree. But since the true normative priors are unavailable to them, they are forced to substitute “reasonable” guesses as to how just how the configurations generated by each grouping hypothesis are distributed. Just how these imaginary distributions are constructed will be the main topic of interest.

*Preview of the theory* We will first propose that minimal dot groups (triplets of dots) are divided into two qualitatively distinguishable categories. Then, each category will have a separate “modal” class-dependent density, centered on a “most typical” case. (We call such distributions “modal” because the density peaks correspond to qualitative types.) Within each mode, the highest (most typical) point is stipulated (heuristically and non-normatively) to occur at a point that is maximally distant, and thus maximally distinguishable, from adjacent modes. The observer thus implicitly imputes a certain convenient, regular, structure to the world—a structure that is highly speculative and not necessarily “true,” but which nevertheless aids in drawing robust and useful inferences about regularities in the world (like tendencies of dots to cohere to one another in particular ways). Thus the clustering scheme described here can be thought of as depending on a version of the “kind world” assumption (cf Bennett, Hoffman, and Prakash (1989)).

The next section begins by situating the three dots case, and the detection of collinearity in fields of dots, in both perceptual psychology as well as in a clustering framework.

### 1.1 Two kinds of dot clusters

Going back to Fig. 1, notice that the two clusters in the figure seem to be of two perceptually distinct types: the cluster on the left seems to scatter randomly in two dimensions, while the cluster on the right seems to be an ordered chain of dots falling along some sort of curvilinear path. The Gestalt psychologists first observed the two apparently distinct grouping principles at work here, calling them the principle of *proximity* and *good continuation*, respectively. Gestalt grouping principles have more of a descriptive than an explanatory flavor, however, in that they seem not to provide any justification or reasoning under which these types of grouping are superior to any other. Moreover Gestalt principles provide no prediction of the exact magnitude of the tendency for dots to cohere in various ways.

The extraction of curvilinear structure from fields of dots has been approached from a number of directions since. Much attention has naturally focused on large fields of dots, in particular Glass patterns (Glass, 1969), leading to an emphasis on the process by which local evidence of collinearity and parallelism is collected and assembled into a global percept of structure. This process has been modeled computationally (Stevens, 1978, Zucker, 1985). Many researchers (Glass, 1969; Caelli and Julesz, 1978; Prazdny, 1984; Brookes and Stevens, 1991) have found evidence that human grouping inferences depend at least in part on orientation-tuned local operators. On this view the perceptual salience of a group of nearly collinear dots depends on its literal resemblance (shared low spatial frequency component) to an oriented segment, the ideal form preferred by some cortical cells. However, evidence also exists (Stevens and Brookes, 1987) that such operators are not sufficient to account completely for human percepts, suggesting that symbolic computations on location tokens might also be necessary. Excellent discussion of the difference between these two approaches and suggestions for how they might be reconciled may be found in Foster (1984) and Stevens and Brookes (1987). As pointed out by Zucker (1984), the “how” question can

be distinguished from a “why” question, which asks what sort of distal environmental structure the system is actually looking for, and how this search might be aided by the detection of particular kinds of proximal structure. In this light Witkin and Tenenbaum (1983) have argued that low-level structure and regularity in the image serves as an underpinning for a “meaningful” representation of causal forces at work in the world. The “why” question also serves to relate perceptual grouping to models of statistical clustering, in which it is natural to consider exactly how assumptions about the probability distributions governing the observer’s environment entail particular inference mechanisms. In this spirit, the current paper focuses on relatively small numbers of dots, attempting to ask what tacit assumptions the human perceptual system appears to be making when it ascribes collinearity to a collection of dots. Thus this question can be seen as orthogonal to the question of how these decisions are implemented, and indeed it seems entirely plausible that the resulting decision schemes might be carried out by local operators.

All models of collinearity detection assume a maximal response to perfectly collinear stimuli, with response decreasing as deviation from collinearity increases. But there has generally not been a motivated way to predict the exact magnitude of the off-angle response—the function governing how quickly the cohesiveness of dot clusters degrade as they range from perfect collinearity to extreme non-collinearity. The analysis below will provide exact (and apparently empirically correct) predictions for the magnitude of the off-angle response, which emerge essentially from an analysis of the structure of the space of grouping hypotheses.

The goal, then, is to construct a concrete model of the inference scheme that underlies the grouping reflex. Ideally, such a scheme would attach to each configuration of dots an exact numerical estimate of the probability that the configuration merits a particular category label—that is, that it was produced by a certain type of dot-generating process.

First of all, we assume that exactly two kinds of dot-generating process inhabit the dots world: 2-D patches in which dots are generated uniformly (like a Poisson process); and

1-D curves along which dots are generated at random intervals of arclength<sup>1</sup> (see Fig. 2). The curvilinear process may be thought of as a kind of “regularity” in this domain (see Feldman 1992a for further discussion of this notion), while the 2-D patch process may be thought of as a region of non-regularity, i.e. a region in which dot locations are all generated independently of one another, rather than being causally co-related. These two processes are by no means the only types one might imagine in the dots world, and in one view restricting attention to them corresponds to ascribing only a narrow, contingent belief to the observer. On the other hand, the close correspondence between these two processes and intrinsic mathematical objects (e.g. curves and surfaces) as well as basic perceptual primitives (e.g. edges and regions) suggests that they represent a fundamental case (Zucker, 1985). In any case, these types of probabilistic process underlie all the discussion below, though the two types will be represented in several other ways, all of which are in the cause of understanding what observers perceive as “typical” outcomes of each generating process.

We denote individual dots in the plane by  $\mathbf{x} \in R^2$ , and an observed set of  $N$  dots  $\{\mathbf{x}_1 \dots \mathbf{x}_N\}$  by  $\mathbf{X}^N$ , omitting the  $N$  whenever unambiguous (e.g. later when we have fixed on the case  $N = 3$ ). Thus to each set of dots  $\mathbf{X}$  we would like to attach two inferred quantities: the probability of a “chain” category,  $p(C_c|\mathbf{X})$ , and the probability of a “random cluster” category,  $p(C_r|\mathbf{X})$ . Notice that it is only under our assumptions about the dots world—that it contains two types of regular process, the dot-producing curve and the internally unstructured cluster—that these posterior probabilities represent meaningful inferences about something going on in the world.

## 1.2 Molecular clusters

In order to construct these functions we begin with the smallest observation from which anything about curvilinear processes can be inferred. We would like to ignore the effects on  $p(C_i|\mathbf{X})$  of translation, rotation, and scale of  $\mathbf{X}$ . Since all pairs of points are equivalent

---

<sup>1</sup>The fact that we restrict our attention in this paper to small dot clusters (triplets) relieves us of worrying about global properties of these two generating patches, such as the border of the 2-D patches and the connectivity and complexity of the 1-D curves.

under these operations, we take triplets as a minimal set. Since, in this sense, a triplet is the smallest set of dots that can still meaningfully exhibit “chain”-ness or “cluster”-ness, we call it a “molecular” chain or cluster respectively. A larger chain would then be made up of consecutive interlocking 3-chains, e.g.  $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ , etc. Note that for the moment we postpone choosing a convenient parameterization for each triplet—such as by its main angle and the lengths of its two virtual “legs”—until one can be motivated more carefully in Sec. 2.1.

A minimal chain, which we will refer to as a “3-CHAIN”, is a set of three dots generated by the curvilinear process; a minimal cluster (a 3-CLUSTER) is a set of three dots produced by the patch process (Fig. 2). We assume that these two processes occur in the world with probabilities  $p(\text{3-CHAIN})$  and  $p(\text{3-CLUSTER})$  respectively.

Naturally we do not expect the two generating processes to produce identical-looking dot triplets. Rather, we expect 3-CHAINS to “typically” be straighter than 3-CLUSTERS, as in the figure. To be more precise, for any three dots, we pick the straightest of the three possible orderings of the dots; and then measure how far even this straightest ordering is from perfect collinearity. We then expect that the distribution of this angle among 3-CHAINS would be centered relatively near zero (collinearity), and that of 3-CLUSTERS relatively far from zero. (The parameterization we will introduce later is designed to help make this idea more explicit.)

The observer seeking to classify dot triplets requires the two quantities  $p(\text{3-CLUSTER}|\mathbf{X})$  and  $p(\text{3-CHAIN}|\mathbf{X})$ . We have assumed that these are the only two categories into which a triplet of dots can be put (this assumption will be spelled out in more detail and defended below). Hence by stipulation we immediately have

$$p(\text{3-CLUSTER}|\mathbf{X}) = 1 - p(\text{3-CHAIN}|\mathbf{X}) \quad (1)$$

Hence for the remainder of the paper we will be concerned simply with the shape of the posterior function  $P(\text{3-CHAIN}|\mathbf{X})$ . Intuitively, the question then is: in order for a triplet of dots to be considered a 3-CHAIN, how straight is straight enough; and why?

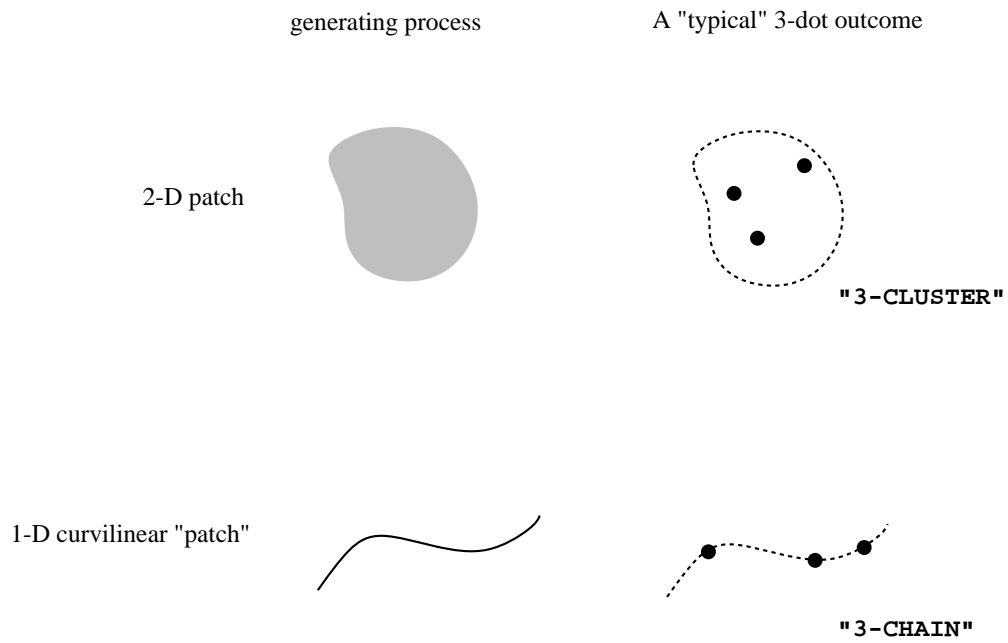


Figure 2: A schematic depiction of the two generating processes or “regularities” under consideration, showing for each the form of the underlying generating process, and a “typical” 3-dot outcome (i.e. a 3-CLUSTER or a 3-CHAIN).

By Bayes' theorem we know that the posterior function *ought* to take the form

$$p(\text{3-CHAIN}|\mathbf{X}) = \frac{p(\text{3-CHAIN})p(\mathbf{X}|\text{3-CHAIN})}{p(\mathbf{X})}, \quad (2)$$

where the denominator  $p(\mathbf{X})$  is composed of contributions from the two generating processes:

$$p(\mathbf{X}) = p(\text{3-CHAIN})p(\mathbf{X}|\text{3-CHAIN}) + p(\text{3-CLUSTER})p(\mathbf{X}|\text{3-CLUSTER}), \quad (3)$$

where again  $p(\text{3-CHAIN})$  and  $p(\text{3-CLUSTER})$  are the prior probabilities with which these two categories occur in the world. Whether human observers *actually* draw their category judgments in this normative manner, though, and what values they give to the various priors, are largely empirical questions, which will be taken up as such in the experiments reported below.

