DIMACS Series in Discrete Mathematics
and Theoretical Computer Science
Volume 19, 1995

PERCEPTUAL MODELS OF SMALL DOT CLUSTERS

JACOB FELDMAN

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 ("molecular") 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 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 use internal models of each generating process to construct prior distributions for the shape of both random and "ordered" triplets. 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 direction inspection (on the other) is an appealing one—and yet the two process are peculiarly different. Consider, for example, Fig. 1 (below). To a human observer (such as the reader), 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

1991 *Mathematics Subject Classification.* Primary 92J30

I am grateful to Whitman Richards and Allan Jepson for many helpful discussions of the ideas in this paper, and to an anonymous reviewer for comments on an earlier draft; and to Janice Taylor for running the subjects and coding their responses. A final version of this paper will be submitted for publication elsewhere. This research was supported by the Rutgers University Center for Cognitive Science.

must be supplied to the algorithm (see [4] or [15] 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 [13], 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 [10]):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, can be seen as a minimal kind of categorical inference, in that (as shall be argued below) 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 [9] (cf [7]) call a "regularity".

Describing the human inference in this domain turns turns critically on the exact nature of human perceivers' *prior probabilities*—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 a 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, deferring a more complete theory of this overall inference to a longer future paper. For the three dots case, it will turn out that the human subjects draw inferences in a recognizably Bayesian (though unorthodox) way. It will be argue that the exact assumptions prior they make in order to do so, as suggested above, 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 first present some quasi-

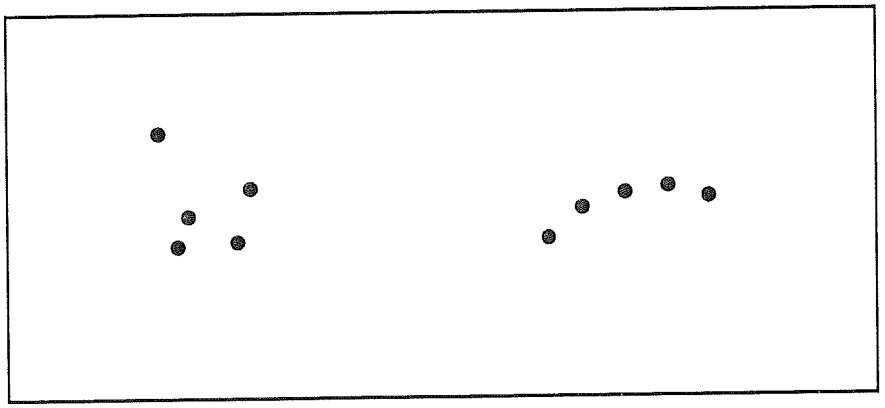


FIGURE 1. Two clusters, of different perceptual "types".

normative facts about dot densities due to the statisticians Kendall and Kendall [10] These densities, of which subjects are presumably completely unaware (they are in fact counterintuitive in several respects) serve as kind of neutral backdrop for observers' naive hypotheses. Section 2.2 lays out these basic statistical "facts." In grouping dots, as will be discussed below, observers need values need to know how just 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 topic of the next section.

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 the most typical point 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 the observer 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 kind of "kind world" assumption (cf [1]).

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 prin-

ciples 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, in particular Glass patterns, has been approached from a number of directions since. Some (e.g. [18]) have modeled it using low-level orientation-tuned operators similar to the type believed to operate in visual cortex. 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. Other work [16] has asked how curvilinear groups of dots can be efficiently computed, arguing that perceptual phenomena can be seen as emerging from the heuristics the visual system implicitly employs. Witkin and Tenenbaum [17] have argued that this sort of low-level structure and regularity in the image serves as an underpinning for a "meaningful" representation of causal forces at work in the world, a view more in sympathy with the model developed below.

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 can be generated uniformly (like a Poisson process); and 1-D curves along which dots are generated at random intervals of arclength[1] (see Fig. 4). The curvilinear process may be thought of as a kind of "regularity" in this domain (see [6] 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 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 $\mathrm{x} \in R^2$, and an observed set of N dots $\{\mathrm{x}_1 \ldots \mathrm{x}_N\}$ by X^N, omitting the N whenever unambiguous. 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, $p(C_r|\mathbf{X})$.

[1]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.

FIGURE 2. "Typical" examples of (a) a 3-CHAIN and (b) a 3-CLUSTER

Notice that it is only because of 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 under these operations, we take triplets as a minimal set. Since, in this sense, a triplet is the smallest set of points 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. $\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}$, etc. Note that for the moment 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. 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. "Typical" examples are depicted in Fig. 2). 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 idea explicit.) The two kinds of generating process and their expected outcomes are depicted schematically in Fig. 3.

For most of the remainder of the paper we will be concerned with the two quantities $p(\text{3-CLUSTER}|\mathbf{X})$ and $p(\text{3-CHAIN}|\mathbf{X})$. We assume that these two interpretations 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. 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?

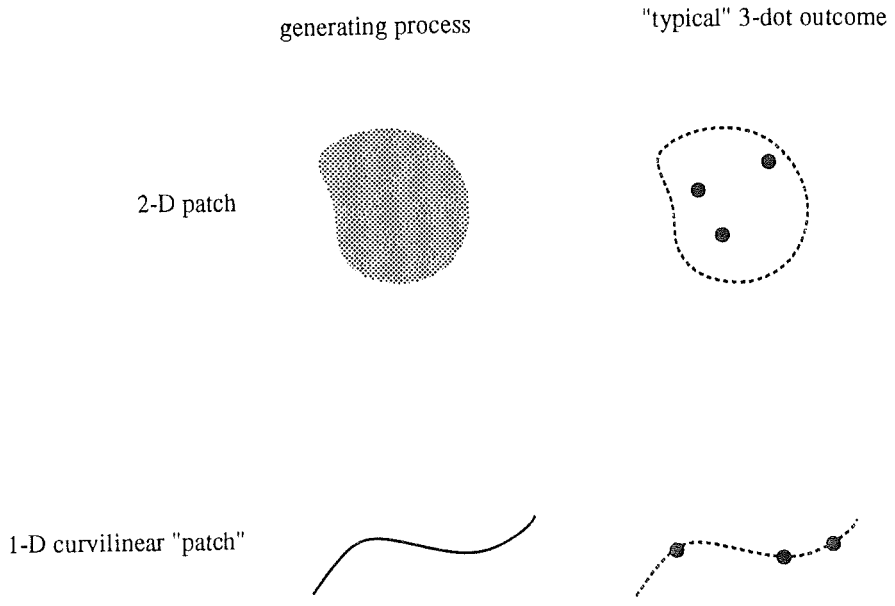


FIGURE 3. A schematic depiction of the two generating processes or "regularities" under consideration, showing the form of the underlying generating process, and a "typical" outcome.

