RuCCS TR – 6

Perceptual Categories & World Regularities

Jacob Feldman
Center for Cognitive Science
Department of Psychology
Rutgers University
jacob@ruccs.rutgers.edu

Technical Report #6, September 1993
Rutgers University Center for Cognitive Science
Rutgers University
PO Box 1179
Piscataway, NJ 08855

Acknowledgments

This paper was originally a dissertation submitted to the M.I.T. Department of Brain and Cognitive Sciences. The dissertation is reproduced verbatim here as part of the Rutgers University Center for Cognitive Science (RuCCS) Technical Report Series, with the sole alteration being this abbreviated acknowledgements section.

I am most especially grateful Whitman Richards, who served as supervisor to the original dissertation and inspired the work. I am also especially grateful to the members of the thesis committee, Steven Pinker, Shimon Ullman, and Alan Yuille; and to Allan Jepson, who served as a sort of unofficial reader and whose insights greatly influenced me. Others to whom I am grateful for extensive comments on drafts are Aaron Bobick, Marc van Hulle, Michael Leyton, and Eric Saund. I also greatly benefitted from discussions with Paul Bloom, Susan Carey, Stan Dunn, Suzie Johnson, Ellen Hildreth, Michael Jordan, Sandy Pentland, Molly Potter, Sandeep Prasada, Josh Stern, Karin Stromswold, Michael Tarr, Tony Wilkie, and Karen Wynn. I am grateful to a number of people for carrying out the thankless task of procuring subjects: Marie Coppola, Jane Gwiazda, Tatiana Schnurr, Sara Solnick, Karin Stromswold, and Josh Tenenbaum. Special thanks also to Bill Gilson and Jan Ellertsen.

Abstract

What makes a good category? Perceptually natural categories—object classes in which an infinity of distinct forms collapse compellingly into a unary description, such as *triangle*, or *dot on a line*—impose structure onto our perceived world. This thesis investigates the formal composition of simple category models, and the properties that distinguish such categories from arbitrary incoherent sets of unrelated objects. The goal is a formal characterization of human category inferences, including the rather subtle relationship between a perceiver's existing concepts and entailed inductive hypotheses. A critical issue is the formal relationship between mental models and actual world regularities (i.e., covariation in the world among logically orthogonal properties, or "natural modes"). The main formal structure is a lattice of category models, a relational structure that enumerates the various distinct uniform category models in a model class. The lattice serves as a kind of category hypothesis generator, providing the observer with a closed class of distinct models from which to select, each of which corresponds to a coherent "causal" model of the induced category. A computer program is developed to check the validity of the theory, and to generate the lattice of category models for complex families.

A series of experiments are reported in which subjects were asked to induce simple categories from a very small set of unfamiliar sample objects (either one or three objects), and generate novel examples of the category. The results corroborate the lattice theory, and lobby against a view of categorization as any kind of a statistical summary of environmental frequency distributions. In several conditions, subjects' produced a frequency distribution that actually contained a larger number of modes (peaks) than there were objects in the sample set; in another condition, subjects' frequency distributions exhibited a mode in a region of the model space where they never observed any examples; and in another condition, subjects produced a frequency distribution that was distinctly modal in a region of the model space in which distribution they observed was carefully arranged to be perfectly flat. In all these cases, the frequency modes corresponded neatly with nodes on the theoretical category lattice computable from the sample set.

Table of Contents

- Chapter 1 -

Introduction: Perceptual Categorization as a Window onto World Structure

1.0. Introduction

This thesis investigates the formal properties that distinguish *coherent* perceptual categories—collections of objects that all constitute relatively uniform variants of some underlying model—from incoherent ones. An account is proposed for the compelling and intuitive quality manifested by some simple inductions—namely, the ones that human observers are willing to assign to small collections of sample objects. The motivating presumption is that as we attempt to build a veridical and useful representation of the world, coherent perceptual categories represent relatively good bets to correspond to real, causally coherent object-generating processes in the world; that is, they pick out the world regularities most likely to be of interest to us.

1.1. Motivation

Human observers have a striking capacity to collapse a small number of observed objects into one simple, abstracted model, somehow sidestepping the inductive ambiguity inherent in such an inference. Amazingly, such a model can in many cases be constructed from just a single example, as in Fig. 1-1.

Fig. 1-1. *"One-shot categorization"*. Even with only this one example, many observers generalize to exactly the same category model: *segment with a dot on it*. That is, we guess that in other examples of the same category, the dot will always appear somewhere on the segment.

Here, when asked to predict what they would expect other examples of the "same" category to look like, most observers guess that they would also have the dot exactly *on* the segment, though its position *along* the segment might vary (Fig.1-2).

Fig. 1-2. More examples of the category model induced from Fig. 1.

Indeed, most subjects make exactly this generalization—which is, in a way, a wild inductive leap—after just one example (about 90%, and 100% after 3 examples; see Ch. 5). That is, after one example, observers have guessed with confidence that certain properties of the given example—e.g. the fact that the dot is touching the line exactly—are intrinsic to the category, while other properties, even some that are of exactly the same type—e.g. the exact position of the dot *along* the segment—are transitory artifacts of the sample and should not be anticipated in future examples of the same category. Still other properties, like the size of the dot, the length of the line, and the time of day at which the object was observed, might also be deemed unlikely to be preserved in future examples. Distinguishing these properties from fixed ones from just one example is intrinsically difficult—even just a second example would provide *some* hint as to which aspects of the sample are probably the intrinsic properties and which the variable ones. That a model can be constructed from just one example suggests that our categorizing apparatus operates under *extremely* tight constraints as to both the composition of induced categories are their relationship to observed objects. This thesis will explore some of these constraints.

Some specificity is in order about exactly what is meant by a *categorical interpretation*. Shepard (1987) speaks of identifying an object with its "consequential region," meaning the region of the parameter space nearby the observed object, within which other objects have been observed to behave equivalently to the observed object, and hence across which the observer is well served to generalize. The act of categorization, Shepard argues, rather than being merely a "failure of discrimination" (as had been argued by behaviorists) is a purposeful, inferentially sensible process whereby an observer seeks to discover object properties. This notion gives some purpose and form to the otherwise vague idea of the "correct" categorization of an observed object, i.e. the category to which the observed object can most naturally be said to belong.

In speaking of identifying a category—as if categorical status were an independent, objective fact—it is prudent to sidestep any dependence on categories' possessing some sort of intrinsic existence outside of the mental apparatus of some perceiving system. This view, sometimes referred to (perhaps ironically) as a "realistic" one, finds no comfortable place within computational theories of perception. Unlike the case of inferring, say, a depth value of a physical point on an object, "inferring" the category membership of an observed object threatens to become a circular notion—in which the observer ascertains a world property that can only be defined within its own internal models. Thus in order to create a computational model of categorization, a particular conception of what is meant by world "stucture" must be adopted and made as explicit as possible; then, the route by which this conception of structure tends to gives rise to distinct "categories" in the world must be explained; and finally the relationship between the apparent properties of an object and its correct categorical identification must be made explicit. Only with definitions of "regularity" and "category" in place, in other words, can the inference loop be grounded.

To motivate this idea about the ontogeny of categories, consider how the identification of a model from the observed object in the line-on-dot example seemed to involve the ascription of a certain kind of *causal story* to the example—a certain account, albeit a coarse and extremely speculative one, of the constraints that govern the object's structure. For example, when we interpret that the dot will always be somewhere *on* the line, that is because we think it is on the line in the example for some reason—some reason we think is liable to remain in force in other instances of the "same kind." Perhaps the dot is stuck to the segment; perhaps it is an insect climbing a branch. In any case we assign importance to the property "on" in this case, because we suspect that this property did not arise by accident, but rather is genuinely the result of some causal forces that brought it about. Hence we can feel assure that these same causal forces, and hence typically this same observed property, will *also* appear in other objects of the same "kind."