Because  $p(\text{3-CLUSTER})$  and  $p(\text{3-CHAIN})$  are scalars, most of the structure in the posterior function will come instead from the two class density functions  $p(\mathbf{X}|\text{3-CLUSTER})$  and  $p(\mathbf{X}|\text{3-CHAIN})$ . Various theories about the proper structure of these densities will be taken up in detail in the next section. We will then be in a position to ask whether there exist priors and class-conditional densities such that human observers' category decisions actually follow the Bayesian formulation, and what tacit principles observers seem to follow in order to choose these priors.

## 2 Models of dot-producing processes: statistical and perceptual

In probabilistic theory, and by extension in the theory of clustering, it is natural to think in terms of a "generating process," some agent that gives rise to a particular set of observed outcome events with particular probabilities. Such an idea has, perhaps peculiarly, been slower to gain prominence in the theory of perceptual inference. Recently, however, Leyton (1984, 1989) has drawn attention to the central role of generative models in perceptual

description, in particular in the theory of shape representation. Each shape, in Leyton's model, is conceptualized as having been produced by the deformation of some simpler shape. The resulting representation of a complex object is as a nested series of generative operations from which the object is interpreted as having been generated. This kind of representation thus corresponds to a kind of "causal history" by which the observed object is interpreted as having been generated. This process imparts to the interpretation more of a quality of explanation rather than mere description (see also Witkin and Tenenbaum, 1983).

In earlier work (Feldman, 1992b) it was observed that the hierarchical organization of such generative processes can be usefully captured by a certain discrete structure called the *category lattice* (see also Jepson and Richards, 1993a and 1993b). In such a lattice, each node represents a class of objects in the overall space. Moving down the lattice, one moves from general and inclusive classes to more structurally constrained and specific classes. The generative history of an object corresponds, in this scheme, to the exact route by which the object's history has transformed it from the completely structurally fixed class at the bottom of the lattice (the simple primitive object). A critical idea in this scheme is the idea of *genericity*: loosely, that each observed object should be associated with a node on the lattice in which it is generic, that is in which all the generative operations in its description have been carried out to some non-zero degree. Our use of this scheme in representing the two categories 3-CHAIN and 3-CLUSTER will be relatively simple and self-contained, but it is worth noting that it is consistent with this more fully articulated and complex scheme which covers a wider range of perceptual categories in other types of object spaces. Moreover, the lattice scheme has the power to productively enumerate distinct types of cluster beyond the two considered here, thus placing these two in their proper formal perspective (Feldman, 1995).

A few remarks are in order to place the category apparatus into the context of existing perceptual theory. The lattice scheme is predicated on the search for "regularities" in the observed world, a world which exhibits more consistently regular structure than is in principle necessary (see also Richards, Jepson, and Feldman 1993). An observer detects

this structure in order to help recover the causal forces at work in the environment, and thereby to infer reliable and useful facts about the world. In Feldman (1992a) the notion of regularity was cast, broadly and abstractly, as a lower-dimensional manifold through some higher-dimensional embedding space of objects. In the dots world this notion reappears concretely as our 1-D curvilinear process. As will be detailed below, this is our central model of a chain-producing process, whose presence in the environment means that the inference of a “chain” is correct. Thus while the theory below is intended to have direct application to the problem of psychological clustering, it may also be seen as an attempt to focus probabilistic machinery on regularity concepts and category concepts that were originally conceived to apply to a wider range of perceptual inferences. For our purposes here, the category structure and genericity notions are necessary to provide an underpinning for a more focused set of probabilistic questions: questions concerning the class-conditional densities associated with each of the modal categories, and about the decision procedure by which human observers place a set of observed dots into one of the categories.

## 2.1 A generative model of dot triplets

Following Feldman (1992b), we now construct a generative model for dot triplets, which will serve to determine a parameterization for dot triplets. This in turn will determine a category lattice formally expressing the (admittedly simple) relationship between the categories 3-CLUSTER and 3-CHAIN. Each category on the lattice can be thought of as the “orbit” of some set of constructive operations (cf Hoffman, 1966; Dodwell, 1983; Leyton, 1984, 1989).

We begin with a reference dot,  $\mathbf{x}_b$  (Fig. 3). We imagine a “translation process” that positions a dot  $\mathbf{x}_c$  along some translation vector  $\vec{t}$  away from  $\mathbf{x}_b$ ,  $\mathbf{x}_c = \mathbf{x}_b + \beta_1 \vec{t}$  for some scalar  $\beta_1$ . A third dot,  $\mathbf{x}_a$ , is required to determine a reference direction for  $\vec{t}$ , i.e.  $\vec{t}$  lies in the direction of the vector  $\mathbf{x}_b - \mathbf{x}_a$ . Thus we can think of an entire space of dot triplets that are spanned by the translation process alone, consisting of the exactly collinear triplets in which  $\mathbf{x}_c - \mathbf{x}_b = \beta_1 \vec{t}$ .

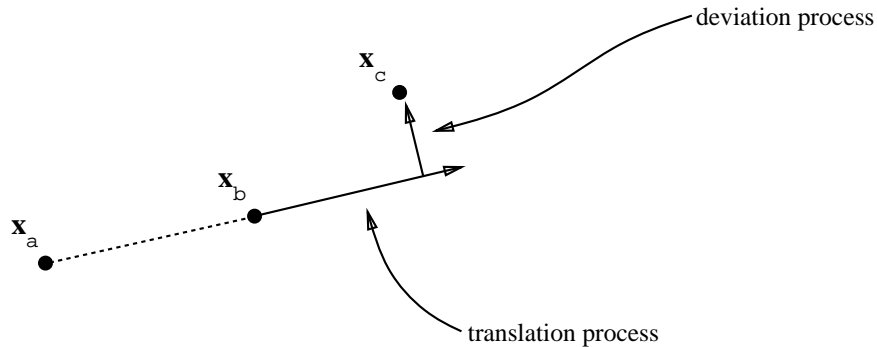


Figure 3: The “generative model” for dot triplets. A “translation process” carries  $\mathbf{x}_c$  out from  $\mathbf{x}_b$ , along the line joining  $\mathbf{x}_a$  to  $\mathbf{x}_b$ . A transverse “deviation” process carries  $\mathbf{x}_c$  away from this line.

To express dot triplets outside this special subspace, we need a second process, which we call the “deviation” process  $\vec{d}$ ; this process carries the third dot in a direction normal to the tangent direction  $\mathbf{x}_b - \mathbf{x}_a$ . Critically, all dot triplets can be expressed as a weighted sum of these two processes, i.e. any triplet can be written as  $\{\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_b + \beta_1 \vec{t} + \beta_2 \vec{d}\}$  for some scalars  $\beta_1$  and  $\beta_2$ . The two operations are of course isomorphic to the familiar tangent and curvature vectors of a curve respectively, except reconceived constructively as producing discrete dots at intervals of non-vanishing arclength.<sup>2</sup>

In combination, these two processes can produce any triplet. Triplets produced by the combined operation of some non-zero amount of *both* operations ( $\beta_1 \neq 0, \beta_2 \neq 0$ ) are called *generic* in the space defined by the two operations, i.e. in the triplet space. We refer to this category as  $\vec{t} + \vec{d}$ , dropping the coefficients  $\beta_1$  and  $\beta_2$  because we are collapsing over the exact amount to which each operation is applied. Triplets produced by the translation process *only* are by definition non-generic in the overall space, but can be thought of as generic in a subcategory  $\vec{t}$ . Notice that there are no triplets defined by  $\vec{d}$  without  $\vec{t}$ , because  $\vec{d}$  is defined only with respect to a tangent vector. Triplets defined by zero magnitudes on

<sup>2</sup>Readers may also note the relevance of the “fundamental theorem of curves” (see Do Carmo, 1976) which tells us that in a certain sense these two vectors specify all there is to know about the shape of the underlying curvilinear process near the point  $\mathbf{x}_b$ .

both vectors are degenerate. Hence the only categories we need to consider are  $\vec{t}$  and  $\vec{t} + \vec{d}$ .

The space spanned by the two processes, i.e. the set of all dot triplets, can of course be conveniently parameterized by magnitudes to which the two processes were carried out,  $\beta_1$  and  $\beta_2$ , for which we substitute the more meaningful (but diffeomorphic) parameters angle  $\alpha$  and line-length ratio ( $\frac{\|\mathbf{x}_b \mathbf{x}_c\|}{\|\mathbf{x}_a \mathbf{x}_b\|}$ ). This natural parameterization of the space by leg-length ratio (proportional to  $\beta_1$ ) and angle  $\alpha$  (proportional to  $\beta_2$ ) is depicted in Fig. 4. The idea is that this parameterization is not simply an arbitrary choice, but rather corresponds straightforwardly to the particular generative model of triplets we have adopted.

For the remainder of the paper we will restrict attention to the angle  $\alpha$ , the parameter on which the regularity turns out to be defined; the line-length ratio is devoid of special values and hence of special subcategories. Therefore we will usually write  $p(3\text{-CHAIN}|\alpha)$  rather than  $p(3\text{-CHAIN}|\mathbf{X}^3)$ , etc.

The close relationship between the chosen parameterization and the chosen generative model provides a clean formalism for distinguishing among subcategories of triplets that have been produced by qualitatively different generative histories—namely, either by the full model  $\vec{t} + \vec{d}$  (i.e., generic triplets) or by the submodel  $\vec{t}$ , at which the angle  $\alpha$  is 0 and the deviation process completely drops out. At this point, the curvature of the inferred underlying generating process vanishes. Hence the category  $\vec{t}$ , containing only collinear triplets which are non-generic in the overall space, is a very special subcategory. It represents a kind of idealized form for the output of the curvilinear process—not only a central tendency, but a kind of “best case:”, one from which detection of the curvilinear process is maximally (though not perfectly) reliable. (See Feldman, 1991, for some subtleties on this point). By this reasoning straight triplets constitute an idealized prototype of triplets generated by a curvilinear process.