Notice that we are implicitly defining a 3-CLUSTER only by default with respect to a 3-CHAIN: a 3-CLUSTER is a triplet of dots *not* straight enough to be considered a 3-CHAIN. That is, $p(\text{3-CLUSTER}|\mathbf{X}) = 1 - p(\text{3-CHAIN}|\mathbf{X})$ by stipulation. The structure in the observer's category decision will come instead from the two class densities $p(\mathbf{X}|\text{3-CLUSTER})$ and $p(\mathbf{X}|\text{3-CHAIN})$. Various theories about the proper structure of these densities will be take up in detail in the next section. Once priors have been determined, and assuming the two categories themselves are mutually exclusive[2] and jointly exhaustive, Bayes dictates that the posterior probability of the 3-chain category should be calculated as

$$p(\text{3-CHAIN}|\mathbf{X}) = \frac{p(\text{3-CHAIN})p(\mathbf{X}|\text{3-CHAIN})}{p(\mathbf{X})} \tag{1.1}$$

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

$$p(\mathbf{X}) = p(\text{3-CHAIN})p(\mathbf{X}|\text{3-CHAIN}) + p(\text{3-CLUSTER})p(\mathbf{X}|\text{3-CLUSTER}) \tag{1.2}$$

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 draw their category judgments in this normative manner, though, and what values they give

[2]Care must taken on this point, since the two distributions will in fact overlap. What we mean by "mutually exclusive" is that both category labels cannot be applied *correctly* to the same observed triplet.

to the two scalar priors, are largely empirical questions, which will be taken up as such in Section 3.

2. Statistical and perceptual models of dot-producing processes

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 ([11],[12]) 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 [17]).

In earlier work ([6],[7]) 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 [8] and [9]). 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.

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 [14]). 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 [6] 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 curvilinear process is our model of how chain clusters come about; and, critically, is the process whose presence in the environment means that the inference of a "chain" is actually 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 concepts of regularity and categorization that were originally conceived to apply to a wider range of perceptual inferences.

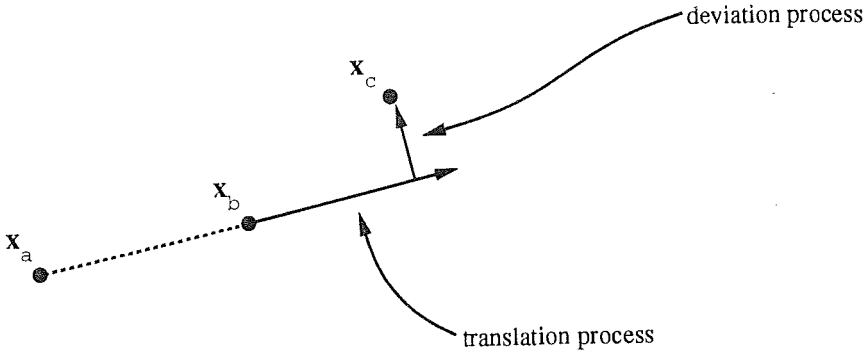


FIGURE 4. The "generative model" for dot triplets. A "translation process" carries x_c out from x_b, along the line joining x_a to x_b. A transverse "deviation" process carries x_c away from this line.

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.

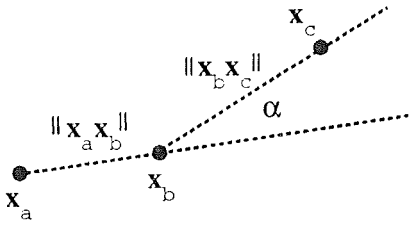
2.1. A generative model of dot triplets. Following [7], we now construct a generative model for dot triplets, which will serve to determine a parameterization for dot triplets. This in turn will determine a category lattice formally expressing the admittedly simple relationship between the categories 3-CLUSTER and 3-CHAIN.

We begin with a reference dot, x_b (Fig. 4). We imagine a "translation process" that positions a dot x_c along some translation vector $\vec{t}$ away from x_b, $x_c = x_b + \beta_1 \vec{t}$ for some scalar β_1. A third dot, x_a, is required to determine a reference direction for $\vec{t}$, i.e. $\vec{t} \sim x_b - x_a$. Thus we can think of an entire space of dot triplets that are spanned by the this translation process alone, consisting of the exactly collinear triplets in which $x_c - x_b = \beta_1 \vec{t}$.

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 $x_a x_b$ (so that $\vec{d} \perp \vec{t}$). Critically, all dot triplets can be expressed as a weighted sum of these two processes, i.e. any triplet can be written as $\{x_a, x_b, x_b + \beta_1 \vec{t} + \beta_2 \vec{d}\}$ for some scalars β_1 and β_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.[3]

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. 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 subspace.

[3]Readers may also note the relevance of the "fundamental theorem of curves" (see [3]) 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 x_b.



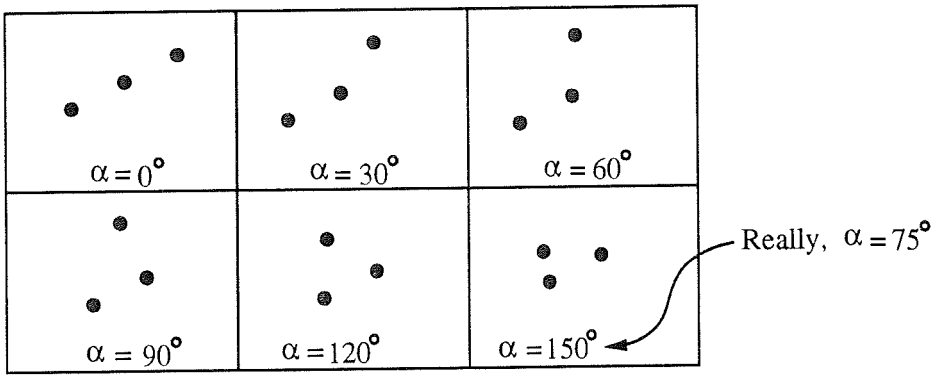


FIGURE 5. We can parameterize any triplet, up to similarity, by α (the angular deviation from straight) and $\frac{\|x_b x_c\|}{\|x_a x_b\|}$, the ratio of the length of the second "leg" to the length of the first leg. Below, examples are given of dot triplets with α at various values (and $\frac{\|x_b x_c\|}{\|x_a x_b\|} = 1$). Notice that α cannot exceed 120°, since it is always measured as the angle nearest to straight (illustrated in the bottom right example).