1.1.1. Critical dependence on world regularity

In the example, the intuition is strong that our induced model—*dot on a line*—really is liable to be correct in some sense. We feel that unobserved examples of the class, if encountered in anything like a natural setting, really would be *likely* to have the dot right on the line just, as in the example, though possibly at a different position along the line.

This intuition begs some sort of justification. It is basic to the inductive set-up that such a speculation about the general properties of the class from which the sample was drawn is not deductively guaranteed to be correct. Indeed, in a completely unstructured world, in which coherent classes with consistent properties did not exist, there would *be* no reason to expect it to be correct—no reason for the world's probability density function to behave in the manner we seem to expect it to. Yet our feeling is that in practice, it will. Clearly, this is because the world we live in is *not* an unstructured chaos of causally unrelated properties, but rather is a highly regular, organized place in which some properties causally influence other properties in a consistent, reliable and useful way. In order to create a concrete theory of category structure, it follows, we need to commit to some sort of model of just what we mean by "structure," i.e., what sorts of categories we would like to detect.

When does a regularity not look like a regularity? Our conception of "regularity" can be focused somewhat by considering a pseudo-regularity that, despite being well-defined, "simple" (in some sense), and formally similar to the one we induced in Fig. 1, is apparently not readily accessible to human categorization faculties (Fig. 1-3).

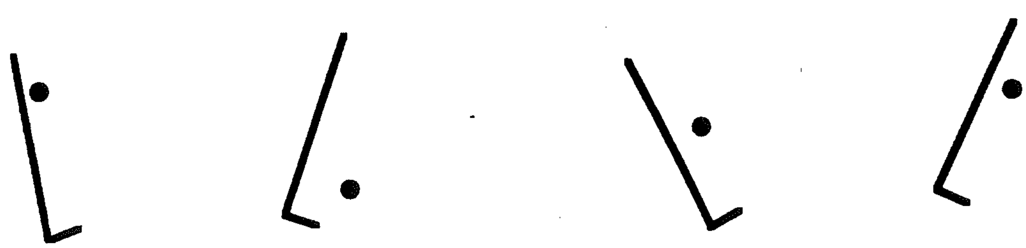


Fig. 1-3 What is the category?

What is the generalization here, or, and what would we expect other examples of this category look like? The apparent common property seems to be that in each example the dot is to the right of the segment.

In truth, though, there is a more particular regularity present: the dot always falls on the same imaginary line that descends from the top of the L at a 30° angle (Fig. 1-4). While we detect some regularities, that is, we completely miss others.

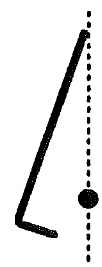


Fig. 1-4. The actual generating process of the objects in Fig. 1-3 : the dot always falls on the (imaginary) line (shown dotted). A regularity that we systematically miss?

Of course, it is hardly surprising that we fail to detect every conceivable sort of regularity. There is always an infinity of different "regularities" consistent with any sample, after all, the vast majority of which are arbitrary, complex, and invoke concepts and patterns that no naïve human observer would recognize as in any sense "regular." (Compare the theory concerning the complexity of strings of symbols of Kolmogorov, 1965; Martin-Lof, 1966; and Chaitin, 1966, in which in any given description language, most descriptions of a given string are longer than the string itself, and hence only a small minority of strings of a given length can be summarized compactly.) In this example, though, we missed one that is equivalent in form to another (the dot on the line) that we caught easily—both are straight, 1-dimensional subsets of the space of all line-dot juxtapositions.

Note, in fact, that the second "regularity," no less than the first, meets all the usual canons of a good scientific hypothesis: it is simple with respect to the number of data, it is unlikely to be corroborated by accident, and so forth. Indeed, once discovered, it seems perfectly credible; the problem is that our perceptual apparatus does not readily discover it. Unlike the first example, it seems in some sense critically out of alignment with the first-order hypotheses that our categorization system entertains—perhaps because it depends on a virtual structure, the imaginary line, that lacks a direct structural basis. In Ch. 2 will sharpen this idea somewhat by proposing essentially that world regularities can be picked up correctly as categories only if their geometric analogs coincide in a particular well-defined way with a kind of "category coordinate frame," an organizational structure that will be defined by means of special values along certain kinds of parameters. In essence, this condition guarantees that the categorical hypotheses we as observers chose, beyond being simply consistent with observation, are consistent with some sort of causally coherent generative story; or, putting it the other way, we do not consider category interpretations for which we have no ready coherent account.

1.2. Some Background

By way of fitting these problems into context of prior research, the next five sections consider themes of past investigations that serve to situate the theory proposed in this thesis. Some of them were touched on obliquely in the section above. Needless to say, these five are neither exhaustive nor entirely independent of one another. Broadly, Secs. 1.2.1-1.2.3 describe research on categorization per se, while Secs. 1.2.4 and 1.2.5 describe several ideas that have recently arisen in formal perceptual theory. This thesis may be seen as an attempt to bring these mathematical ideas to bear on the categorization problem in a context of perceptual inference.

1.2.1. Inductive nature of categorization

The philosopher David Hume is usually credited with being the first to point out in a clear fashion that reasoning from finite examples to general rules ("induction") cannot be guaranteed to be correct in the strict manner that one associates with deductive reasoning. Hume pointed out that believing such inductions—as people consistently do—is tantamount to relying on the "Uniformity of Nature": our simple inductions turn out to be true because the world is basically a simple place. (See Sober, 1975, and Quine, 1963, for good discussion of some subtleties of this point.). The basic dilemma has been more recently sharpened by Charles Peirce (amplified by Keil, 1981) who proposed that the route to choosing an induction from among all possible is in imposing constraint on the space of hypotheses we consider.

Today the need to constrain inductive hypotheses finds its most explicit application in the study of the acquisition of linguistic rules by children (see Gold, 1967, and Osherson, Stob, & Weinstein, 1986, for perspectives on how formal constraint on the inductive hypotheses can be employed to render infinite sets, such as human languages, learnable from finite examples).

In the realm of visual categories, the problem is perhaps best illustrated by Bongard's (1970) list of visual induction problems (see Richards, 1988, for a discussion). In each "Bongard problem," six pictures on the left are contrasted with six on the right, with the reader invited to figure out what the left pictures have in common as contrasted with those on the right (Fig. 1-6).