The category lattice depicting this structure is shown in Fig. 5. The parametric collapse from  $\vec{t} + \vec{d}$  to  $\vec{t}$  (we can omit the weights because we are not distinguishing among generic members of the same category) is realized as a transition from the *generic* dot triplet category at the top of the lattice to the lone special subcategory at the bottom. This

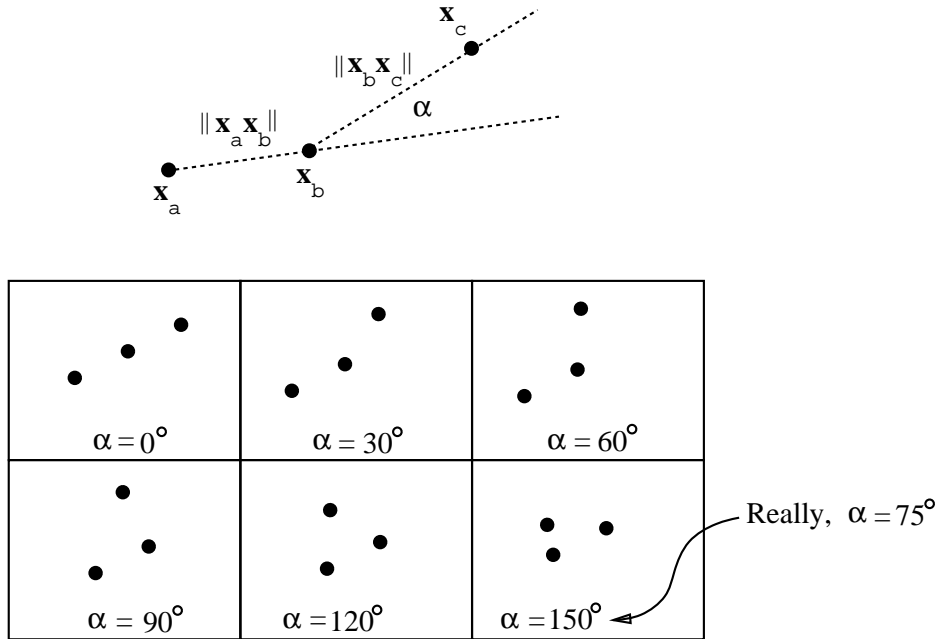


Figure 4: The parameterization for dot triplets corresponding to the translation-deviation generative process. We can fully express any dot triplet (up to similarity) by  $\alpha$ , the angular deviation from straight, and  $\frac{\|x_b x_c\|}{\|x_a x_b\|}$ , the ratio of the length of the second “leg” to the length of the first leg. Below, examples are given of dot triplets with  $\alpha$  at various values (and  $\frac{\|x_b x_c\|}{\|x_a x_b\|} = 1$ ). Notice that  $\alpha$  cannot exceed  $120^\circ$ , since it is always measured as the angle nearest to straight. Hence, in the bottom right example, where the angle grows to  $150^\circ$ ,  $\alpha$  should now be measured at one of the other two vertices where it is now closer to straight; instead of  $150^\circ$ , it now becomes  $75^\circ$ .

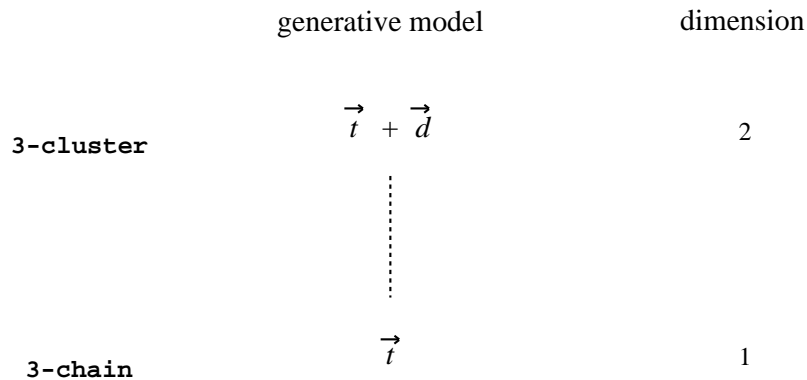


Figure 5: The category lattice depicting the relationship between 3-CHAIN and 3-CLUSTER. For each category, the lattice gives the symbolic representation of its generative model, making explicit the dimensional relationships among distinct categories.

category transition carries a change in dimension from the 2-D space spanned by both parameters to the 1-D subspace spanned by the translation parameter only.

Now that we have a model of how members of each category 3-CLUSTER and 3-chain are generated, we can ask what “typical” products of each generative process would look like. This question may be regarded two ways: statistically or psychologically. The statistical question is: given the mathematical definitions of these two generative processes, what will be the resulting distribution of angle? The psychological question is: given the mental models of these two generative processes, what will be the resulting subjective expectation of dot triplets? The statistical “facts of the matter” can be seen of as a kind of normative backdrop against which human observers’ beliefs can be compared. As it happens, the mathematical answer for the generic category 3-CLUSTER was provided by Kendall and Kendall (1980), whose results are summarized in Section 2.2.1. The non-generic category 3-CHAIN requires different treatment, and yields a somewhat less clear-cut answer; this will also be taken up below. Before considering either the normative or the psychological questions, though, it is helpful to consider in more detail the mathematical structure of the triplet space, towards the end of clarifying the difference between the mental and mathematical constructions of the triplet distributions.

### 2.1.1 The maximally generic point in triplet space

In Feldman (1992b), evidence was presented (in a different domain) that subjects construct prior probability distributions—that is, they create their own beliefs about the statistics of the world—in such a way that there are peaks (modes) at each distinct category in their model of it. In the dot triplets domain, this would mean simply that there would be two peaks in their mental density function: one at the non-generic (collinear) subcategory, i.e. at  $\alpha = 0$ , and one at a point construed to be *maximally* generic—the prototypical 3-CLUSTER (Nosofsky, 1984). Putatively, following Feldman (1992b), maximally generic simply means maximally *distant* from nearby non-generic points. In the dot-triplet domain, the most generic angle is simply the angle most distant from the non-generic point  $\alpha = 0^\circ$ : namely  $120^\circ$ , the largest value the angle can take, in accordance with the symmetry argument presented above. Though it is easy to see that  $120^\circ$  is the angle most distant from straight, it is worth expanding on this point a bit in order to clarify the internal structure of the triplet space.

Consider the space of dot triplets as a geometric object. We can display its structure in a convenient form by the contrivance of placing the first and second dots,  $\mathbf{x}_a$  and  $\mathbf{x}_b$ , at points  $(-1,0)$  and  $(0,0)$  respectively, so that the position of the third dot  $\mathbf{x}_c$  is parameterized by the ordinary Cartesian coordinates of the space (Fig. 6). In order to focus only on angle and ignore leg-length ratio, consider the part of this space containing just dot triplets with equal interdot distances,  $\frac{\|\mathbf{x}_b\mathbf{x}_c\|}{\|\mathbf{x}_a\mathbf{x}_b\|} = 1$ . This part of the space falls naturally on the circle  $\|\mathbf{x}_c\|^2 = 1$  (shown in the picture).

Now consider how our requirement that we always measure the straightest angle in any dot triplet imposes a kind of *symmetry structure* on this circular space. Imagine moving the dot  $\mathbf{x}_c$  along the periphery of this circle starting from the rightmost point. In this region, labeled **A** in the figure, the angle of the corresponding dot triplet changes from  $0^\circ$  at the rightmost point to  $90^\circ$  at the topmost point, etc. If however one were to continue past the point  $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$  (where  $\mathbf{x}_c$  is actually drawn in the figure), one would begin to form triplets whose straightest vertex was not equidistant from the other two dots. So instead

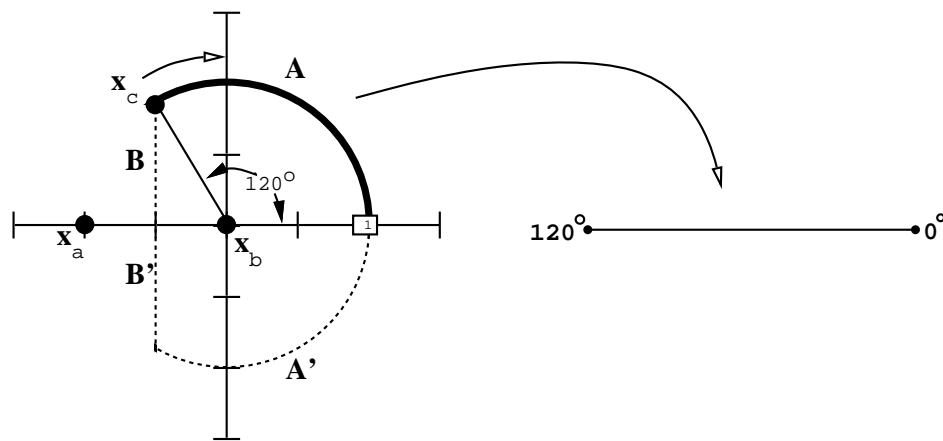


Figure 6: Schematic depiction of the space of dot triplets, making clear how  $\alpha = 0$  and  $\alpha = 120^\circ$  are maximally distant points. The space of equal-leg-length dot triplets can be depicted (left) as a part of a circle embedded in the 2-D space of general dot triplets; the highlighted part (labeled A in the figure) contains each distinct triplet exactly once. This space can be rewritten (right) as a linear space containing two definite endpoints,  $0^\circ$  and  $120^\circ$ . These endpoints, which are maximally distant from each other in a metric-invariant sense, become the prototypes of the two opposed categories: one the non-generic point, the other the maximally generic point.

we can continue moving  $\mathbf{x}_c$  straight *down*, along the segment marked **B**, until we meet the  $x$ -axis. Notice that in segment **B** we encounter each and every triplet from **A** again, only scaled, rotated and translated. The two sections **A** and **B**, that is, contain exactly the same (shaped) triplets. In a similar though less subtle way, the parts labeled **A'** and **B'**, mirror images of **A** and **B** respectively, also contain exactly the same set of triplets, except reflected. Thus one can move continuously from the rightmost point on the circle all the way around and back to the rightmost point, encountering each unique triplet exactly four times.

In order to take “shape” only, then, we can map a single one of these iterations (the figure indicates section **A**, but the choice is an arbitrary one) to a linear space containing each triplet just once (shown on the right in the figure). This space can now be taken as representing the true triplet space in its most intrinsic form. Notice that it has two definite endpoints:  $0^\circ$  and  $120^\circ$ , respectively. Notice that these two points are maximally distant from each other, in a sense that is completely insensitive to the choice of metric. Now since  $0^\circ$  has been identified as a non-generic point in the space, we can confidently identify  $120^\circ$ , the equilateral triplet, as the maximally generic point.

Intuitively, this point in the space is the triplet that is (psychologically) the least straight, in the sense of being maximally remote from the nearest possible attribution of straightness. That is, the ordering of the three dots is as ambiguous as possible, and the triplet’s probability of assignment to the category 3-CHAIN ought to be as low as possible. These two maximally remote cases thus become the ideals or prototypes of the two qualitatively opposed generating processes. The idea, then, is that human observers would construct their probabilistic beliefs so that the complementary category, 3-CLUSTER, would be maximally probable at this point, so that the two categories could be optimally distinguished. Below we will present empirical evidence that subjects do in fact construct their class density functions in exactly this manner.

## 2.2 Statistical models of dot triplets

We now consider ways in which probability distributions for first 3-CLUSTER and then 3-CHAIN may be obtained. For each type of molecular cluster, there is a quasi-normative distribution due to Kendall and Kendall, which we present first. Then we consider alternative distributions constructed around the category theory presented in the previous section. The latter probabilities, which the experiments below suggest subjects actually adopt, seem to reflect their tacit beliefs about the regularities extant in the dots world.

### 2.2.1 The basic statistical facts

(1) *Random clusters of three dots.* Kendall and Kendall (1980; see also Broadbent, 1980) considered the question: when triangles are generated at random (i.e. from selected as 3 points in the plane chosen from a 2-D Gaussian distribution), what distribution of maximum angle should we expect? This is equivalent, of course, to asking what random members of 3-CLUSTER look like. That is, they sought the shape of the function

$$\mathcal{K}(\alpha) = \mathbf{E}[180 - \max(a, b, c)],$$

where  $a$ ,  $b$ , and  $c$  are the three interior angles of the random triangle in degrees, and  $\mathbf{E}[\cdot]$  denotes mathematical expectation. Here as throughout this paper we give the angle of a triplet as its deviation from straight, i.e. as the complement of the largest interior angle.