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, β_1 and β_2, for which we substitute the more meaningful (but diffeomorphic) angle α and line-length ratio ($\frac{\|x_b x_c\|}{\|x_a x_b\|}$). This natural parameterization of the space by leg-length ratio ($\sim \beta_1$) and angle $\alpha(\sim \beta_2)$ is depicted in Fig. 5. 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 α, the parameter on which the regularities turn out to be defined; the line-length ratio is devoid of special values and hence of special subcategories. Therefore we will usually write $p(\text{3-CHAIN}|\alpha)$ rather than $p(\text{3-CHAIN}|\mathbf{X}^3)$, 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 α is 0 and the deviation process completely drops out. At this point, the curvature of the inferred underlying generating process vanishes. This

	generative model	codimension
3-cluster	$\vec{t} + \vec{d}$	0
	⋮	
3-chain	$\vec{t}$	1

FIGURE 6. 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.

nongeneric subspace of the overall space. containing only collinear triplets, is of course 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", a case from which detection of the curvilinear process is maximally (though not perfectly) reliable (see [5] for some subtleties on this point).

The category lattice depicting this structure is shown in Fig. 6. 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 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. (The change in dimension from 2 to 1 is more conveniently expressed as change in codimension[4] from 0 to 1, for reasons outside the scope of the current discussion.)

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 is a question to be addressed by statistical techniques. That is, given a randomly chosen member of the generic triplet category, how straight do we expect the angle between the dots to be; and what is the distribution of angle over many such randomly chosen triplets? (Equivalently, how much of the "deviation" process do we typically expect to have been applied in the construction of a random generic triplet?) The true, statistical "fact of the matter" can be seen as a kind of normative backdrop against which various perceptual and cognitive beliefs about the same question can be compared. As it happens, this question was answered for the generic category by Kendall and Kendall [10], whose results are summarized in the next section. (Section 2.2 also discusses the corresponding statistical "facts" for the curvilinear case (codimension 1 case).) Before looking at the normative question ("what do random members of each generative category look

[4] Codimension is the difference in dimension between a given category and the overall embedding space (the generic category).

like?"), though, we start with psychological question ("what do subjects think they look like?"). To answer this question for the generic (codim-0) category, we begin by asking where in dot-triplet space would this generic category be centered: that is, where in the observer's mental construction of this space would the maximally generic point be located.

2.1.1. *The maximally generic point in triplet space.* In [7], evidence was presented (in a different domain) that subject 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 nongeneric (collinear) subcategory, i.e. at $\alpha = 0$, and one at a point construed to be *maximally* generic. Putatively, following [7], maximally generic simply means maximally *distant* from nearby non-generic points. In the dot-triplet domain, the most non-generic angle is simply the angle most distant from the non-generic point $\alpha = 0°$: namely 120°, the largest value the angle can take, in accordance with the symmetry argument presented above. Though it is easy to see that 120° 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 itself a geometric object. We can display its structure in a convenient form by the contrivance of placing the first and second at dots, $\mathbf{x}_a$ and $\mathbf{x}_b$ at points (0,-1) 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. 7). 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 figure).

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 right most point. In this region, labeled **A** in the figure, the angle of the corresponding dot triplet changes from 0° at the rightmost point to 90° at the topmost point, etc. If one were to continue past the point $(\frac{-1}{2}, \frac{\sqrt{3}}{2})$ (where $\mathbf{x}_c$ is actually drawn in the figure), though, one would begin to form triplets whose straightest vertex was not equidistant from the other two dots. So instead 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. From the point of view of shape alone, that is, the two sections **A** and **B** contain exactly the same 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 to represent the triplet space in its most intrinsic form. Notice

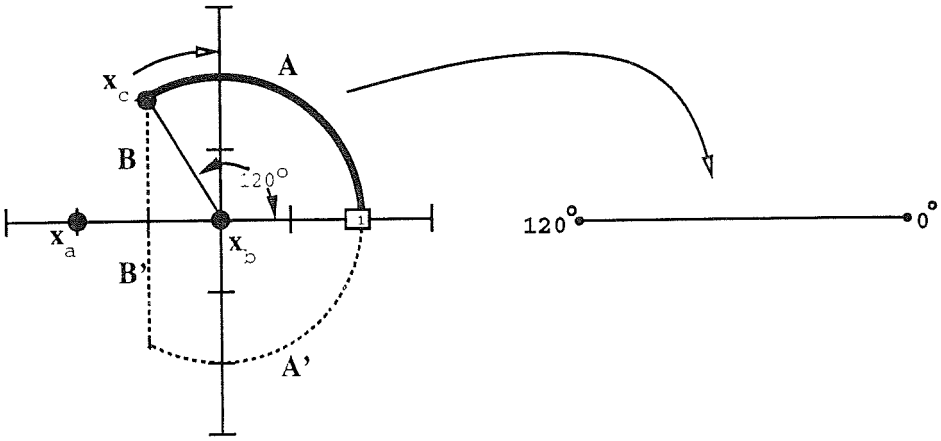


FIGURE 7. Schematic depiction of the space of dot triplets, making clear how $\alpha = 0$ and $\alpha = 120°$ are maximally distant points. The space of equal-leg-length dot triplets can be depicted (left) as a part of a circle embedded in the 2-D space of general dot triplets: the highlighted part (labeled 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° and 120°. 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.

that it has two definite endpoints: 0° and 120°, respectively. Notice further 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° has been identified as a non-generic point in the space, we can confidently identify 120°, the equilateral triplet, as the maximally non-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. Section 3 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 [10], which we present first. Then we consider alternative distributions constructed around the category theory presented in the previous section. The latter probabilities, which