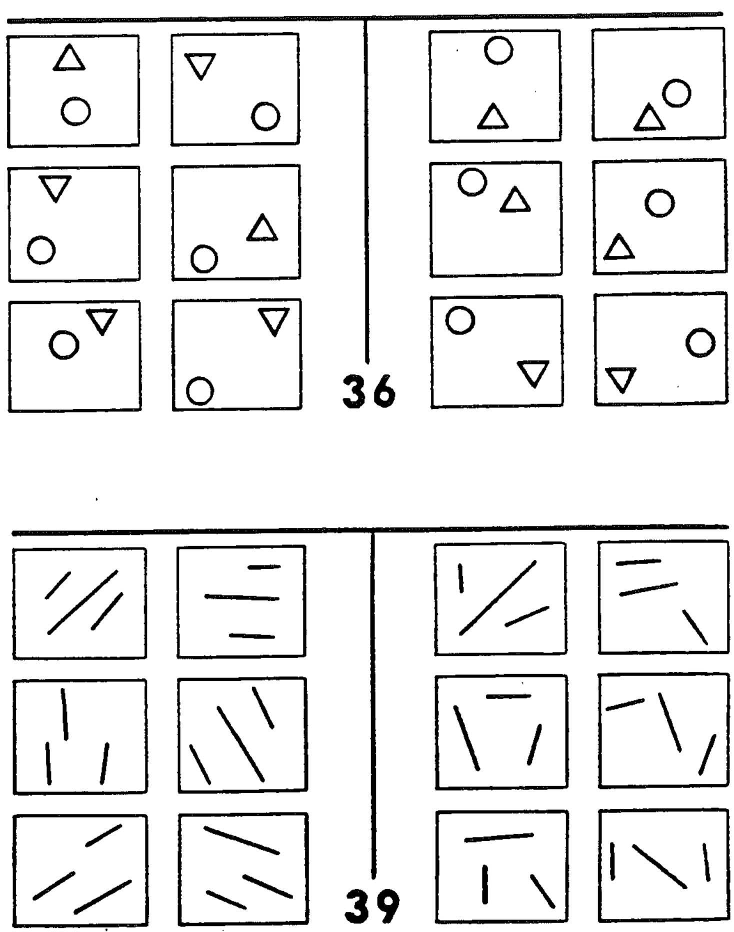


Fig. 1-6 Two sample "Bongard problems" (from Bongard, 1970).

While some of Bongard's problems are trickier than others, all share a compelling quality once the solution has been found. But (as Hume pointed out) the relationship between these finite collections and the infinite classes of which human observers judge them to be subsets—elevating particular infinite extensions above all the other potential extensions that are equally consistent with observation—is mysterious. In Bongard problems, we can expect a compelling solution because we know another person (Bongard) put one there. In a natural categorization framework, it is left to naturally-occurring processes, whatever superficially unfamiliar form they may take, to put the orderliness and regularity into the problem.

1.2.2. Classical vs. cluster categories

Research in cognitive science about the nature of human category representations has exhibited a central tension between two seemingly inconsistent views of the ontogeny of categories: one, in which category membership can be decided by a well-defined, in-or-out rule invoking necessary and sufficient criteria for membership (what Smith & Medin, 1981, call the "classical" model), and another in which category membership is seen as a graded function of distance from some maximally typical category member (the "cluster" model). In the cluster conception, research naturally centers around the structure of the parameter space in which typicality distance is measured, and the gradient with which typicality falls off with distance.

Cluster model of human concepts are generally motivated by the prevalence of graded typicality judgments in human data, first noted by Rosch (1973; see also Rosch & Mervis, 1975). For a wide variety of object domains, human subjects are willing to assign an apparently continuous measure of category membership to candidate objects. These typicality or membership scores, in turn, can be formalized as distances in some psychological parameter space away from some optimal examples of the category, sometimes called a "prototype." Shepard (1987) has investigated formal aspects of the regions of parameter space (which he calls "consequential regions"; see above) that correspond to coherent sets of objects, i.e. regions across which object properties tend to be relatively stereotyped. He showed that typicality gradients within such regions take an exponential form in a wide variety of object domains, while the shape of the regions (e.g. circular or rhomboid) bears a natural relationship to the type of distance metric employed (e.g. Euclidean or city-block).

These cluster model of categories leaves some theorists dissatisfied, however, because it fails to account for the intuition that members of some natural categories genuinely belong together due to some common property (such as the atomic number of the element gold). In this view, sometimes called "essentialism" (see Schwartz, 1979), the category is in effect dictated by the scientific laws governing the domain in question, even though this fact may be obscured by the presence of many ancillary properties that are present in varying degrees. Medin & Ortony (1988) have argued that humans recognize this as a basic organizing principle of categories, in effect favoring some essential defining property (even an unknown one, or one known only by experts) over typicality in deciding membership.

The conceptual minefield relating classical and cluster categories was perhaps brought into clearest relief by the experiments of Armstrong, Gleitman and Gleitman (1983), who replicated the cluster and typicality effects of Rosch using what seem to be necessarily classically-defined categories such as "prime number." The need to reconcile classical definitions (or at least something approaching them) with cluster concept data has thus often been felt acutely (see for example Cohen & Murphy, 1984, Osherson & Smith, 1981, and Nosofsky, 1991). A kind of compromise position has appeared in many sources that distinguishes between a "core" concept, bearing a strict definition, and a peripheral "identification procedure," that yields graded typicality effects as an epiphenomenon. Pinker & Prince (1991), intriguingly, find evidence for two such interacting systems in memory for the past tense forms of verbs, suggesting that the dichotomy may serve as a wide-ranging organizing principle underlying concept representations in a wide variety of cognitive domains.

Statistical clusters. The phenomenon of cluster concepts in human categories finds a natural analog in the statistical problem of finding "good" clusters of points in the space, generally meaning clusters that have low variance compared to the distance separating them (see, for example, Chambers, 1982 or Farlow & Kleiner, 1984). Although in many ways this problem is fairly tractable (for instance there is a known Bayesian upper bound to the performance of any clustering algorithm; see Watanabe, 1985), clustering algorithms notoriously suffer from several characteristic technical faults, as compared to human impressionist clustering of a plot of the data points.

Broadly speaking, clustering techniques have trouble with the idea of dimensionality. That is, natural data-generating processes often seem to have a lower "intrinsic" dimensionality than the dimension of the space in which they are encoded, reflecting some smaller number of actually varying parameters internal to the process. While human observers can often spot, simply by inspection, distributions that would be better modeled in a lower dimensional space, the statistical inference involved in doing so automatically is notoriously difficult (the so-called "curse of dimensionality"; see Fukunaga, 1982, and Jain & Chandrasekaran, 1982). As a result, human judgments about the proper decomposition of some types of distributions is often wildly at odds with the best-fit clusters, in effect because human models of clusters are at once both more flexible and less sensitive than the statistical concepts in use. Matthews & Hearne (1991) recently proposed to sidestep this problem by actually eliminating the distance metric concept from the parameter space, but it is not clear how their technique generalizes.

In a way, it can be argued that the underlying problem in these statistical techniques, just as in the cluster concepts of cognitive psychology, is that it is not entirely clear in the first place why we would expect either data points or objects in mental spaces to behave categorically at all—that is, why we would expect some boundaries in these parameter spaces to mark sharp, reliable, and meaningful boundaries between systematically disparate collections of objects. Again, what is needed is a category theory to enforce and at the same time explicate the tendency of real-world parameter spaces to exhibit "natural" seams.

1.2.3. Categories as correlations of properties in the world

Recently, the Principle of the Uniformity of Nature has turned up in accounts of categories and concepts in a number of guises, all of which recognize in one way or another that cognition and categorization depend on the world being a highly structured place. It is now a common though not uncontroversial view in cognitive science that the existence of categories in the world has something to do with the tendency of interesting properties of objects to covary with one another; that is, for some properties to occur with characteristic regularity and with greater consistency in the presence of certain other properties (see Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976; Smith & Medin, 1981; Quine, 1985).

The regular and predictable quality of the world manifests itself, *inter alia,* in the co-variation of distinct, logically orthogonal properties. Thus, there are things which are red, crunchy, and round (*apples*) and other things which are yellow, soft, and banana-shaped (*bananas*). Critically, the complementary cells of the implied matrix are not filled, or at least, are less populous: while one can *imagine* red, crunchy, cylindrical things very readily—meaning that these terms are logically orthogonal in our conceptualization of them—one tends not to encounter them as often. In this view, categories amount to equivalence classes of co-implying features. In a maximally intercorrelated world, an (arbitrarily large) number of features are all equivalently predictable from the presence of any one member. In a strict view, these equivalence classes are exactly what is meant by the term "prototypes."