Though the problem is simple to state, the true shape of this curve, depicted in Fig. 7, turns out to be somewhat counterintuitive. The distribution is naturally bounded by  $0^\circ$  and  $120^\circ$ , since we are selecting the largest interior angle, and all triangles have at least one interior angle at least as large as  $60^\circ$ . The peak of the distribution is at  $90^\circ$  (at which point there is also a change in functional form)—perhaps counter-intuitively. One intuition about why the peak is not larger is that values larger than  $90^\circ$  require triangles close to equilateral, requiring in effect close accidental alignment between the two independent angles. Another way of thinking about this is that random triplets are straighter than one might guess simply because we are always taking the straightest of three possible paths through them.

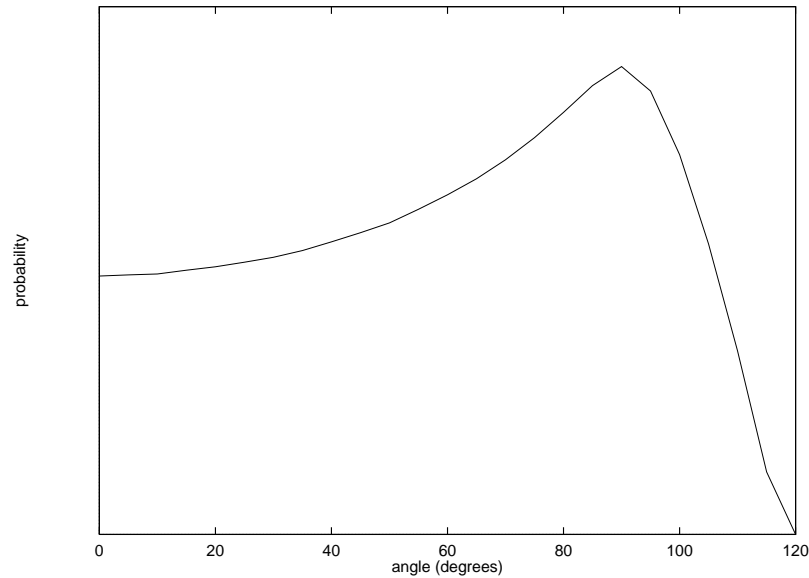


Figure 7: Kendall and Kendall’s (1980) distribution of the straightest angle of random dot triplets. Here we drew  $10^7$  triplets from a Gaussian distribution, and measured their smallest deviation from straight ( $0^\circ$ ); this value is naturally bounded by  $0^\circ$  and  $120^\circ$ . The peak is at  $90^\circ$ .

This counter-intuitiveness will have intriguing perceptual ramifications later.

(2) *Random chains of three dots.* The problem of finding a density for 3-CHAIN is somewhat different from the 3-CLUSTER case. Here, the natural way to obtain a distribution on triplets is to “random” curves from some distribution, and then generate triplets by sampling points along the curve at random intervals of arclength. For a fixed choice of arclength, points of low curvature would tend to produce nearly straight triplets, while points of high curvature would produce arbitrarily large angles. The difficulty is that the idea of a “random” curve is not well-defined; in particular there is no a priori answer to the question of how much curvature a “typical” curve point in such a curve might have.

Nevertheless, the assumption that the curves are *smooth* implies that the distribution of triplets generated from an arbitrary family of curves can be expected to be sharply peaked at  $0^\circ$ . To see why, consider triplets sampled at finer vs. coarser scales (Fig. 8). Within any

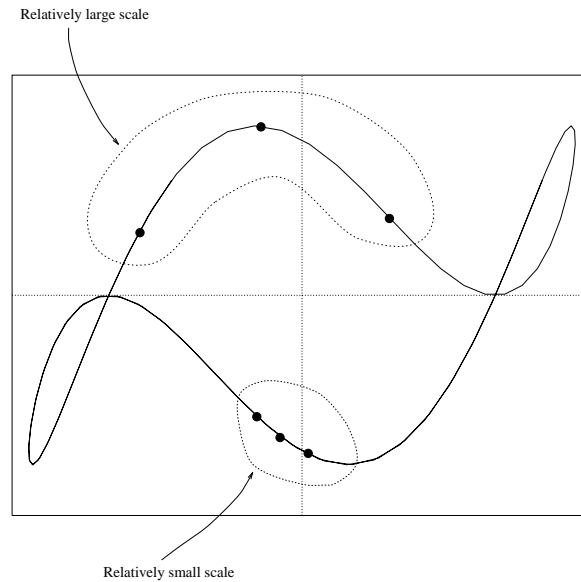


Figure 8: A smooth curve, showing how triplets sampled at a fine scale tend to produce relatively straight triplets, while triplets sampled at a large scale can produce triplets at arbitrary angles. The curve shown here is from the parametric family  $x(t) = A\cos(t) + B\sin(t), y(t) = C\cos(t) * \sin(t)^2$ , chosen to have a balance of high- and low-curvature segments. Here  $A = B = C = 1$ . The same curve family is used in the simulation shown in Fig. 9, with values of  $A, B$ , and  $C$  selected randomly on each trial.

sufficiently small neighborhood, finer scales always lead to more nearly straight triplets; in fact this is virtually the *definition* of smoothness, i.e. that as the neighborhood gets smaller the curve is better and better approximated by a local tangent. By contrast, coarser scales do *not* necessarily yield less straight triplets; this depends on the global structure of the curve, which is arbitrary. Integrating across all scales, smaller scales contribute more univocally, and hence straighter triplets tend to dominate the resulting distribution (see Feldman, 1991).

A simulation was carried out in order to corroborate this argument, and also to uncover the structure of the distribution in more detail. First, a parametric family of curves was defined, constructed so as to include regions of both high and low curvature. On each trial,

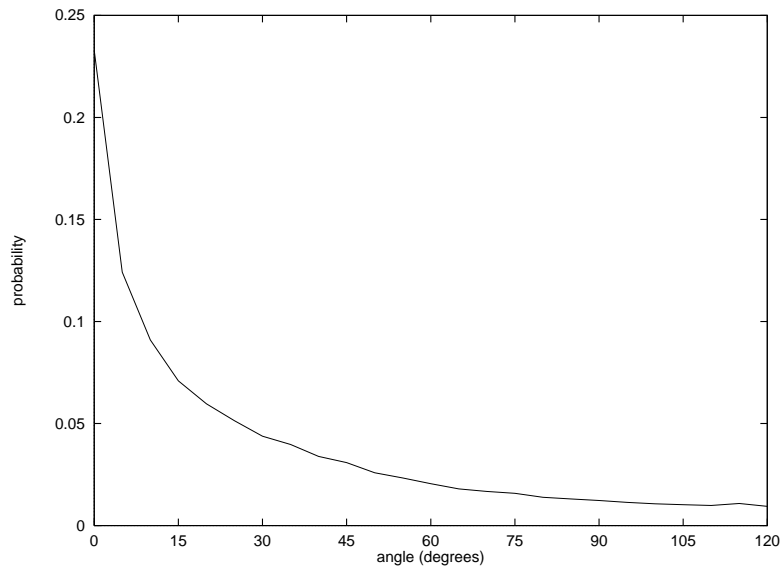


Figure 9: Distribution of 100,000 triplets generated by sampling from smooth curves from the family shown in Fig. 8. On each trial values of  $A$ ,  $B$ , and  $C$  were chosen uniformly from  $(0, 1)$ , and a scale was chosen randomly from  $(0, \frac{\pi}{3})$  (to prevent cycles).

a triplet was generated by sampling points at a randomly chosen scale from a randomly chosen member of the curve family. The results of 100,000 such trials are illustrated in Fig. 9. The curve is clearly peaked at  $0^\circ$ , and trails off monotonically at greater angles. Notice that large angles (e.g.  $120^\circ$ ), while relatively rare, do occur.

It is perhaps counterintuitive that under “reasonable” assumptions, the distribution does not vanish even at relatively large angles. Imagining triplets generated from a curve, one imagines only relatively straight triplets. That is, one imagines only triplets that one would actually classify as having been generated this way—that is, as having been generated by 3-CHAIN. But because the tail does not vanish, the class actually produces some triplets that human observers would not put into this class, but would rather classify as “random”—i.e., as members of 3-CLUSTER.

In both cases, 3-CLUSTER and 3-CHAIN, literally random members of each category turn out to be distributed in ways that the intuition finds “incorrect”—reflecting, presumably,

that our intuitions are based on category models such as the ones describe in Section 2.1, rather than on tutored statistical beliefs. Psychologically typical members of each category, by contrast, are generated in such a way as to respect the structure of the psychologically qualitative categories. The next section describes one scheme by which such densities might be generated.

### 2.2.2 Constructing probability densities for the category models

As noted above, ordinary human observers are presumably ignorant of Kendall and Kendall’s distribution of random triplets, as well as of the distribution of triplets sampled from smooth curves, and substitute some kind of heuristic expectations. The critical intuition concerns what human observers, as opposed to statisticians, mean by “random”. In this context, a statistician means “chosen without bias on the space,” like Kendall and Kendall’s triplets. Under the category theory presented above, though, what the perceptual system means by “random” is: *not regular*, or, more specifically, *not non-generic*. That is, the non-genericity  $\alpha = 0$  represents a class of triplets that have some regular structure, and which trigger an inference of an underlying curve. Psychologically random triplets, putatively, should be biased towards triplets that do *not* trigger such an inference.

Under this reasoning, it makes sense (as suggested above) to construct the class density  $p(\mathbf{X}|\text{3-CHAIN})$  by placing a peak at the straight mode  $\alpha = 0^\circ$ , and then construct  $p(\mathbf{X}|\text{3-CLUSTER})$  by placing a peak at the angle furthest in triplet space from straight. This point, the most generic point in the space, was established in Sec. 2.1.1 to be at  $\alpha = 120^\circ$ . We assume, arbitrarily but reasonably, that these two densities are each normal in shape. Notice incidentally that this assumption gives  $p(\mathbf{X}|\text{3-CHAIN})$  a qualitatively “correct” shape (mode at  $0^\circ$ , monotonically less at larger angles, but never vanishing), but an “incorrect” shape for  $p(\mathbf{X}|\text{3-CLUSTER})$  (mode at  $120^\circ$  instead of  $90^\circ$ ). If these were “real” densities, determined by the world, then the world would also give them different heights (corresponding to the scalar priors,  $p(\text{3-CHAIN})$  and  $p(\text{3-CLUSTER})$ , and different widths (standard deviations). Moreover, in the world there might be some other regularities that occur—other

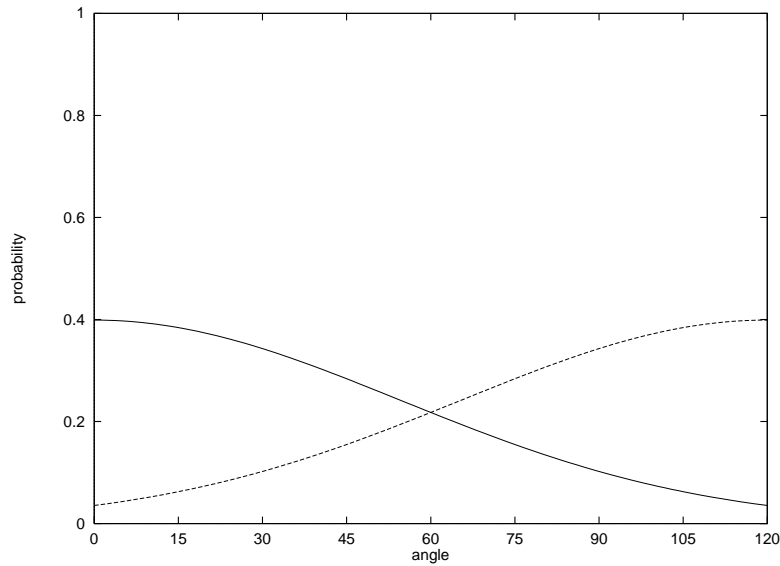


Figure 10: The two hypothetical prior “modes”, one at  $0^\circ$  (straight), the other at  $120^\circ$  (generic). Standard deviations are each  $53.059^\circ$ , the fitted value drawn from Experiment 1.

factors contributing to the density  $p(\alpha)$ . However, since these densities are heuristically constructed, the observer can unilaterally disallow all of these complicating factors, about which it has no knowledge. This stipulation corresponds to the observer imputing no additional structure to the world other than what is necessary to distinguish the two categories of interest. The resulting class densities consist entirely of two normal peaks, one centered on each regularity, each with the same height and width (Fig. 10).