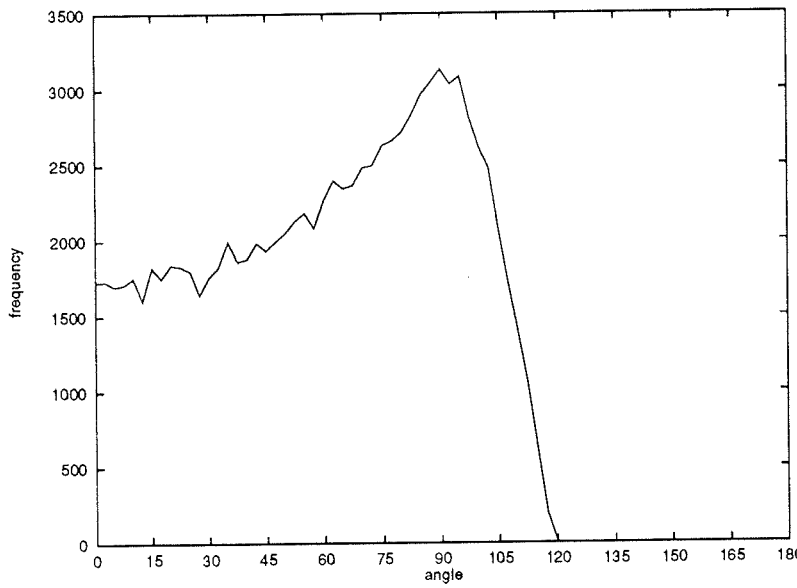


FIGURE 8. Simulation of the distribution (derived in [10]) of the straightest angle of random dot triplets. Here we generated 100,000 triplets (drawn from a Gaussian distribution), and measured their smallest deviation from straight (0°). (Thus this value is naturally bounded by 0° and 120°.) The theoretical peak is at 90°.

the experiments in Section 3 suggest subjects actually adopt, seem to reflect their tacit beliefs about the regularities extant in the dots world.

2.2.1. *Random clusters of three dots.* Kendall and Kendall [10] (see also [2] considered the question: when triangles are generated at random (i.e. with dots drawn from a 2-D Gaussian distribution in the plane), what distribution of largest internal angle should we expect? This is equivalent, of course, to asking what random members of 3-CLUSTER look like. Though the problem is simple to state, the answer turns out to be somewhat counterintuitive. A simulation of the resulting curve is presented in Fig. 8. (Notice that in the figure the angle measures deviation from straight, and is thus the complement of the largest interior angle in the random triangle.) The distribution is naturally bounded by 0° and 120°, since we are selecting the largest interior angle, and all triangles have at least one interior angle at least as large as 60°.

The peak of the distribution is at 90°—perhaps counterintuitively. One intuition about why the peak is not larger is that values larger than 90° come from triangles close to equilateral, requiring in effect a close accidental alignment between the two independent angles.

2.2.2. *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 of triplets is to impose a distribution on the expected curvature of the underlying curvilinear process. For example, we might stipulate a distribution on the family of curves; then choose a curve randomly from this distribution; then

generate each triplet by choosing three points at random along this curve; and then compute the density of triplets generated in this manner. This scheme amounts to assuming something about the world, though—that is, we get out a distribution that reflects, in effect, how much bend per arclength we expect curve processes in the dots world to contain. Thus we will not get a single normative answer as in the 3-CLUSTER CASE, but rather a family of answers keyed to assumptions.

A loose but simple argument suggests that this distribution can be expected to be approximately normal in form. Since in a completely arbitrary curve, nearby points can be thought of as independent (i.e. the curve can bend an arbitrary amount between adjacent sampled points), we can expect the density of the deviations from tangency at various sampled points along the curve to act as the sum of a number of independent, identically distributed random variables—and thus, by the central limit theorem, normal in the limit. The mean is presumably zero by symmetry (strictly on the contingent assumption that curve processes in our world are unbiased with respect to handedness). But the standard deviation is arbitrary—it depends on the assumptions about the distribution of curves and inter-sample arclengths, and thus on the distribution of the random variable.

Hence we can conclude that randomly chosen members of 3-CHAIN will have angles chosen from a normal distribution with mean 0° and *arbitrary* width. Like the 3-CLUSTER case, this is in some respects counterintuitive. Imagining random members of this category, one imagines only relatively straight triplets (as do subjects: see Experiment 2 below). That is, the only triplets one imagines generated are those that one would actually classify as belonging to this category—that is, as having been generated by 3-CHAIN. But under certain assumptions about the curve, the tails of this distribution will actually produce 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.

2.3. 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, and substitute for it some kind of heuristic expectation. The critical intuition concerns what human observers, as opposed to statisticians, mean by "random". In this context, a statistician means by the word random "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 (cf [1] on the need for bias on the space of perceptual hypotheses).

Under this reasoning, it makes sense (as alluded to above) to construct the class

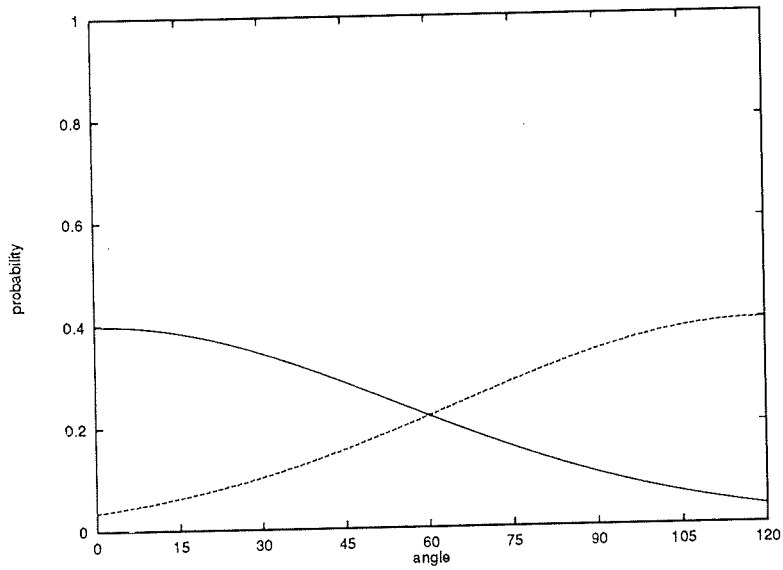


FIGURE 9. The two hypothetical prior "modes", one at 0° (straight), the other at 120° (generic). Standard deviations are each 53.059°. the fitted value drawn from Experiment 1.

density $p(\mathbf{X}|3\text{-CLUSTER})$ by placing a peak at the angle furthest in the space from straight, a point established in Sec. 2.1.1 to be at $\alpha = 120°$. Similarly, we construct $p(\mathbf{X}|3\text{-CHAIN})$ with a peak at $\alpha = 0°$. We assume, arbitrarily but reasonably, that these two densities are each normal in shape.