One apparent paradox posed by the view that categories are loci of covariation of properties is that one can hypothesize categories, as shown above, from even *single* examples: can we justifiably infer *covariation* without witnessing *variation*? Under certain circumstances, the answer turns out to be yes. One of the goals of this thesis is to resolve the apparent paradox, and show how even one example can be taken as a reliable cue that some set of properties covaries with one another—so long as the category inferences are organized in a particular way, and the world in which the categorizer finds itself has a particular kind of structure.

In perceptual theory, the idea of the Uniformity of Nature has materialized in the notion that inference of world properties depends critically on structure, regularity and over-constraint in the physical world (see Marr, 1970, as well as Witkin & Tenenbaum, 1983 and 1986), a view expanded into more generalized property contexts and category models in the Natural Modes hypothesis of Bobick (1987), formalized in Jepson & Richards (1992). Richards & Bobick (1988) have shown that the notion of world regularity, when fleshed out by even a surprisingly short list of known physical, biological and mechanical (etc.) lawful correlations (e.g., larger animals make lower-pitched sounds), makes it possible for a perceiver with some underlying knowledge of its environment to convert a few simple, observable facts about an unknown object into an extremely informative categorization. This thesis, in a sense, looks at the obverse issue; rather than enumerating particular, concrete, regularities, the idea here is to enumerate the qualitatively different mathematical forms that models of regularities can take, as well as constraints that our perceptual models must obey, in order to successfully identify categories that genuinely obey a world regularity.

1.2.4. Genericity and dimensional issues in perceptual inference

Much recent work in perceptual theory has highlighted, sometimes implicitly, the role of dimensional differences between some solution space and its false target space, which guarantee that the latter will always be a measure-zero subset of the former. Numerous 3-D structure inference schemes, for example, allow correct recovery from all viewpoints except *non-generic* ones, i.e. viewpoints that line up just perfectly with object structure in some way, such as when the plane containing one side of a cube also contains the viewing direction. In various forms, the notion that schemes for inferring structure must never depend on such accidental arrangements—because they are so atypical—is called a *generic viewpoint constraint.*

Genericity constraints in more general form appear whenever some inference applies in all but a systematically atypical set of configurations, in particular when this false-target set is actually a lower-dimensional subspace embedded in the general solution space (as is typically the case for non-generic viewing orientations). Various kinds of "non-accidental" features (Binford, 1981, and Lowe, 1985 and 1987) fall into this class as well; these are structural features (such as parallelism in the image) that are unlikely to occur unless a related distal feature (such as parallelism in the world) also occurs. Reuman and Hoffman (1986) have generalized this idea somewhat, showing in general how canonically the inference of a world feature depends on the vanishing probability of false targets, and the idea of genericity has emerged as central in the more comprehensive theories of Bennett, Hoffman & Prakash (1989), Clark & Yuille (1990), and Knill & Kersten (1991).

Since perceptual genericity invokes the topological relationship between spaces and their lower-dimensional subspaces, the inferential arrangement can be clarified and amplified by a more comprehensive and abstract analysis of the this kind of relationship (Poston & Stewart, 1978). In the theory proposed in this thesis (see also Jepson & Richards, 1991,1992; Feldman, 1991b, 1992; Feldman, Jepson & Richards, 1992; and Richards, Feldman & Jepson, 1992), *modal non-genericities* (meaning special configurations that we consider structurally important, and occur commonly in the world) turn out to be a central organizing principle.

The idea here is to apply the notion of genericity to category models, so that members of a given category can typically (i.e, generically) manifest an entire cohort of predictable properties in common with each other. In any given category model, wherever a central, structure-supporting parameter collapses (creating a modal non-genericity), the model itself collapses to a submodel, i.e. a recognized subcategory.

1.2.5. The causal story or generative process model

In a series of investigations on the interpretation of shape, Leyton (1984, 1988, 1989,1992) has emphasized the idea of a generative process model, a kind of account of the causal history that gave rise to visible structure, as goal of inference. This idea plays a central role here in the idea of a category model (see also Feldman, 1991b). When parameters are interpreted as the magnitudes of generating operations, the non-genericities—the places where a parameter collapses—become structurally important: these are transitions between discretely different process-histories (cf. Bobick & Richards, 1986).

The need for ascribing a process history to a shape, or here a generative model to a category, parallels some recent trends in category theory as well, in which some sort of causal story, i.e. a series of comprehensible, interpretable events, is needed to underwrite a category inference. Murphy & Wisniewsky (1989), for example, inquiring whether subjects actually attend to property covariations in their observations, found that they did so only for those property pairs for which they expected a correlation and for which they had a ready coherent account, in effect ignoring even perfect correlations between properties for which they could not conceptualize a causal relationship. The idea of the central explanatory story plays a role in the category theory of Murphy & Medin (1985) as well, as of course in some perceptual theory (e.g. Rock, 1975). In the theory proposed in this thesis, since each category model is itself a kind of causal story, only non-genericities for which we have some sort of a coherent structural account (i.e., only modal ones) support category inferences.

In earlier work (Feldman, 1991a), I examined a single, particular pattern model—the smooth curve, as perceived in a field of dots in the plane—in an attempt to characterize the formal properties that make this type of pattern a particular good and coherent one to extract from an observed field. In such a curve, the dots exhibit a certain kind of "constrained covariation" in that positions of the constituent dots, rather than being mutually independent, are largely predictable from one another. Here, I expand the scope of the investigation to arbitrary pattern classes; and though the reasoning is more elaborate, and the basic types multiply both in number and in complexity as higher dimensions and more types of parameter are considered, the reasoning is similar. Just as curvilinear collections of dots in the plane all non-accidentally fall on a lower dimensional curve in a higher dimensional space, in the analysis that follows, coherent object categories will always amount to collections of objects whose defining parameters are constrained to obey some smooth, lower-dimensional constraint in a some kind of higher-dimensional embedding parameter space. The formal analysis of Chs. 2 and 3 simply seeks to enumerate the qualitatively different ways in which this can happen. Ch. 4 works the theory out by simulation for several very simple model classes, and Ch. 5 seeks corroboration of it in human subjects' category inferences. Ch. 6 summarizes the research presented, arguing that human categorization judgments can be viewed as well-defined inferences designed to orient our knowledge of the world around meaningful interpretations.

- Chapter 2 -

Regularities and Lattices of Category Models

2.0. Chapter preview

This chapter lays out the basic definitions and machinery (model classes, category models, and lattices) that will be employed throughout the thesis.

Sec. 2.1 develops a definition of a "regularity" in the world: a manifold of positive codimension in some embedding space of world parameters. This puts into a single general formal framework the idea of a causally coherent collection of objects, that is, a collection of objects that are interpreted as having all been produced by a common generating process. Such an arrangement represents some degree of stucturedness in the world, with respect to a more "naïve" (less constrained) model, corresponding to the larger configuration space in which the constraint manifold is embedded.

Sec. 2.2 develops a definition of a "category model," motivated by the discussion in Ch. 1. It is then shown how regularities, in the sense defined in Sec. 2.1, naturally give rise to category models, each corresponding to a collection of objects obeying a common constraint and generated by a common world process. Briefly, a category model is defined as a set of parameters along which an object in a category can translate freely (i.e. parameters that are intrinsic to the model), and another set of parameters on which members of the category must have some fixed value. A *genericity* constraint restricts objects within each category model to points that are mathematically typical in it, relegating non-generic objects to a discretely different, lower-dimensional model.