How is the standard deviation of each “hump” determined? Loosely speaking, the idea is that the curve descends from a maximum at its central mode down to some “significance” level, near zero, at the location of the other mode. (This value is literally a significance level in that it denotes the probability of the observed triplet having been generated by the alternative “null” hypothesis). A theory whereby these widths can be derived will be presented in the discussion of Experiment 1, below. Loosely, the idea will be that variances are selected that capture the categorical distinction between the two classes as cleanly as possible, and support an optimally transparent posterior classification. The widths used to

draw Fig. 10 are drawn empirically from Experiment 1, but match those predicted by this scheme almost exactly.

Summarizing, we postulate an observer who has the following tacit assumptions about the dots world under observation:

- A1.** 3-CHAINS have a class density that is Gaussian with mean  $0^\circ$  and s.d.  $\sigma_c$ , and occur with probability  $p(3\text{-CHAIN})$ ;
- A2.** 3-CLUSTERS have a class density that is Gaussian with mean  $120^\circ$  and s.d.  $\sigma_r$ , and occur with probability  $p(3\text{-CLUSTER})$ ;
- A3.**  $p(3\text{-CHAIN}) = p(3\text{-CLUSTER})$ ; and
- A4.**  $\sigma_c = \sigma_r$ .

These assumptions should be viewed as *default* beliefs on the part of the observer, and certainly not infeasible; indeed it is interesting to investigate just how flexible they are in the light of evidence that contradicts them (see Experiment 2 below). The underlying hypothesis, again, is that however these assumptions might be altered so as to create more realistic priors, the human observer always persists in classifying dot patterns in a Bayesian manner. Note that the extremely bland and uncontroversial nature of these assumptions should be understood to be an advantage for the theory, in that we can generate very specific empirical predictions by ascribing to the observer a minimum of tendentious assumptions.

### 2.3 The posterior probability function

With the observer's prior assumptions about the two generating processes explicitly in place, we are now in a position to construct the a posteriori function  $p(3\text{-CHAIN}|\alpha)$ , by which observers actually assign a category label to an observed triplet. (Recall that by Eq. 1 we are free to consider only the chain case). Substituting our assumptions **A1-A4** into Bayes' rule we have

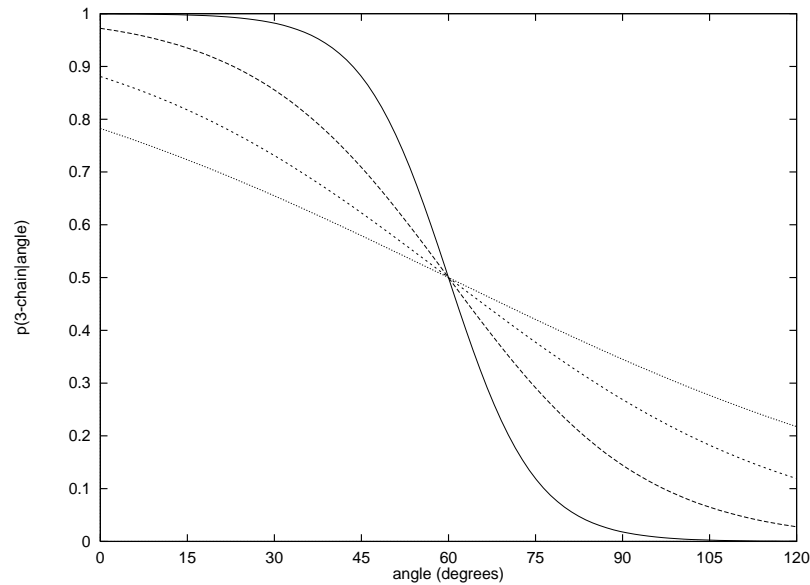


Figure 11: The hypothetical a posteriori curve  $p_\sigma$  based on **A1-A4**, shown at four different standard deviations ( $\sigma = 30^\circ, 45^\circ, 60^\circ$ , and  $75^\circ$ ).

$$p_\sigma(\text{3-CHAIN}|\alpha) = \frac{N(0^\circ, \sigma)}{N(0^\circ, \sigma) + N(120^\circ, \sigma)} \quad (4)$$

where  $N(\alpha, \sigma)$  is the normal distribution with mean  $\alpha$  and standard deviation  $\sigma$ . This equation is actually a one-parameter family of curves parameterized by  $\sigma_c (= \sigma_r)$ , the standard deviation of each “hump;” below we will refer to this family simply as  $p_\sigma$ . Notice that all the free parameters in **A1-A4** other than  $\sigma_r$ , namely the two scalar priors, have canceled out. Each curve in the family  $p_\sigma$  (four examples of which are plotted in Fig. 11) has the familiar sigmoid shape characteristic of many psychometric functions, with a transition from one category to the other whose sharpness depends on  $\sigma$ . Thus the hypothesis that observers hold **A1-A4** as default assumptions, coupled with the hypothesis that they draw category conclusions in a Bayesian manner, together amount to the claim that they will classify dot triplets according to a curve drawn from this family.

Alternatively, if subjects actually had normative statistical knowledge about random members of **3-CLUSTER**, we might in place of **A2** substitute

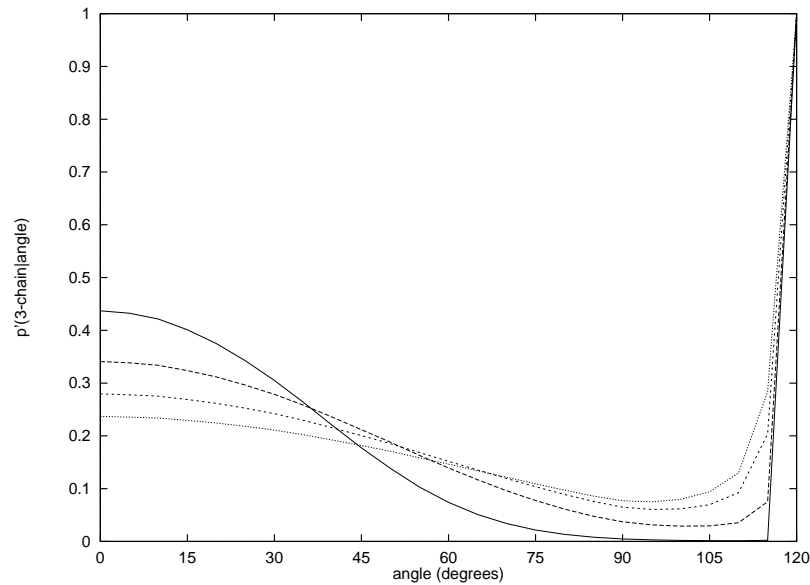


Figure 12: The alternative a posteriori curve  $p'_\sigma$  based on **A1**, **A2'** and **A3**, shown at four different standard deviations ( $\sigma = 30^\circ, 45^\circ, 60^\circ$ , and  $75^\circ$ ).

**A2'**. **3-CLUSTERS** have a class density that follows  $\mathcal{K}(\alpha)$ ,

incidentally rendering **A4** obsolete, since in this case  $p(\mathbf{X}|\mathbf{3-CLUSTER})$  is no longer determined by a standard deviation at all. (Notice that it makes no sense to replace **A1** as well, because as discussed above there is no unique choice of distribution to replace it with; moreover the resulting posterior distribution would have no degrees of freedom, which makes it too easy to reject the quasi-normative model of subjects' decisions.) With this variant of the assumptions, we now have an altered posterior function  $p'(\mathbf{3-CHAIN}|\alpha)$  following

$$p'_\sigma(\mathbf{3-CHAIN}|\alpha) = \frac{N(0^\circ, \sigma)}{N(0^\circ, \sigma) + \mathcal{K}(\alpha)}. \quad (5)$$

This is again a one-parameter family of curves parameterized by  $\sigma_c$ , which we call  $p'_\sigma$ ; examples of  $p'_\sigma$  for four values of  $\sigma_c$  are shown in Fig. 12.

All versions of this curve have a characteristic spike at  $120^\circ$ . This is *not* an artifact, but

rather reflects the fact that under these assumptions 3-CLUSTERS actually get rare much faster than do 3-CHAINS as the angle approaches  $120^\circ$ . Hence nearly-equilateral triplets are actually more likely to have been generated by the chain process than by the cluster process—a probabilistically legitimate entailment of replacing **A2** by **A2'**. The fact that this is quite counterintuitive simply reflects our intuition that the equilateral triplet is the *most* prototypical case of 3-CLUSTER—i.e. that we tacitly assume **A2** rather than the normative **A2'**.

We now report several experiment designed to directly probe human subjects' intuitions about membership in the categories 3-CHAIN and 3-CLUSTER. In Experiment 1, subjects were presented with triplets of dots, and asked to judge whether they should be classified as having been produced by a curved 1-D generating process or a 2-D patch process (via suitably comprehensible instructions, described below). Experiment 2 followed the same paradigm but varied the frequencies of triplets with various angles, in a manner described below. It should be noted that subjects were instructed that two categories of dot triplets were under consideration, so neither experiment tests the assumption that perceivers automatically assign prior probability to only these two classes; rather, the experiment is a way to reveal the exact form of the class density functions subjects unconsciously assign to the two classes under consideration, and to probe the tacit assumptions that these class densities reflect.

## 3 Experiment 1

### 3.1 Method

#### 3.1.1 Subjects.

Sixteen subjects drawn from the university community were paid for their participation.

### 3.1.2 Procedure and design

Subjects were instructed that they were to be presented with triplets of dots, to be interpreted as the markings on two equinumerous species of bottom-dwelling ocean creatures: a kind of *worm*, whose three markings were randomly placed along its curvilinear body; and an approximately round *flatfish*, whose three marking were randomly placed on its approximately round surface. The instructions explicitly drew subjects' attention to the fact that any triplet could be either type of creature: worms can curl up arbitrarily, while the dot markings on flatfish might just "happen" to fall nearly in a line. As illustration, subjects were shown two worms and two flatfishes, arranged so that exactly the same two dot triplets were given as an example of both categories.

Thus subjects understood that their task was in essence to determine how likely each dot configuration was to be a worm based on how nearly collinear the three dots were. For each triplet, subjects responded by choosing a number from 1 to 5, meaning (1) "almost definitely a flatfish", (2) "probably a flatfish", (3) "either one, with about equal probability", (4) "probably a worm", and (5) "almost definitely a worm".

After practice trials, subjects were presented with 396 such trials, crossing 33 angles (from  $-120^\circ$  to  $120^\circ$  in increments of  $7.5^\circ$ ) with 12 levels of line-length ratio (from 1.00 to 2.85 in integral powers of 1.1). Triplets were each rotated in the plane by a random angle. Dots were black circular patches 5 pixels in diameter (subtending  $0.23^\circ$  of visual angle at 35cm distance) on a white screen with high contrast. Interdot distance ranged from 100 pixels ( $4.5^\circ$ ) to 285 pixels ( $12.9^\circ$ ). Subjects were free to take as long as they wanted to respond.