If these two peaks were "real" densities, determined by the world, then the world could also give them different heights (corresponding to the scalar priors, $p(3\text{-CHAIN})$ and $p(3\text{-CLUSTER})$), and different widths (standard deviations). Moreover, in the world there might be some other regularities that occur—other factors contributing to the density $p(\alpha)$. Here, however, the densities are being constructed heuristically by the observer rather than dictated by the environment—that is, they are "subjective" densities, constructed internally as part of an inferential machine. Hence the observer can unilaterally disallow the complicating factors of varying heights, varying widths, and additional regularities, about which it has no knowledge. The resulting rather severe set of regularizing assumptions—that the peaks (1) have equal height (2) have equal width, and (3) mutually exhaust the set of events that occur in this domain—enforce the observer's desire to impose on (or impute to) the world as little structure as possible, and no more than is necesary to distinguish the categories of interest. The resulting class densities consist entirely of two normal "humps," with the same height and width, one centered on each regularity (depicted in Fig. 9).

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 node. (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 of no structure). A theory whereby these widths can be derived will be presented in the discussion

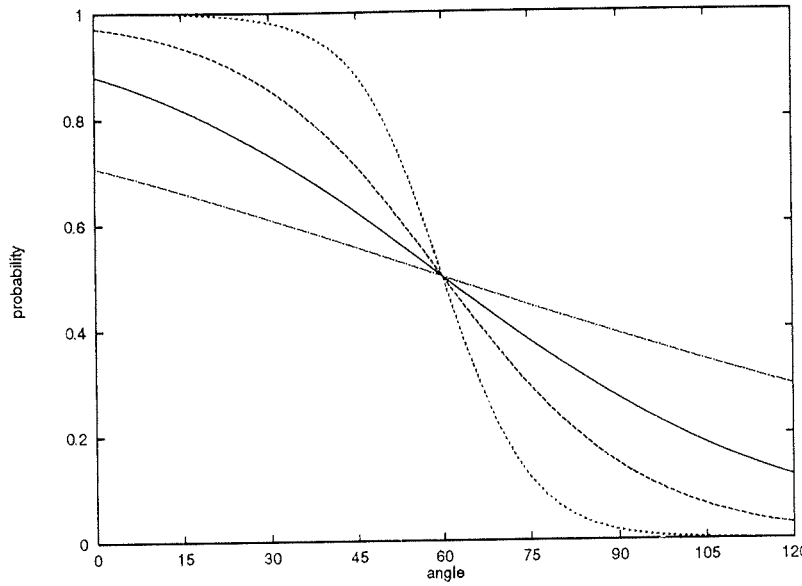


FIGURE 10. The hypothetical a posteriori curve, shown at four different standard deviations ($\sigma = 15°, 45°, 75°$, and $105°$).

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

With the two conditional densities established as described above, we are now in a position to construct the a posteriori functions $p(\text{3-CHAIN}|\alpha)$ and $p(\text{3-CLUSTER}|\alpha)$, by which observers actually assign a category label to an observed triplet. First, by stipulation that these two categories are the only ones under consideration, $p(\text{3-CHAIN}|\alpha) = 1 - p(\text{3-CLUSTER}|\alpha)$, so we can look only at the chain case. Substituting in the assumption of normal class densities, we have

$$p(\text{3-CHAIN}|\alpha) = \frac{N(0°, \sigma)}{N(0°, \sigma) + N(120°, \sigma)} \tag{2.3}$$

where $N(\alpha, \sigma)$ is the normal distribution with mean α and standard deviation σ.

This curve (plotted for four values of σ in Fig. 10) has the familiar sigmoid shape characteristic of many psychometric functions, with a transition from one category to the other whose sharpness depends on σ. Thus the category theory predicts that human observers will draw category distinctions between 3-CHAIN and 3-CLUSTER according to a curve drawn from somewhere in this family. The next section presents an experiment that tests this empirically.

3. Experiments on human observers

We now report an experiment designed to directly probe human subjects' intuitions about membership in the categories 3-CHAIN and 3-CLUSTER. In the main part of the experiment (labeled 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). In a follow-up to Experiment 1 (labeled Experiment 2), the same subjects were asked to generate a number of "typical" members of each of the two categories, thus probing their class-conditional densities in a more directly productive manner. It should be noted that subjects were instructed that two categories of dot triplets were under consideration, so neither experiment tests our assumption that perceivers automatically assign prior probability to only these two classes; rather, the experiments are a way to reveal the exact form of the class density functions subjects unconsciously assign to the two classes under consideration.

3.1. Experiment 1: Perceptual classification of dot triplets. This experiment investigated subjects' beliefs about the dots world by directly asking them to classify dot triplets as the products of either a 2-D patch process or a 1-D curvilinear process.

Method

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

Procedure. 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 flat 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 $-120°$ to $120°$ in increments of $7.5°$) with 12 levels of line-length ratio (from 1.00 to 2.85 in integral powers of 1.1).

Results

The results of Exp. 1 are plotted in Fig. 11. The data can be taken as representing subjects' posterior classification function, $p(\text{3-CHAIN}|\alpha)$. 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 corresponds to 24 observations per subject, by 16 subjects (except for $\alpha = 0$ which is only 12 observations, since there is no sign).

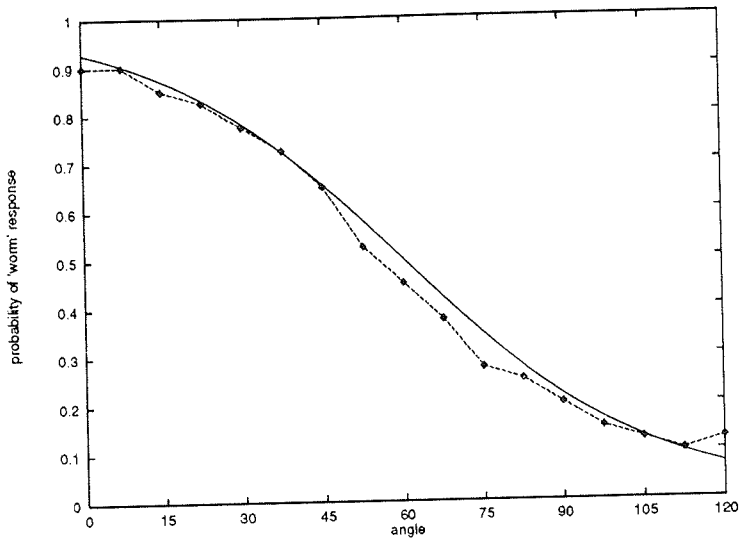


FIGURE 11. Data from Experiment 1. On the Y-axis are subjects' numeric responses, converted to a judgment of probability that the observed triplet was a "worm". On the X-axis is the angle of the dot triplet, collapsing over line-length ratio and sign of angle. Also plotted is the best-fit curve from the likelihood family described in the text, in which the two prior peaks each have standard deviation $= 53.059°$.