Sec. 2.3 shows how a regularity in the above sense gives rise to a *lattice* of category models, each corresponding to an equivalence class of objects all of which are generic in the same subspace of the constraint manifold. Each category model naturally has a *dimension* (i.e. the number of parameters along which it is generic), and a *codimension* (the number of dimensions along which it is non-generic, or the difference between its dimension and the total number of dimensions in the space). The codimension in effect represents the degree of "constrainedness" of the category model. The category lattice, putatively, contains all the sanctioned categories the observer can entertain as hypotheses about observed objects; that is, it enumerates and interrelates all the category hypotheses that the observer takes to be causally meaningful.

Finally, with all the basic machinery for constructing and organizing categorical hypotheses in place, Sec. 2.4 derives the inference procedure whereby an observed object is actually identified with one of the available hypotheses. The requirement of genericity forces the observer to adopt a kind of "minimum principle," in which the lowest-dimensional hypothesis on the lattice is selected. It is argued that use and motivation of the minimum principle in this connection avoids some of the problems (principally problems of justification and motivation) many authors have noted in other contexts.

2.1. Regularities as manifolds: the category continuity hypothesis

This section introduces a single, general model of a regularity in the world that will remain in force throughout the thesis. Of course the idea that such structure actually exists in the world must at bottom be viewed as an empirical hypothesis, about which evidence drawn from the various special sciences might be brought to bear (see Bobick, 1987, for a discussion of this point, as well as evidence that such structure does in fact exist in natural domains). Of course, "order" may manifest itself in many different forms: compare the form assumed here with those in Markus, Muller, Plesser & Hess (1987), and in Marroquin (1976).

The remainder of this chapter will show how the hypothetical presence of structure and regularity in the world, in the form to be described, gives rise to the existence of coherent categories—collections of objects that share common properties, in virtue of having arisen out a common, regular constraint. When human subjects make guesses about the generating category of observed objects, the argument goes, they are in effect assuming that observed objects obey a regularity of the kind to be defined, and then simply guessing *which* regularity.

Our central model of a single, unary category, produced by a coherent, causally uniform generating process, is a smooth manifold in some suitable parameter space (see Saund, 1987). That is, imagine that some set of objects in the world can be described by two parameters ϕ_1 and ϕ_2, i.e. objects in this space are points in a space $\Phi = \langle \phi_1, \phi_2 \rangle$. Imagine further that for all the objects in this collection, the measurements ϕ_1 and ϕ_2 always fall on some smooth curve Ψ through Φ (Fig. 2-1). That is, ϕ_1 and ϕ_2 obey some smooth relation $f_\Psi(\phi_1,\phi_2) = 0$.

Rather than interacting orthogonally, ϕ_1 and ϕ_2 are behaving as if they influence one another in an extremely regular way. As a result, the points in the category can actually be reparameterized using fewer dimension: in this case, just the one dimension of arclength (or some function of it) along the smooth curve.

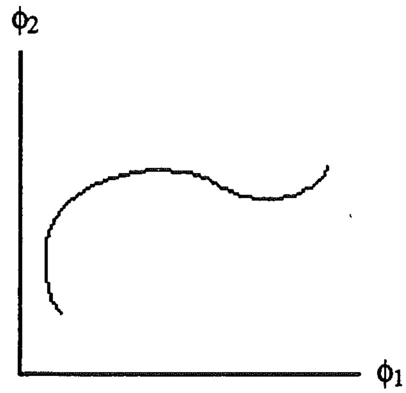


Fig. 2-1. A constraint manifold in $<\phi_1,\phi_2>$-space. Each point on the curve denotes an object that obeys the regularity that defines the category.

Example (i). As a very simple example, consider a set of objects consisting of two line segments in the plane with a common endpoint. Rather than the two lengths being independent, though, say that there is a constraint: one segment is always (say) 1.37 times the length of the other. Call this category of V-like objects $V_{1.37}$. Some examples drawn from $V_{1.37}$ are given in Fig. 2-2.

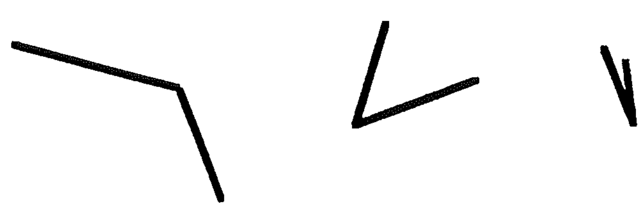


Fig. 2-2. Some examples of the constrained category $V_{1.37}$.

Each object may be expressed as a point in a Cartesian frame fixed to the smaller of the two segments, which is assigned unit length and zero orientation. Then the endpoint of the other segment falls at a location (x,y) in this frame (Fig. 2-3)

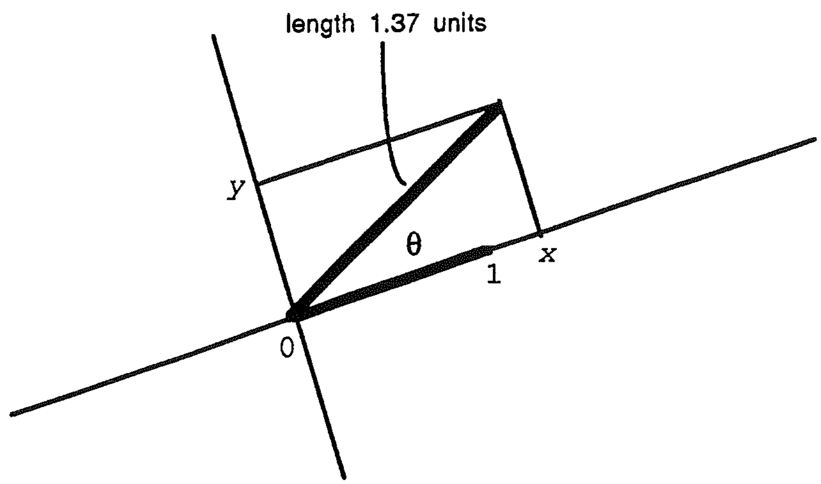
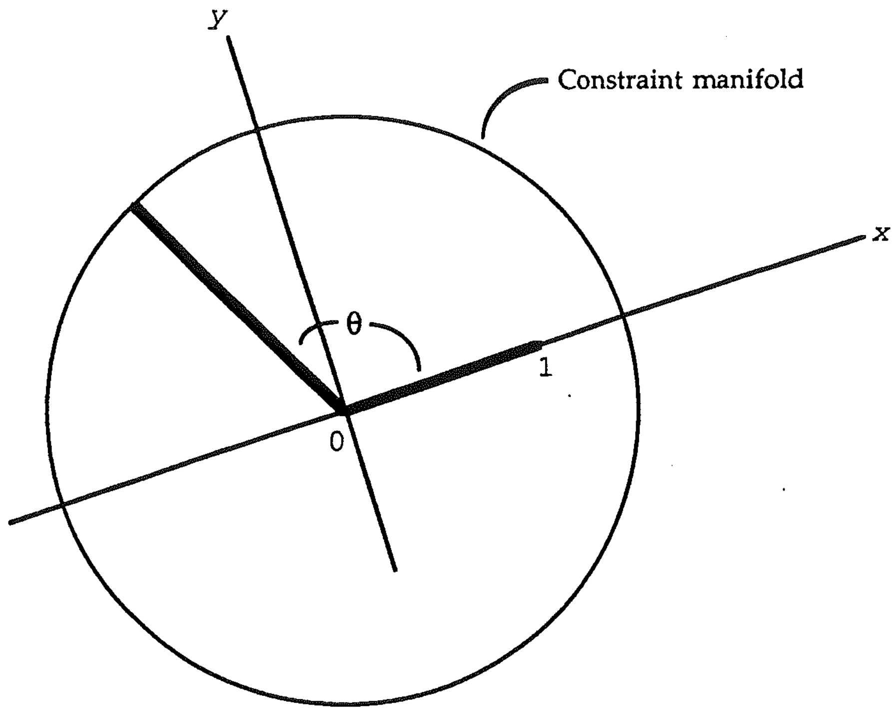


Fig. 2-3. A parameterization of the category in question that fails to take into account the constraint in force.