### 3.1.3 Results

The results of Exp. 1 are plotted in Fig. 13. The data can be taken as representing subjects' posterior classification function,  $p(3\text{-CHAIN}|\alpha)$ . In the plot, the Y-axis is subjects' numeric response, normalized to the interval (0,1) (i.e. converted into a probability judgment). The X-axis collapses over line-length ratio and over sign of the angle. Thus each plotted point

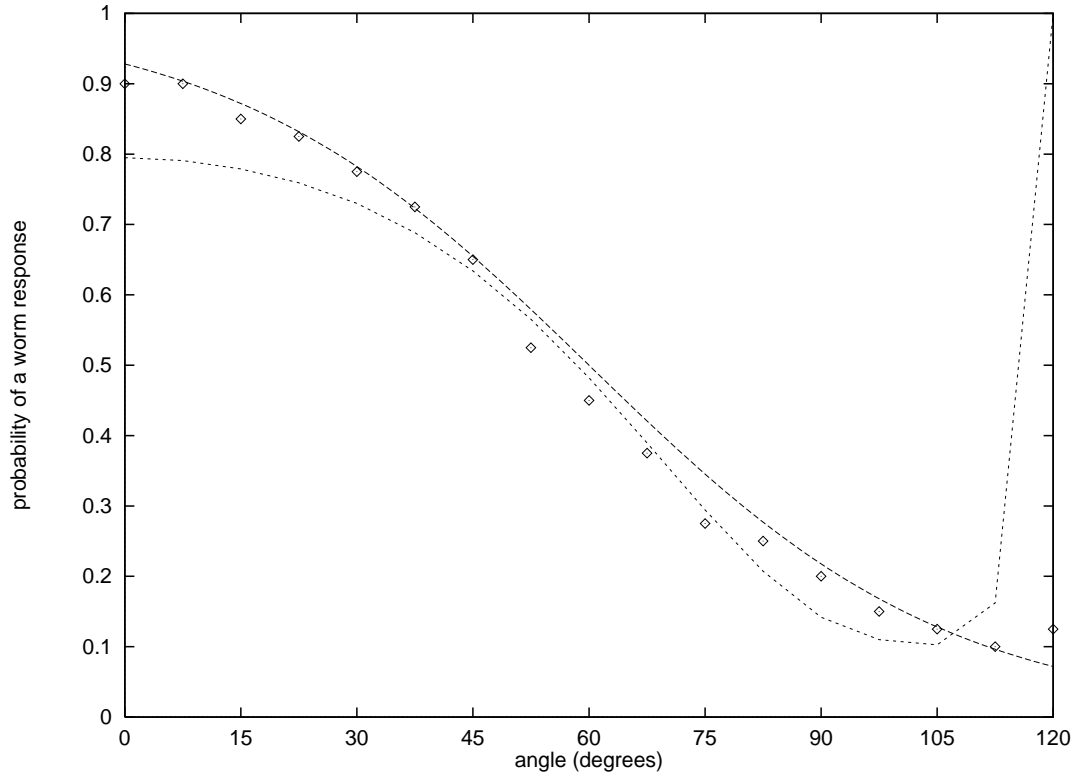


Figure 13: Results of Experiment 1. On the X-axis is the angle of the dot triplet, collapsing over line-length ratio and sign of angle. On the Y-axis are subjects' numeric responses, converted to a judgment of probability that the observed triplet was a “worm.” Also plotted are the best-fit curves from the likelihood families  $p_\sigma$  (in which  $\sigma_c = 53.059^\circ$ ) and  $p'_\sigma$  (in which  $\sigma_c = 39.686^\circ$ ), described in the text.  $p_\sigma$  clearly fits better.

corresponds to 24 observations per subject, by 16 subjects (except for  $\alpha = 0$  which is only 12 observations, since there is no sign).

Also plotted in the figure is the best fit (least-squared error) curve from each of the two families of posterior functions described above:  $p_\sigma$  (Eq. 4), corresponding to assumptions **A1-A4**, and  $p'_\sigma$  (Eq. 5), corresponding to assumptions **A1, A2'** and **A4**. As predicted,  $p_\sigma$  fits nearly perfectly ( $R^2 = 0.988$ ,  $F_{1,15} = 1208.1$ ,  $p \ll 0.01$ ). (Note that significance here simply means that each model is better than no model.) The fitted value of  $\sigma_c$  is  $53.059^\circ$  (the value used above for illustration of the two class priors). By contrast, the fit of  $p'_\sigma$  is

much poorer ( $R^2 = 0.448$ ,  $F_{1,15} = 12.2$ ,  $p < 0.01$ ). Here the fitted value of  $\sigma_c$  is  $39.686^\circ$ . As is clear from the picture,  $p_\sigma$  fits better than  $p'_\sigma$  (note that these models have the same number of degrees of freedom); unsurprisingly, the subjects do not judge nearly-equilateral triplets to be “straight” as would have been dictated by **A2**’. Note that there is no separate height parameter in either curve, making the close fits at the point-predictions on  $p_\sigma$  for  $\alpha = 0^\circ$  and  $\alpha = 120^\circ$  particularly impressive.

### 3.1.4 Discussion

It is clear from the results that subjects classified dot triplets according to  $p_\sigma$  (Eq. 4) that is, building class-conditional densities for each mode according to **A1-A4**, with standard deviations of about  $53^\circ$  for each mode, and then classifying triplets in a Bayesian manner given these priors.

Furthermore, it is clear that they do not construct the mode for “generic” triplets in anything like the normative way, as evidenced by the gross mismatch between their judgments and the curve  $p'_\sigma$  in the region near  $120^\circ$ . As discussed above, this mismatch is a not an artifact, but rather reflects the fact that subjects regarded near-equilateral triplets to be *very unlikely* outcomes of the curvilinear process, but *very likely* outcomes of the generic (patch) process—whereas mathematically such configurations are actually *extremely* unlikely outcomes of a truly random patch process, but only *mildly* unlikely outcomes of the curvilinear process. (Notice that this is true whether or not the 3-CHAIN mode is normal or is closer to the distribution of Fig. 9.) Putting it another way, near-equilateral triplets are possible though extremely unrepresentative outcomes of the curvilinear process; but to subjects, given their mental model of a generic configuration, such triplets are the very prototype of the patch process.

It is worth remarking on the fact that subjects’ induced density functions (Fig. 10) bear a plain resemblance to the canonical signal detection set-up. In both cases, a single discriminant parameter is inhabited by two Gaussian peaks of equal variance. In effect the subjects have fabricated two class densities exactly as if to simulate signal and noise, with the triplets

generated by the generic category 3-CLUSTER serving as noise, and  $d' = 2.2599 (= 1/0.4435)$ . In a sense this is perfectly consistent with the idea that the generic category serves as a random backdrop or distractor to the regular (curvilinear) target category 3-CHAIN. The difference is an epistemic rather than a technical one. There is no extrinsic noise source in the dot triplet space; the variance in both categorical densities is a subjective invention on the part of subjects, rather than anything like literal uncertainty in the physical detection of a signal. The probabilities, that is, are subjective ones; the “signal” is a concept of order or regularity in an abstract hypothesis space, and the “noise” is a concept of disorder, irregularity, and causal independence in the generation of the dots.

In summary, subjects were in effect placed into a novel, imaginary domain of creatures of which they had no prior knowledge whatsoever. They were thus left to create class-conditional density functions wholly out of their own intuitions. They apparently did so by appealing to the kind of regularity- and genericity-based apparatus described in Section 2.1. This apparatus thus accounts for the off-angle response function in the minimal case in the process of grouping dots.

*Accounting for the distribution widths.* To complete the account of these data, we would like to be able to say something about the sole remaining fitted parameter,  $\sigma$ , the standard deviation of each of the two prior modes. What value would we have predicted for this parameter?

One prediction for the value of  $\sigma$  can be derived from the premise that the observer is attempting to construct priors so that the ensuing classification will be as simple as possible to calculate. To do this, the observer simply has to assume that all triplet configurations are actually equiprobable in the world; that is,

$$\mathbf{A5.} \quad p(x) \equiv 1.$$

Under this assumption, the denominator (Eq. 3) in the a posteriori equation (Eq. 2) drops out (notice that we can ignore the two constant terms by virtue of **A3**). This reduces

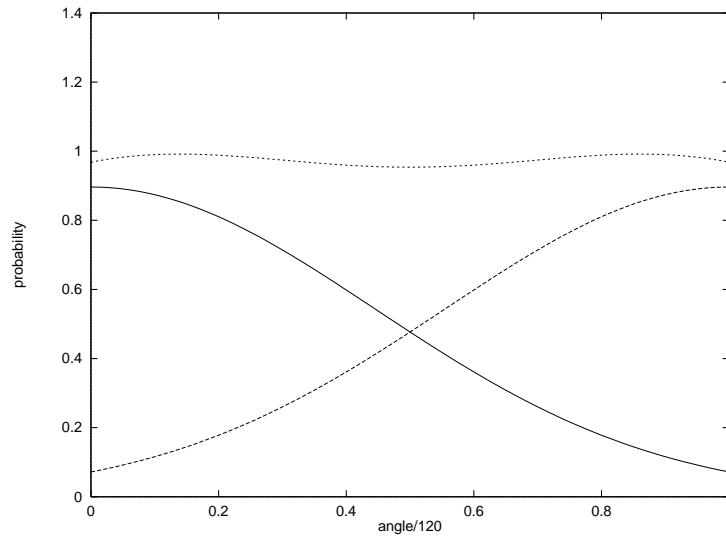


Figure 14: The two class densities sum identically to about unity, meaning that all triplet angles are about equally likely, and prior and posterior probabilities are as nearly identical as possible.

the inference to a simple maximum likelihood calculation: infer that generating class under which the observed triplet was the most probable. That is, given an observed triplet with angle  $\alpha$ , the classification probability  $p(\text{3-CHAIN}|\alpha)$  can be read directly off the known *prior* probability density  $p(\alpha|\text{3-CHAIN})$ . Pictorially, **A5** holds when the two triplet modes or “humps” summed together produce a flat horizontal line of height unity (Fig. 14).

Hence in order to derive a prediction for the standard deviation of the two humps, we seek the value of  $\sigma$  for which the function  $f(x) = N(0^\circ, \sigma) + N(120^\circ, \sigma)$ , in the interval  $[0^\circ, 120^\circ]$ , deviates as little as possible (in a least-squares sense) from the line segment  $f(x) = 1$ . That is, we seek a value of  $\sigma$  that as nearly as possible satisfies the equation

$$\frac{d}{d\sigma} \left[ \int_0^1 \left[ e^{-\frac{x^2}{2\sigma^2}} + e^{-\frac{(x-120)^2}{2\sigma^2}} - 1 \right]^2 dx \right] = 0. \quad (6)$$

An analytic solution to this minimization is problematic because of the Gaussian terms, but numeric estimation techniques provide a figure of  $\sigma \approx 0.4435(120^\circ - 0^\circ) = 53.225^\circ$ , or only about 0.31% higher than the empirical figure of  $53.059^\circ$ . Another way of expressing the

closeness of this fit is to plug this value back into a model of the subjects' data: the model now has 0 degrees of freedom (a point prediction) but accounts for 98.7% of the variance (and has infinite  $F$ -ratio).

It is worth heading off an objection that these results simply reduce to the unsurprising discovery that the classification curve is Gaussian. In fact the curve is not a Gaussian, but rather is constructed from Gaussian priors in such a way as to make it resemble one (closely, but not precisely).