We can use these data to estimate the means of the two (implicit) prior peaks from which the posterior function is drawn. Using a conservative model in which the two standard deviations and the two heights are allowed to vary independently, the two means are estimated at 1.439° (99% confidence interval from −0.168° to 3.046°) and 118.539° (99% confidence interval from 116.909° to 120.169°). Using the more theory-dependent model in which the two peaks have the same height and width, the two means are estimated at −2.566° (99% confidence interval from −5.061° to −0.071°) and 116.372° (99% confidence interval from 113.924° to 118.820°). The peaks clearly come close to the 0° and 120° predicted by the category model, though the more theory-dependent estimates leave some room for doubt as to the precise locations of their centers, in each missing the predicted values by a small amount. This slight deviation may well be due to some unknown additional factor, though it might just as well be due to a small degree of kurtosis in the subjects' densities. In any case the subjects' "generic" peak is quite distant from the normative value of 90°.

The figure shows the best fit (least-squared error) curve from the family of posterior functions described in Eq. 2.3, $R^2 = 0.988$, $F_{1,16} = 1215.8, p \ll 0.01$. Here we use the theoretical model in which both peaks have the same height and width, so that the only free parameter is their common standard deviation, σ. This value has

been estimated at 53.059°. Note that because this posterior curve has no separate height parameter, the close fits at the point-predictions for $\alpha = 0°$ and $\alpha = 120°$ are particularly impressive.

Discussion

It is clear from the results that subjects classified dot triplets according to Eq. 2.3, to an extremely close approximation, using standard deviations of about 53° for each mode. Subjects were in effect placed into a novel, imaginary domain of bottom-dwelling 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—i.e., putatively, wholly out of the sort of regularity-driven category-generating apparatus described above.

Accounting for the distribution widths. To fully characterize these data, we would like to be able to say something about the sole remaining fitted parameter, σ, 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 σ can be derived from the premise that the observer is attempting to construct priors so that an optimally simple maximum-likelihood scheme can be employed in order to calculate posterior classifications. That is, given an observed triplet with angle α, the observer would like to be able to read off the posterior probability of membership in the class 3-CHAIN, $p(\text{3-CHAIN}|\alpha)$, directly from the prior density $p(\alpha|\text{3-CHAIN})$. Since these two quantities are actually related by Eqs. 1.1 and 1.2 (in which we can ignore the constant terms $p(\text{3-CHAIN})$ and $p(\text{3-CLUSTER})$), equating them simply requires that the denominator (Eq. 1.2) sum identically to unity. Pictorially, the prior and posterior probabilities become identical when the two "humps" sum together to produce a flat horizontal line of height unity (Fig. 12). Putting it another way, this condition requires that between the two generating processes, all shapes of triplets are equally probable in the world.

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

$$\frac{d}{d\sigma}\left[\int_0^1 [e^{-(\frac{x^2}{2\sigma^2})} + e^{-(\frac{(x-1)^2}{2\sigma^2})} - 1]^2 dx\right] = 0. \tag{3.4}$$

An analytic solution to this minimization is problematic because of the Gaussian terms, but numeric estimation provide a figure of $\sigma \approx 0.4435(120° - 0°) = 53.225°$, almost exactly the empirical figure of 53.059° (only about 0.31% high). Anther 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 remarking on the fact that subjects' induced density functions (Fig. 9) 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 height and equal variance. In effect the subjects here have fabricated two class densities exactly as if to simulate signal and noise, with the triplets generated by the generic

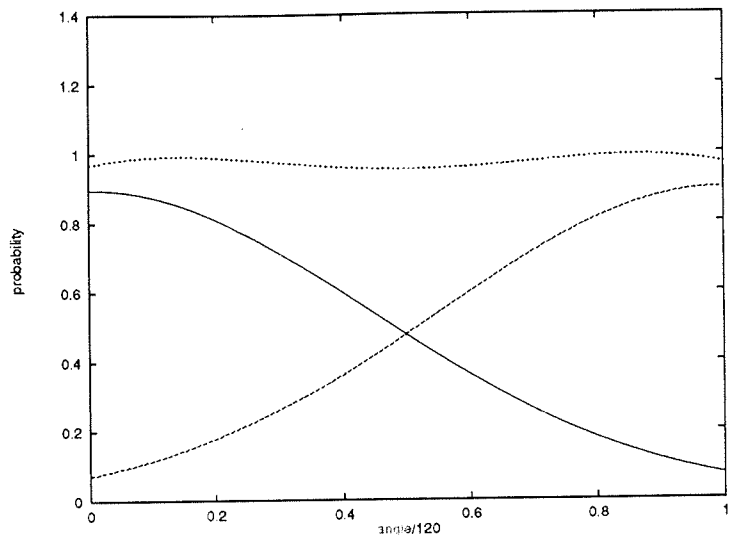


FIGURE 12. The two class densities sum identically to about unity, meaning that all triplet angles are about equally likely, and prior and posterior probabilities are as nearly identical as possible.

category, 3-CLUSTER serving as noise, and $d' = 2.2599(= 1/0.4435)$. In a sense this is consistent with the idea that the generic category serves as a random backdrop or distractor to the regular (codimension-1) 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 or irregularity; or, more precisely, of causal independence in the generation of the dots.

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. Their model of collinearity constitutes a kind of theory about a basic form that regularity in this unknown domain might take. They are making the quite plausible assumption, in effect, that observed objects in the new domain are *not* generated completely randomly; but rather that they are subject to causal forces that, while mysterious in exact nature, are coarsely predictable in form.

3.2. Experiment 2: Generation of dot triplets. Experiment 1 allowed us to infer the form of subjects class conditional probability densities indirectly. Experiment 2 takes a different tack by probing these densities directly (though less sensitively), by asking the same subjects from Experiment 1 to generate "random" members of each set.

Method

Immediately after completing Experiment 1, the subjects were given questionnaires, asking them to draw 10 different examples of each type of creature, "worms" and "flatfish". It was feared that if subjects were asked to produce a large number of examples, they would tend to increasinly produce deliberately (an unrepresentatively) atypical ones On the other hand, asking for only a very small number would provide an overly impoverished picture of their probabilistic beliefs. The number chosen (10) is a somewhat arbitrary compromise between these concerns. It should be noted though that any progression in the subjects' drawings from more to less prototypical has been ignored in the analysis below. This fact may well be taken to cast a certain amount of doubt on the idea that the procedure directly elicits the subjects' priors. Due to the coarseness and origins of the data, then, the plots below are to be viewed more impressionistically than analytically, and more for insight into subject's priors than for final adjudication of theoretical questions.