Then, because of the constraint, the collection of these objects, rather than filling out the Cartesian space, is entirely restricted to a circle with the equation $x^2 + y^2 = 1.37^2$ (Fig. 2-4). That is, the entire collection fits on a particular lower-dimensional curve through the space.



Notice that this means that one can think of the collection as being constructed by taking one object in the set, some point on the circle, and applying some transformation to it that carries it strictly along the circle (i.e. in this case a transformation that preserves the leg-length ratio). Given the definition of the object, what kind of transformation would do this? (As always the similarity transforms, translation, rotation and scale, are excluded.) Clearly, the answer is a transformation that changes only the angle between the two segments, since that is the only shape property that changes around the circle. It may be some *function* of angle (i.e., as we move at a constant rate around the circle, the angle does not change at a constant rate), but angle is the "essential" variable. Angle can be said to be the "natural" parameter expressing what actually does change among objects that obey the constraint.

Note the similarity between this example and the line on the dot from Fig. 1-1: in both cases there is a reduction to a single dimension of allowable variation within the category. This suggests a natural form for category models, which will be explicitly adopted in the next section: a collection of objects that are generated in a regular manner by variation along some recognized parameter of structure.

Generalizing, this situation captures a common form of structural regularities in a wide variety of contexts, expressing in effect a causal relationship between the value of one property and the value of another. This idea can now be cast as a kind of regularizing principle for observers' category inferences, a guiding principle that biases the models observers construct towards those that are liable to correspond to genuinely causally coherent world categories (cf. Ullman & Sh'ashua, 1988, who develop a scheme for picking out salient patterns from fields of visual items, i.e. groupings that seem particularly likely to be important and deserve attention).

This idea is called the "category continuity hypothesis."

Category Continuity Hypothesis:

Causally coherent categories in the world may be expressed as smooth manifolds of positive codimension.

This idea underlies all the machinery developed below for the construction of category models and the inference of category models from examples.

Two critical properties of constraint manifolds, namely *smoothness* and *positive codimension*, can now be described in more detail. These properties as are best regarded as axiomatic, in that they simply crystalize something about what we intuitively mean by a "natural category." Whether or not world categories (as modulated by the actual causal mechanisms in the environment) are actually or objectively "smooth" or "continuous"—whether the category continuity hypothesis is really "true" or "false"—is bracketed as a fine philosophical question not to be answered here. Rather, it is *assumed* that continuous categories, if conceptualized in this particular way, make a good model for causal coherence; speculate that human observers incorporate this assumption into their category inferences; work out the ramifications that incorporating this assumption has on the structure and organization of category inferences; and, later, assemble empirical evidence that human subjects' category judgments have the structure and organization in question—corroborating the role of the category continuity hypothesis, "true" or "false" notwithstanding, in human categorization.

(i) *Continuity and smoothness.* We assume that a world category is some *connected* subset Ψ of the parameter space Φ: this means that for every point $x \in \Phi$ that obeys the regularity (falls in Ψ), there is an entire neighborhood of points arbitrarily close to x that also obey the regularity (are also in Ψ). This axiom satisfies an intuition that objects' membership in a real-world category does not depend in an entirely idiosyncratic way on their structure, but rather that their membership is grounded in their structure via some transparent causal mechanism; equivalently, that the category in question contains objects produced by some *causally coherent* generative process. Focusing on this kind of topological coherence, rather than on more specific aspects of the shape of the category, frees us from worrying about the exact meaning of the various parameters ϕ_1, ϕ_2, etc., since the coherence of the category is preserved under a wide range of transformations of the parameter space—in particular, "homeomorphisms", or continuous, reversible changes in variable.

We further assume, as seems safe for natural categories, that the category is (locally) *smooth* (i.e. has derivatives of arbitrary order). This idea crystalizes the idea of causal coherence in a particular clear way, in that it means that the best local linear approximation of the category, i.e. the tangent line (or plane, etc.), changes orientation continuously. This means that the best simple approximation of the category at one point and that at a neighboring point are always infinitesimally close to one another. This matches our intuition that objects in a "natural" category all embody the *same*, uniform, constraint. Analogously, dots in a plane whose positions obey a smooth constraint curve "pop" out to the eye, appearing to a human observer to constitute a coherent group (see Feldman, 1991a). A smoothly embedded connected subset is normally called a *manifold*, hence our term "constraint manifold."

Continuity and smoothness are highly non-trivial, in a mathematical sense. Most point sets are not connected (much less smooth): being connected is in a strong sense *infinitely* unlikely[1], so even this simple axiom ascribes an enormous amount of structure to the world. Less abstractly, continuity underwrites a certain kind of category *productivity*: for any two points (objects) in the category, there is a path connecting them such that all points

[1]It has infinite codimension in the space of all point sets.

lying on the path are also in the category.[2] In order to conclude the stronger condition that all points with intermediate parameter values are also category members (*convexity*[3]) as well, for which, unfortunately, there is no a priori justification. Convexity and connectedness are equivalent only in one parameter dimension, which, while clearly not the general case, is an important one: in this case, any two points in the category would allow the automatic generation of an endless stream of new examples from the interval between the two known examples.

(ii) *Non-zero codimension*. We would like to express the idea that the category membership rule excludes members in a non-trivial way, as is presumably typically the case for natural categories. In the example given above, notice that since the category had one fewer dimension than the $<x,y>$ space in which we drew it, objects picked randomly from this space were extremely unlikely to satisfy the category constraint by purely accident. The constraint manifold pictured is said to have *codimension* 1, meaning that the difference between the dimension of the embedding space (2) and the dimension of the curve itself (1) was 1.

In general, the codimension (see Poston & Stewart, 1978, for a discussion of geometric aspects; or Jepson & Richards (1992) for application to perceptual theory) represents the number of independent conditions that points on the curve must satisfy. That is, each condition removes one degree of freedom from the data set. Naturally, the codimension is only meaningful with respect to some embedding space, here $<\phi_1,\phi_2>$, which can be thought of as a kind of "naïve" model of the data, like our $<x,y>$ space above, in that it encodes the positions of the points in a manner that fails to take into account the regularity in force.