## 4 Experiment 2

Subjects in Experiment 1 apparently chose a standard deviation by assuming that the two modes sum identically to unity; that is, by assuming that  $p(\alpha)$  is flat. Notice, though, that this was actually *true* in the experiment—that is, all angles were equally likely.<sup>3</sup> Hence solely on the basis of Experiment 1 it is impossible to determine the precise nature of subjects' assumption **A5** that all triplets were equally likely, and how the classification process might be affected by alteration of this assumption.

Experiment 2 seeks to remedy this situation by varying the frequency of triplets of various angles, presumably biasing subjects to change their beliefs about the relative proportions of 3-CHAINS and 3-CLUSTERS in the environment (defeating **A3**). Two scripts were employed. In one, frequency increased with angle so that equilateral triplets were twice as likely as straight triplets; in the other, frequency decreased with angle in a like manner.

### 4.1 Method

#### 4.1.1 Subjects

20 subjects were tested, 10 with each of the two scripts. None of these subjects had participated in Experiment 1.

---

<sup>3</sup>Note though that it is meaningless to say that the two modes summed to unity in the experiment *per se*, since the modes exist only in subjects' heads, not in the stimuli.

#### 4.1.2 Procedure and design.

The procedure and instructions were identical to Experiment 1; only the scripts varied. In Script 1, there were 8 triplets at  $0^\circ$ , and then 1 additional triplet for each  $15^\circ$  increment away from  $0^\circ$  in both positive and negative directions, to a maximum of 16 triplets at both  $120^\circ$  and  $-120^\circ$ . This scheme was crossed with two levels of inter-dot distance: equal, and unequal with a ratio chosen randomly from 1 to 2. This amounts to a total of 416 trials. Script 2 followed the same scheme except with frequency decreasing from 16 at  $0^\circ$  to 8 at  $120^\circ$  and  $-120^\circ$ , again multiplied by two levels of inter-dot distances; this amounts to a total of 400 trials (less than Script 1 because the most numerous case there occurs twice ( $120^\circ$  and  $-120^\circ$ ), while in Script 2 the most numerous case ( $0^\circ$ ) only occurs once).

#### 4.1.3 Results

The results for both scripts are displayed in Fig. 15. Along each set of data, the best-fit curve is plotted from the family of posterior curves in which the relative height of the classes  $h$  ( $= p(\text{3-CLUSTER})/p(\text{3-CHAIN})$ ) is free to vary:

$$p(\text{3-CHAIN}|\alpha) = \frac{N(0^\circ, \sigma)}{N(0^\circ, \sigma) + hN(120^\circ, \sigma)},$$

but in which  $\sigma$  is fixed at the empirical value of  $53.059^\circ$  established in Experiment 1.

This model fit both sets of data extremely well. In the rising case (Script 1), the curve fit with  $R^2 = 0.999$ ,  $F_{1,7} = 12,930$ ,  $p \ll 0.01$ , with  $h$  fit to 0.478. In the falling case (Script 2),  $R^2 = 0.992$ ,  $F_{1,7} = 891.904$ ,  $p \ll 0.01$ , with  $h$  fit to 0.311.

#### 4.1.4 Discussion

The results are consistent with the hypothesis that subjects classify triplets in a Bayesian fashion, having constructed suitable priors in the manner described. Under the influence of these new scripts, subjects have simply varied their estimate of the relative frequency of 3-CLUSTERS and 3-CHAINS in their environment. That is, they have constructed the two class densities exactly as before—contraposing generic and non-generic categories—and

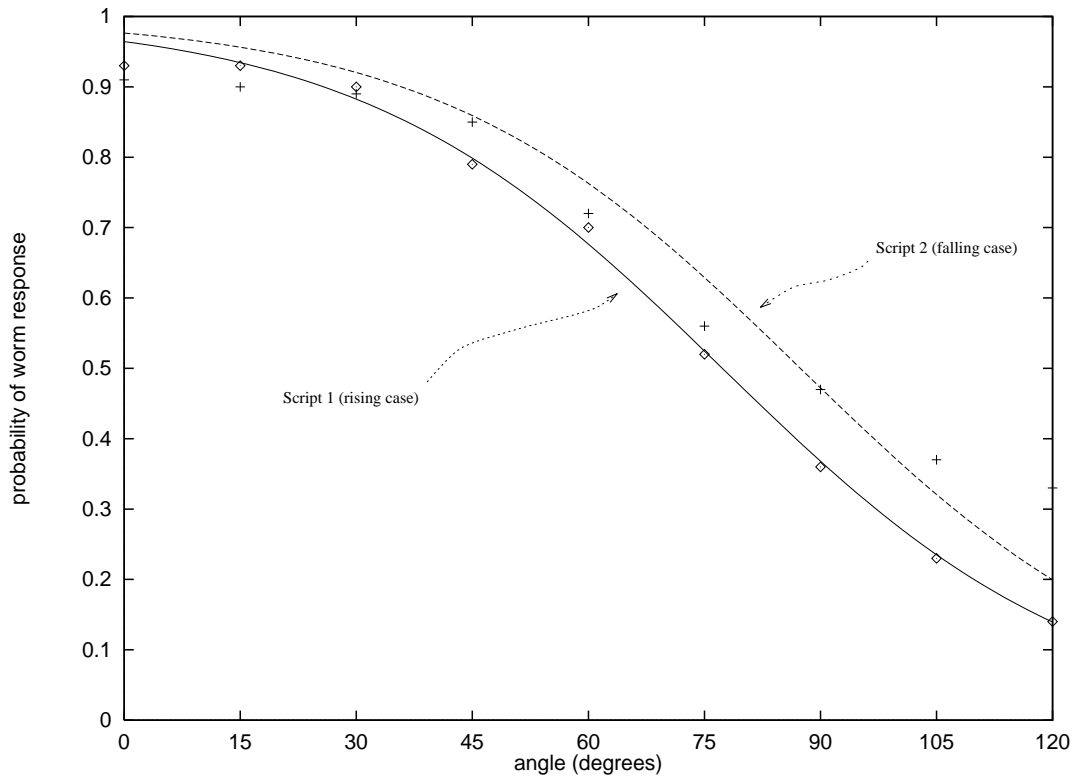


Figure 15: Results from Experiment 2, plotted as in Fig. 13, including both Script 1 (circles) and Script 2 (pluses). For each case, also plotted is the best-fit models from the posterior family parameterized by  $h$  as described in the text.

then simply pushed around the height of the generic “bin” as compared to the non-generic (curvilinear) bin in order to better reflect the conspicuously non-uniform environment in which they found themselves situated. Oddly, they seem to have estimated the former to be more probable in both conditions; hence while they move bins around in reaction to the environment, they do not seem to do so in an accurate manner. They varied an existing parameter in the posterior calculation, but otherwise preserve unchanged the basic structure of the categorization uncovered in the previous experiment.

## 5 General discussion of experiments

The experiments paint a hybrid picture of human observers’ models of dot clusters, in which posterior inferences are drawn using distinctly Bayesian machinery, and yet in which critical antecedents such as the class-conditional densities are virtually fabricated out of whole cloth. Subjects’ responses in both experiments clearly demonstrate that they classify dot triplets in an approximately normative manner, *given* that they construct category densities in the distinctly extra-normative manner outlined above: namely in which generic and non-generic categories are placed at extreme opposite ends of triplet space, the “random” category being centered around the *maximally generic* point in the parameter space, as per the category lattice scheme. In the simplest terms what this means is that subjects seem to make a particular set of tacit assumptions about the regularities governing the dots world. They then use these regularity notions to structure their beliefs about what is likely and what unlikely given various hypotheses about the state of the world.

Because the subjects’ responses are at odds with the normative distributions within each class, it is tempting to conclude from this they are simply *inaccurate* in their attempts to guess what typical members of each class, selected randomly, would look like. It makes more sense, though, to point out that observers are following a useful strategy. They are in effect theorizing about what basic forms regularities in this unknown domain might take. In effect they are assuming, quite plausibly, that observed objects in the new domain are *not* generated completely randomly, but rather are subject to some causal forces that, while

mysterious in exact nature, are coarsely predictable in form.

## 6 Composing molecular clusters into larger groups

With the exact class-conditional densities for molecular clusters in hand, an observer is theoretically in a position to determine likelihoods for arbitrarily large sets of dots, by composing molecular decisions into larger groups. In this manner a psychologically correct decomposition of dot configurations such as that in Fig. 1 could be computed in a Bayesian way. We now briefly consider how this might be done, and show that the normative way of doing so does not work: the human intuitions are, again, somewhat more complicated.

Consider a chain of four dots  $\{\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d\}$  defined by two angles  $\alpha_1$  and  $\alpha_2$  (Fig. 16), again ignoring relative inter-dot distances. We ought to be able to treat the two triplets  $\{\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c\}$  and  $\{\mathbf{x}_b, \mathbf{x}_c, \mathbf{x}_d\}$  as separate trials, computing a likelihood  $p(\text{3-CHAIN}|\alpha_i)$  for each triplet separately. The key question concerns the independence of these two “trials”. If the two triplets did not originate from a common process, the probabilities that each one is actually a 3-CHAIN ought to be independent. On the other hand, if the four in a row were all generated by a common curvilinear process, then we would not expect their class labels to be independent. Rather, in a world in which curvilinear regularities exist, very straight triplets might be expected to be commonly followed by additional straight triplets. Hence whereas in the 3-dot case subjects’ beliefs about regularities were probed by measuring means and variances, the corresponding questions in the 4-dot and larger cases need to be probed by measuring means and covariances (work that is in progress). Regularity in the world then corresponds literally, not surprisingly, to covariance among image features.

If the two constituent triplets in the ordered four-tuplet are treated as independent events, the likelihood of the overall configuration under a chain interpretation would be

$$p(\mathbf{X}^4|\text{curvilinear process}) = p(\alpha_1|\text{3-CHAIN})p(\alpha_2|\text{3-CHAIN}). \quad (7)$$

Substituting Gaussian densities for each triplet density, this become approximately

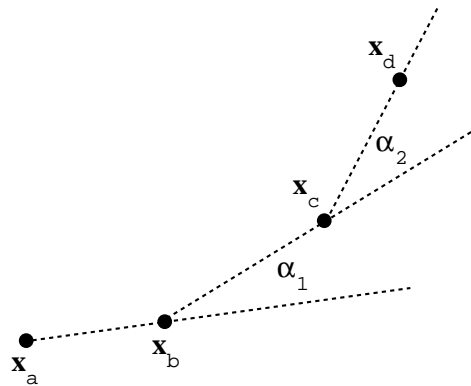


Figure 16: Parameterization for four dots—a multi-molecular cluster.

$$p(\mathbf{X}^4 | \text{curvilinear process}) \approx \frac{1}{\sqrt{2\pi}} e^{[-\frac{1}{2\sigma^2}(\alpha_1^2 + \alpha_2^2)]}. \quad (8)$$

This equation ought to tell us the power of a given chain extraction to explain a given set of dots. Assuming completely neutral priors (chains and clusters equally likely, and groups of all sizes equally likely: cf the celebrated “Principle of Indifference”), this allows us to compute a maximum a posteriori interpretation for each field of dots. This interpretation in principle represents the inferred generating process for the observed dots that, among all models under consideration, is actually the most likely to be the correct.

For many dot configurations, the maximum a posteriori interpretation does in fact result in the clustering that human observers extract. For example, consider the field of eight dots shown in Fig. 17. The groups discovered by the computer are indicated: dots belonging to a chain are linked by line segments in sequence, and dots belonging to a cluster are each linked to a virtual dot (drawn as an unfilled circle) at the centroid of the cluster.