Results

Each of the resulting dot triplets were coded by hand as an angle (the positive deviation from straight, as above) and a line-length ratio. Again, we collapse over line-length ratio, and plot frequency as a function of angle. Fig. 13 and Fig. 14 show the results for "worms" and "flatfish" respectively. The peaks at 0° and 120° are clearly visible[5], indicating the subjects produced many exactly straight "worms" and many almost exactly equilateral "flatfish". Intriguingly, though, both plots seem to have one additional peak: a second "worm" peak at about 45°, and a second "flatfish" peak at about 90°. Unfortunately, the limited amount of data available (constrained by the factors discussed above) prevents these data from being analyzed other than impressionistically.

Discussion

An argument from [7] suggests an explanation for the additional peaks. Each generic triplet, putatively, is generated by starting from some regularity and moving some distance off through the space, a distance far enough to make it distinguishable from objects at the regularity, but not so far as to encroach into the next categorical region of the space. Thus additional qualitative categories spin off in the region near a regularity, exactly one per "qualitative" region. Hence the four spikes in the two frequency plot can perhaps best be described (in order of increasing angle), as $0, 0+\epsilon_1, 120-\epsilon_2$, and 120. Again, the categorical scheme provides a formalism in which each qualitative category is assigned a unique symbolic description. The origin of the exact values of the ϵ's is unknown. Rather, the point is simply that the qualitative symbolic description generated by the category theory is isomorphic to the modal peaks evident in the subjects' data.

3.3. General discussion of experiments. Taken together, the two experiments paint a hybrid picture of human observers' clustering interpretations, in which posterior inferences are drawn using distinctly Bayesian machinery, and yet in which critical antecedents such as the class-conditional densities are fabricated virtually

[5]Note that the slight migration of the 120° peak towards smaller angles is attributable to subjects' inaccuracy in drawing the figures. That is, *any* deviation from an exact equliteral triangle will be *decrease* the measured angle, since we always measure at the largest interior angle.

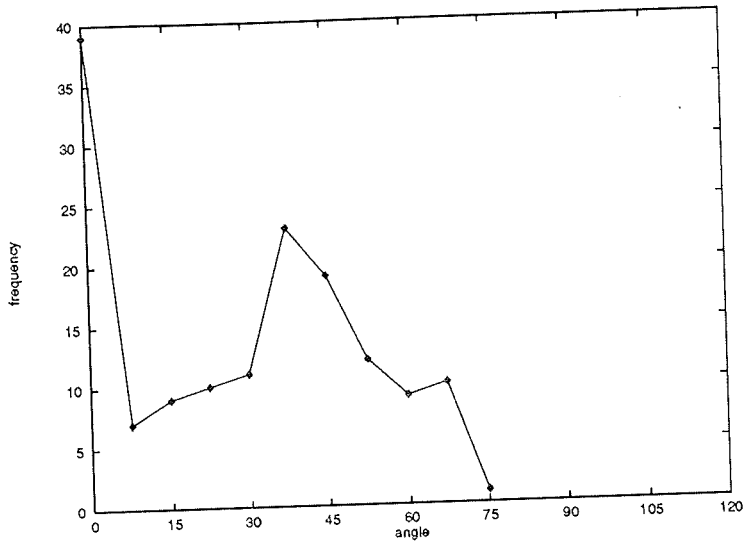


FIGURE 13. Data from Experiment 2 ("worms"). Frequency is plotted as a function of the angle of the dot triplet. Two modes are visible, one at straight, and one at slightly off from straight.

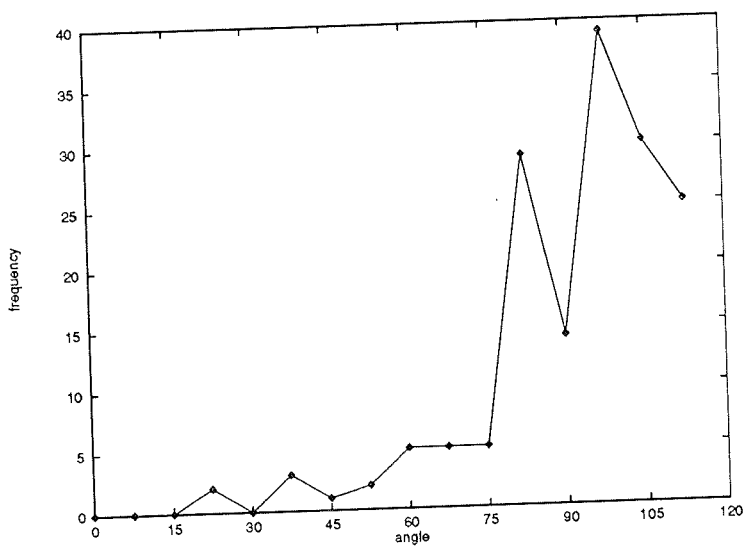


FIGURE 14. Data from Experiment 2 ("flatfishes"). Frequency is plotted as a function of the angle of the dot triplet. Two modes are visible, one at about equilateral (120°) and one at about 90°.

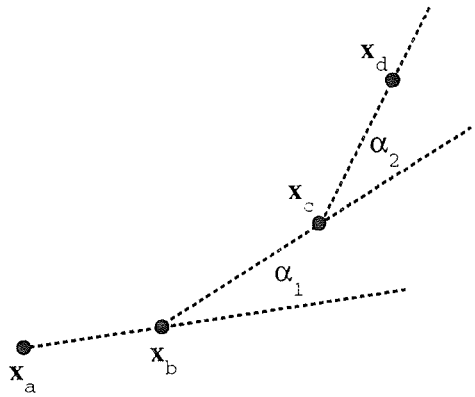


FIGURE 15. Parameterization for four dots—a multi-molecular cluster.

out of whole cloth. While the relationship between the densities drawn from Experiment 2 and those implied by Experiment 1 is unclear, subjects' responses in Experiment 1 clearly demonstrate that they classify dot triplets in an approximately normative, *given* that they draw category densities according to 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. From these beliefs in turn they can infer which hypotheses about the state of the world are more and less likely to be true.

4. 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 α_1 and α_2 (Fig. 15), 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. Critically, such an expectation is justified *only* in such a world; the expectation is not justified in a world lacking such a regularity (see [14] on this point).