Hence within the given naïve configuration space Ψ, the embedded constraint has positive codimension:

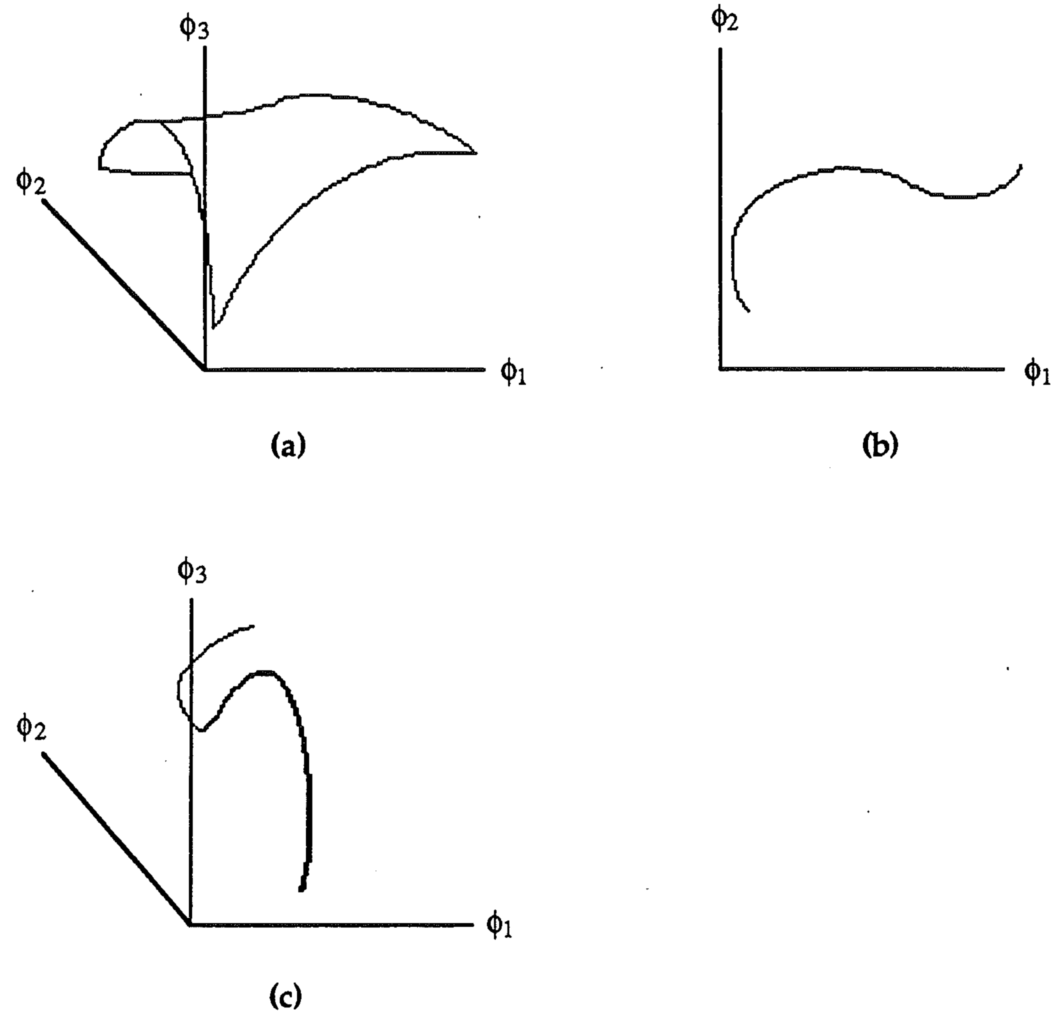
$$\text{codim}\ \Psi\Phi \geq 1.$$

---

[2]This may be seen, locally, as an application of the so-called"intermediate value theorem" from high-school calculus.

[3]Convexity is commonly assumed for well-behaved data sets. Though it seems reasonable when categories are conceived as ovoid clusters, it makes little sense when categories are conceived as curving manifolds.

As discussed above, a single functional condition on points in n-space produces an n-1-dimension surface; for instance, one condition on points in 3-space produces a 2-D surface of some kind,[4] two conditions produces a 1-D space curve, and so forth[5]. (Fig. 2-4).



[4]In fact, the term *surface* is often generalized to mean "manifold of codimension 1".
[5]Notice that classifying manifolds by their codimension gives some useful information not provided by their absolute dimension. For instance, a codim-1 surface always divides the embedding space into two regions,; two codim-1 surfaces (e.g. two line in the plane, two surfaces in 3-space) typically intersect in a codim-2 object; on the other hand two codim-2 objects (e.g. points in the plane, curves through 3-space) typically do not intersect at all, and so forth.

Fig. 2-4. Constraint manifold of various dimensions and codimensions.

(a) *dim* $\Phi = 3$, *dim* $\Psi = 2$, *codim* $\Psi_\Phi = 1$.

(b) *dim* $\Phi = 2$, *dim* $\Psi = 1$, *codim* $\Psi_\Phi = 1$.

(c) *dim* $\Phi = 3$, *dim* $\Psi = 1$, *codim* $\Psi_\Phi = 2$.

This axiom guarantees the non-triviality of the regularity, in that a manifold of non-zero codimension normally has measure zero (takes up infinitesimal volume) in its embedding space, meaning that the chance of a point in the larger space falling on the manifold by accident is about zero (up to some resolution). This corresponds to our intuition that a substantive constraint is a stringent one—one that is not likely to be satisfied by accident. Hence objects that appear to be members of the category are likely to be so "genuinely," i.e. because they obey the same causal forces affecting all other members of the category, rather than by accident. Hence objects interpreted as obeying a regularity can be reliably expected to exhibit whatever other properties are associated with category members.

Example (ii). Now consider another example of a constrained class of V's, V_1. Again objects are constructed from two line segments sharing a point, and again the line segments exhibit a constraint on their relative length, but this time instead of the arbitrary-seeming 1.37 ratio, the two segments are required to be of *equal* length (ratio = 1 Fig. 2-5).

Examples of V_1

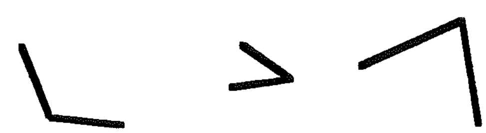


Fig. 2-5. A better behaved category?

Again, the regularity can be drawn as a circle through the <*x*,*y*>-space centered on one segment (Fig 2-6). This time, though, the circle passes through the endpoint of both segments.

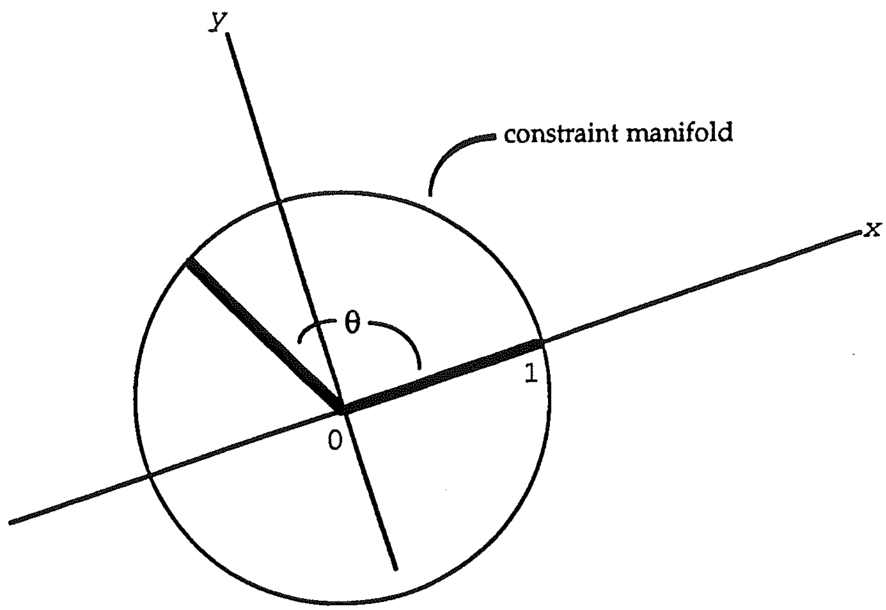


Fig. 2-6. A better-behaved constraint manifold?