Here, the computer’s interpretation is the same as the intuitive one (compare Fig. 1). It is not difficult to find examples, however, in which human intuitions conflict with the computed solution, for interesting reasons. An example with six dots illustrates the problem (Fig. 18). Again, the chains found by the program are linked up by line segments.

Most human observers probably see the configuration as one long but very coherent

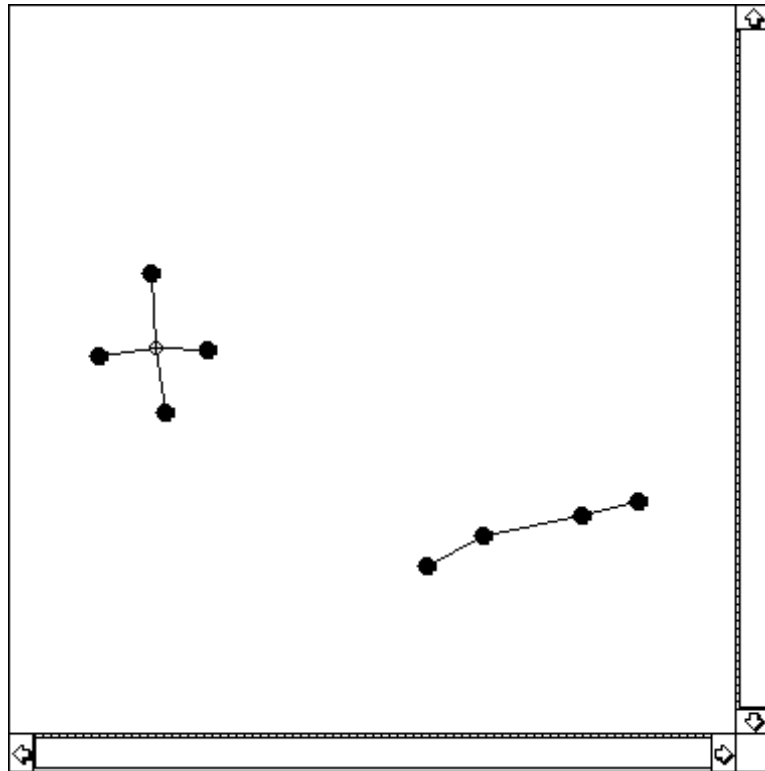


Figure 17: Using a maximum a posteriori grouping procedure as described in the text, the computer finds two groups: an unordered cluster of four dots on the left (drawn linked to their centroid) and a chain of four dots on the right (drawn with links from dot to dot).

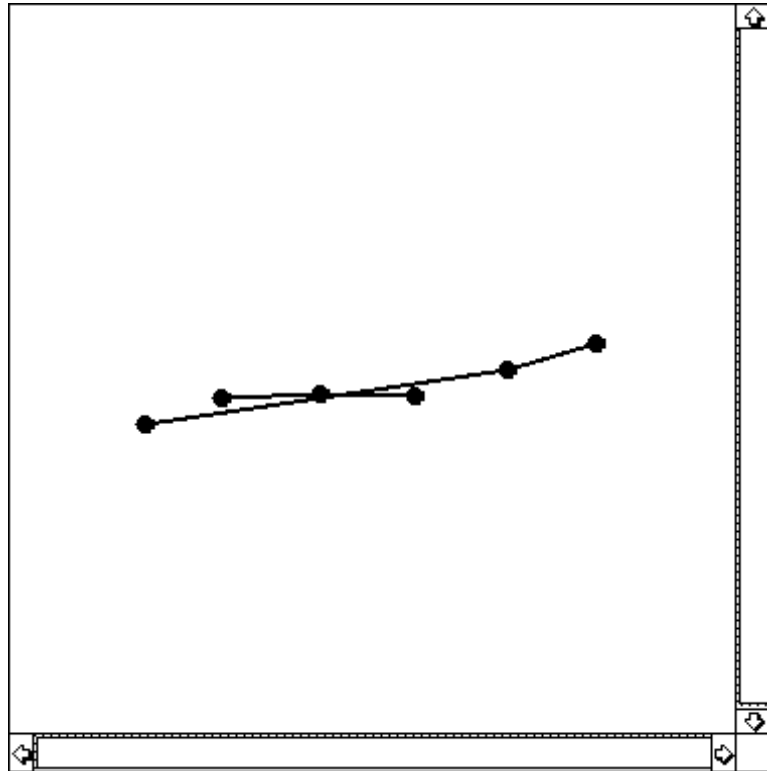


Figure 18: The maximum a posteriori clustering of these six dots (given the assumptions in the text) is as two distinct but intertwined chains, not the single long chain that human observers prefer.

chain of six dots. The decomposition discovered by the program, however, separates the dots into two distinct intertwined triplets. That is, each of these two triplets is internally so nearly collinear, and thus accounts for its subset of the data so completely, that the combination beats out the more curvaceous but also more intuitive six-dot chain.

What strikes the observer as implausible about this decomposition, perhaps, is that the two triplets are not in a generic configuration with respect to each other—they are nearly parallel and nearly coincident. This configuration is highly special, and highly unlikely if the two triplets are independent; but of course it is actually to be expected if all six dots actually originated from one long curvilinear process. In order to render the cluster composition fully competent to parse fields of dots, then, the genericity idea must be extended up to slightly more abstract level: the *relative* configuration among sub-decompositions. Just as three dots within a perceived cluster must be in general position with respect to each other (given the inferred category), the various clusters (cells) in a hypothetical dot decomposition must all be in general pose with respect to one another, in some suitably defined sense. A full dot decomposition theory, correctly accounting for observers' decompositions of arbitrary fields of dots, requires such a definition; again see Feldman, 1995.

## 7 Conclusion: causal models of dot groupings?

The dot interpretation machinery presented in this paper may seem to focus a large amount of apparatus on a very simple domain. The data presented above, however, make it quite clear that human observers' intuitions about this domain contain a considerable amount of structure not given by the problem. The triplet classifications, for example, reveal a nest of tacit assumptions about what ought to be seen as "typical" in the dots world: the centers and widths of the class-conditional densities are drawn from the head, not from the world, and do seem to be built around the categorical structure described here. Even the reader's perception of a dot chain in Fig. 1 reveals the reader's tacit belief in the existence of "regularities" in the dots world. Perhaps counter-intuitively, if there were no curvilinear dot-generating processes in the domain, then *any* perceived chain, no matter how compelling,

would be a false target—an accidental alignment. The fact that one does, automatically, perceive such chain groupings suggests that our perceptual system in some sense believes that causal forces in the environment will, with some non-zero prior probability, cause dots to be generated along curves (again see Richards, Jepson, and Feldman, 1993 on this point).

Theorists of clustering per se, in evaluating the utility of novel clustering techniques, are often necessarily guided by their intuitions about what clustering solutions make “sense” and thus should be recovered by the technique. At the most abstract level, the kind of formalism described in this paper can be seen as shedding light on just in what respect some clustering intuitions make more sense than others, and on the logical structure of the inference mechanism that underlies such intuitions.

#### Acknowledgments

I am very grateful to Whitman Richards, Allan Jepson, and several anonymous reviewers for many helpful discussions of the ideas in this paper, and to Janice Taylor and Sheryl Maniar for running the subjects. This research was made possible by the Center for Cognitive Science at Rutgers University, New Brunswick, NJ.

#### References

- Ashby, F.G. and Gott, R.E. 1988 Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14** 33-53
- Bennett, B., Hoffman, D. and Prakash, C. 1989 *Observer mechanics: A formal theory of perception* (London: Academic Press)
- Broadbent, S.R. 1980 Simulating the ley hunter. *Journal of the Royal Statistical Society A* **143** Part 2, 109-140
- Brookes, A. and Stevens, K.A. 1991 Symbolic grouping versus simple cell models. *Biological Cybernetics* **65** 375-380.

Caelli and Julesz, 1978 On perceptual analyzers underlying visual texture discrimination: Part I. *Biological Cybernetics* **28** 167-175.

Dodwell, P.C. 1983 The Lie Transformation Group model of visual perception. *Perception and Psychophysics* **34** 1-16.

Do Carmo, M 1976 *Differential geometry of curves and surfaces*. (Englewood Cliffs, N.J.:Prentice-Hall)

Duda, R. and Hart, P. 1973 *Pattern classification and scene analysis* (New York: John Wiley)

Feldman, J. 1991 Perceptual decomposition as inference: continuous curvilinear processes. M.I.T. Cognitive Science Center Occasional Paper #42

Feldman, J. 1992a Constructing perceptual categories. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Conference, Champagne-Urbana, Illinois, June 1992

Feldman, J. 1992b Perceptual categories and world regularities. Ph.D. dissertation, Department of Brain & Cognitive Sciences, M.I.T.

Feldman, J. 1995 Regularity-based perceptual grouping. Technical Report, Center for Cognitive Science, Rutgers University.

Foster, D.H. 1984 Local and global computational factors in visual pattern recognition. In *Figural Synthesis*, P.C. Dodwell and T. Caelli (Eds.). Hillsdale, N.J.: Erlbaum, 1984.

Glass, L. 1969 Moiré effects from random dots. *Nature*, **223** 578-580.

Hoffman, W.C. 1966 The lie algebra of visual perception. *Journal of Mathematical Psychology* **3** 65-98.

Jepson, A. and Richards, W. 1993a What is a percept? University of Toronto, Dept. of Computer Science Tech Report RBCV-TR-93-43. (Also MIT Cognitive Science Memo 43, 1991.)

Jepson, A. and Richards, W. 1993b What makes a good feature? *Spatial Vision in Humans*

- and Robots*, L.Harris and M. Jenkin (Eds). Cambridge:Cambridge University Press. See also M.I.T. A.I. Lab Memo 1356 (1992)
- Kendall, D.G. and Kendall, W.S. 1980 Alignments in two-dimensional random sets of points. *Advances in Applied Probability* **12** 380-424
- Leyton, M. 1984 Perceptual organization as nested control. *Biological Cybernetics* **51** 141-153
- Leyton, M. 1989 Inferring causal history from shape. *Cognitive Science* **13** 357-387
- Matthews, G. and Hearne, J. 1991 Clustering without a metric. *IEEE Transactions on pattern analysis and machine intelligence* **13** 175-184
- Nosofsky, R.M. 1984 Choice, similarity and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **10** 104-114
- Nosofsky, R.M. 1991 Typicality in logically defined categories: exemplar-similarity vs. rule instantiation. *Memory and Cognition* **19(2)** 131-150
- Prazdny, K. 1984 On the perception of Glass patterns. *Perception* **13**:469-478.
- Richards, W., Jepson, A., and Feldman, J. 1993 Priors, preferences, and categorical percepts. *Proceedings of the Workshop on Psychophysics and Computation*, Chatham, MA, January 1993, forthcoming.
- Späth, H. 1980 *Cluster analysis algorithms for data reduction and classification of objects* (New York:John Wiley)
- Stevens, K.A. 1978 Computation of locally parallel structure. *Biological Cybernetics* **29** 19-28
- Stevens, K. A. and Brookes, A. 1987 Detecting structure by symbolic constructions on tokens. *Computer vision, graphics, and image processing* **37** 238-260.
- Witkin, A. and Tenenbaum, J. 1983 On the role of structure in vision. In H. Rosenfeld (Ed.), *Human and machine vision* (New York: Academic Press)

Zucker, S.W. 1984 Two constraints on early orientation selection in dot patterns. In *Figural Synthesis*, P.C. Dodwell and T. Caelli (Eds.). Hillsdale, N.J.: Erlbaum, 1984.

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