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

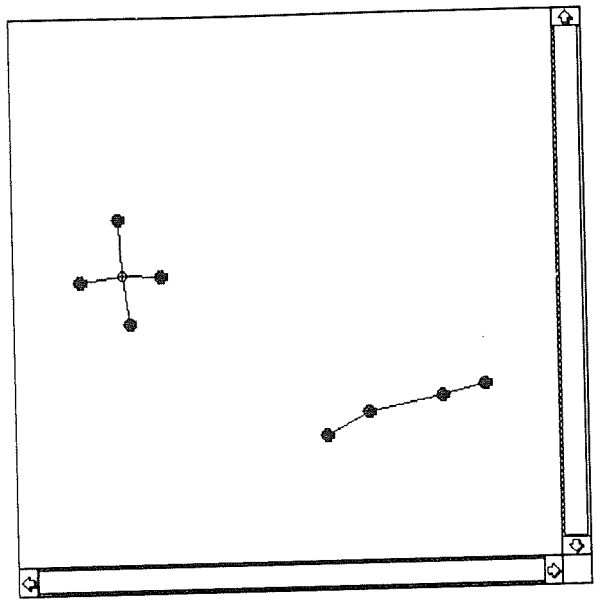


FIGURE 16. Using a maximum a posteriori grouping procedure as described in the text, the computer finds two groups: an unordered cluster of four dots on the left (drawn linked to their centroid) and a chain of four dots on the right (drawn with links from dot to dot).

would be

$$p(\mathbf{X}^4|\text{curvilinear process}) = p(\alpha_1|\text{3-CHAIN})p(\alpha_2|\text{3-CHAIN}). \tag{4.5}$$

Substituting Gaussian densities for each triplet density, this become approximately

$$p(\mathbf{X}^4|\text{curvilinear process}) \approx \frac{1}{\sqrt{2\pi}} e^{[\frac{1}{2\sigma^2}(\alpha_1^2+\alpha_2^2)]}. \tag{4.6}$$

This equation ought to tell us the degree to which a given chain extraction can be taken 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"), we can 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 a 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. 16. 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.

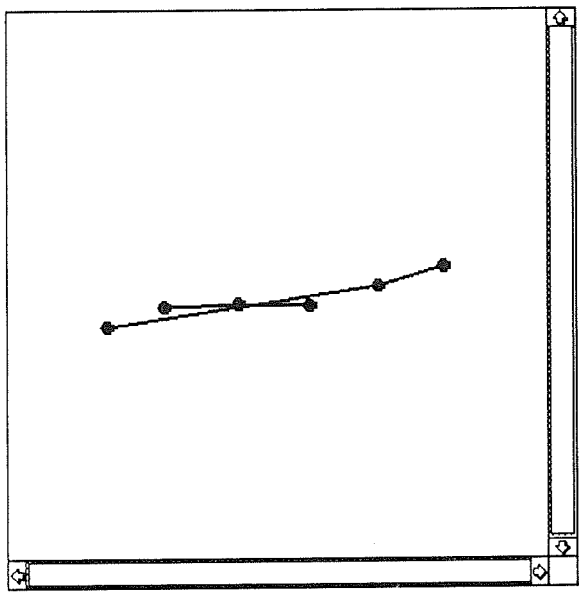


FIGURE 17. The maximum a posteriori clustering of these six dots (given the assumptions in the text) is as two distinct but intertwined chains, not the single long chain that human observers prefer.

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. 17). 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 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 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, which we defer to another paper.

5. Conclusion: causal models of dot groupings?

The inferential machinery presented in this paper focuses a great deal of apparatus on what is, after all, a very simple domain. The data presented in Section 3, 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 counterintuitively, 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 [14] 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 ought to 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.

References

1. Bennett, B., Hoffman, D. and Prakash, C. *Observer mechanics: A formal theory of perception*, Academic Press, London, 1989.
2. Broadbent, S.R. *Simulating the ley hunter*, J. Royal Statistical Society A **143(2)** (1980), 109–140.
3. Do Carmo, M. *Differential geometry of curves and surfaces*, Prentice-Hall, Englewood Cliffs, N.J., 1976.
4. Duda, R. and Hart, P. *Pattern classification and scene analysis*, John Wiley, New York, 1973.
5. Feldman, J. *Perceptual decomposition as inference: continuous curvilinear processes*, Occasional Paper #42, M.I.T. Cognitive Science Center, 1991.
6. ——— *Constructing perceptual categories*, Proc. IEEE Conf. Computer Vision and Pattern Recognition (Champagne-Urbana, Illinois, 1992), 1992, 244–250.
7. ——— *Perceptual categories and world regularities*, Ph.D. thesis, Department of Brain & Cognitive Sciences, M.I.T., 1992.
8. Jepson, A. and Richards, W. *What is a percept?* Tech. Report RBCV-TR-93-43, Dept. Computer Sci., Univ. of Toronto, 1993. (Also Occasional Paper #43, M.I.T. Cognitive Science Center, 1991.)
9. ——— *What makes a good feature?* Spatial Vision in Humans and Robots, (L.Harris and M. Jenkin, eds.), Cambridge University Press, Cambridge (to appear). See also A.I. Memo #1356, M.I.T., 1992.
10. Kendall, D.G. and Kendall, W.S. *Alignments in two-dimensional random sets of points*, Adv. Applied Probability **12** (1980), 380–424.
11. Leyton, M. *Perceptual organization as nested control*, Biological Cybernetics **51** (1984), 141–153.
12. ——— *Inferring causal history from shape*, Cognitive Science **13** (1989), 357–387.

13. Matthews, G. and Hearne, J. *Clustering without a metric,* I.E.E.E. Trans. Pattern Analysis and Machine Intelligence **13(2)** (1991), 175–184.
14. Richards, W., Jepson, A., and Feldman, J. *Priors, preferences, and categorical percepts,* Proc. Workshop on Psychophysics and Computation (Chatham, MA, 1993), to appear.
15. Späth, H. *Cluster analysis algorithms for data reduction and classification of objects,* John Wiley, New York, 1980.
16. Stevens, K.A. *Computation of locally parallel structure,* Biological Cybernetics **29** (1978), 19–28.
17. Witkin, A. and Tenenbaum, J. *On the role of structure in vision,* Human and Machine Vision (J. Beck, B. Hope, and A. Rosenfeld, eds.), Academic Press, New York, 1983.
18. Zucker, S.W. *Early orientation selection: tangent fields and the dimensionality of their support,* Computer Vision, Graphics, and Image Processing **32** (1985), 74–103.

RUTGERS UNIVERSITY CENTER FOR COGNITIVE SCIENCE, PSYCHOLOGY BUILDING NEW WING, BUSCH CAMPUS, RUTGERS UNIVERSITY, NEW BRUNSWICK, NEW JERSEY, 08855
E-mail address: jacob@ruccs.rutgers.edu