Formally, this category may seem analogous to Example (i). Notice, however, that it seems much more plausible in some sense than the earlier one—a fact reflected in the suitability there of the term "arbitrary-seeming." One intuition behind the greater plausibility here is the idea that it is easier to construct a "generative model" for the equal-length category. For instance, V_1 can be imagined to have come about from a population of uniform rods, out of which random pairs have been affixed to each other at an endpoint. To construct $V_{1.37}$, by contrast, some agent would have to take a population of rods of length 1 and 1.37 and cull from it only matched pairs—an inherently more arduous process. The extension of one bar to a particular ratio of the other one, in our attempt to conceptualize it, always seems to require an extra process of some kind—an additional generative operation. By contrast, the 1:1 ratio seems a degree more plausible as a model, and a hypothesis worthy of being entertained with greater care. This notion of a generative model will be formalized in the next section.

Key idea. The critical inferential idea of this chapter, in a nutshell, is that we as observers have a certain inventory of concepts ready-made with which to construct generative models of sets of sample objects we observe: and that these models, and putatively only these models, serve as routes to lower-dimensional reparameterizations of constraint manifolds. That is, of all the possible ways to reparameterize a constraint manifold to its intrinsic dimension, we choose one that corresponds to a generative model that we construct out of available concepts. If no available generative model will serve to reparameterize, we miss the regularity.

In this way, critically, our categorization inferences are oriented around regularities that we *expect* to happen, and to which we attach special meaning—namely, that each one connotes a distinct generative model or causal story that we have some prior framework for comprehending. This idea will be elaborated in Sec. 2.2.

2.2. Category models arising from a regularity

This section formalizes the notion of "generative model" in a simple form, and show how this form can be used as an expression of a reparameterization of a constraint manifold to a lower dimension. This leads to a convenient notation for "category models", in which each model is written as a sequence of algebraic operations taken from a fixed inventory. The critical *genericity* constraint is then introduced, namely that each object in a given generative category can be expected to be formally generic in that category. This constraint underwrites the inference that non-generic objects are actually members of different categories—namely, lower dimensional ones. This in turn allows the construction of the lattice of category models for a closed model class in the next section, which will then be automated (Chs. 3 and 4), and used to construct predictions for categories induced by human subjects (Ch. 5).

2.2.1. Generative models

Above it was argued that categories of related objects can often be thought of as having just a few dimensions of structural variables that can vary from object to object, such as "angle" in V_1. We now propose to think of these parameters as the *dimensions* of some *generative process* or procedure.[6]

To begin with, assume that there is a primitive object of some kind, which acts as a strarting point. For V_1, it might be the unique "right" V, or

[6]Another motivation for parameterizing a class of objects may be found in Shepard & Cherniak (1973).

some arbitrary seeming choice like the V with a 37° angle; for the moment the choice does not matter. Next, we apply some one-parameter deforming operation to the primitive object — stretching, shearing, and so forth. In V_1 the operation would be opening or closing the V (increasing or decreasing the angle). If the operation is chosen suitably, all members of the category can be constructed in this manner.

Leyton (1984; see also Leyton, 1991) has proposed that simple geometric objects can be modeled this way using a fixed algebraic sequence of group-theoretic transformations, and suggests that human subjects use such a scheme to "reference" more complex objects to simpler objects, that is to conceptualize the more complex objects as deformations of the simpler ones (Fig. 2-7). Leyton argues that this nested algebraic generative sequence acts in the observer's representation as a model of the *causal history* of the object. That is, the object's shape is represented by an interpretation of that shape appears to have come into existence.

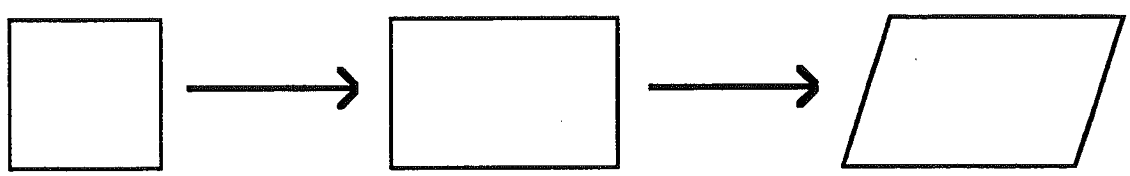


Fig. 2-7. In Leyton's (1984) example, a parallelogram is seen as a deformation of a rectangle, which in turn is seen as a deformation of a square.

Notice critically that conceptualization of (say) a parallelogram as a slanted, stretched square—as opposed to simply representing its shape "in and of itself"—though intuitively compelling, is by no means logically necessary. Rather, this conceptualization amounts to a *theory* about the causal origins of the rectangle (see also Leyton 1989 on this point).

That is, if the primitive object is denoted x_0, and the generative operations $g_1 ... g_n$, we can think of the resulting object x as

$$x = x_0 \cdot g_1 \cdot g_2 \cdot ... \cdot g_n \cdot$$

This sets up the possibility of constructing 2^n different generative sequences, corresponding to each distinct subset of the g_i's. This idea will be the basis for the enumeration of category models in Sec. 2.3. The next section will recast these transforming operations as translations (unit vectors) away from an origin point (the primitive object) through a linear space corresponding to the category parameter space.

In addition to such simple mathematical operations as stretching and shearing, object shapes can be varied by using simple translations, rotations, and scale changes (operations that are normally treated as invariant when they act on an entire object) acting on just one part of an object with respect to the main body, as in V_1 and $V_{1.37}$, in which one leg was rotated with respect to the other leg. Thus one part can be turned with respect to the main axis, or lengthen one part—corresponding to, say, bending an arm, lengthening a neck, and so forth. Transformations of this type will be used in all the examples later in the paper. Tapering, bending and twisting may also performed by simple matrix multiplication (Barr, 1984), and hence make for mathematically convenient generative models. Superquadrics (Pentland, 1986) constitute another parametric family of abstract shapes that make good models of natural forms.

A number of simple mathematical operations have been shown to closely model natural growth of various kinds, which is convenient for the identification of these mathematical operations with natural processes whereby objects' shapes are determined. In a classic inquiry into the geometric properties of natural form, Thompson (1917/1952) noted that the shape change induced by gross size changes between structurally similar animal species often takes the surprisingly simple form of an affine transformation. More recently, Mark, Shapiro, & Shaw (1986) have noted that the mathematical operation cardiodal strain, when carried out upon head-like pictures, is perceived as cranial growth.[7]

[7]Interestingly, they further note that the same operation is not perceived as any kind of growth at all when it is carried out on robot heads made up of straight line segments, rather than on natural-looking line drawings of heads. This accords nicely with the view, elaborated later in this paper, that the chosen parameterization closely follows cues from the observed object at hand.

2.2.2. Model classes as reparameterizations of the constraint manifold

As discussed above, constraint manifolds can be reparameterized in fewer intrinsic parameters than are contained in the naïve embedding space. It is now shown explicitly how *category modes* arise in a straightforward way from such a reparameterization. Later it will be argued that human observers' preference of certain category models over others derives from their preference of certain choices of reparameterization vectors—namely, those that correspond to a generative model constructed from a fixed inventory of object-generating operations.

Since a 1-D constraint manifold has by definition has only one dimension, the choice of new parameters with which to reparameterize it intrinsically is forced to be some function of arclength. In higher dimensions, though, there are an infinity of a priori choices. Starting from some arbitrary origin point **0** on the manifold, each new parameter can be thought of locally as a direction, say a unit vector lying within the manifold: the *i*-th unit vector is notated as $\hat{t}_i$. In general, then, any two $\hat{t}_i$ directions that are *transverse*—cross each other in a "typical" fashion, i.e. independently—will serve to reparameterize the manifold (Fig. 2-8).

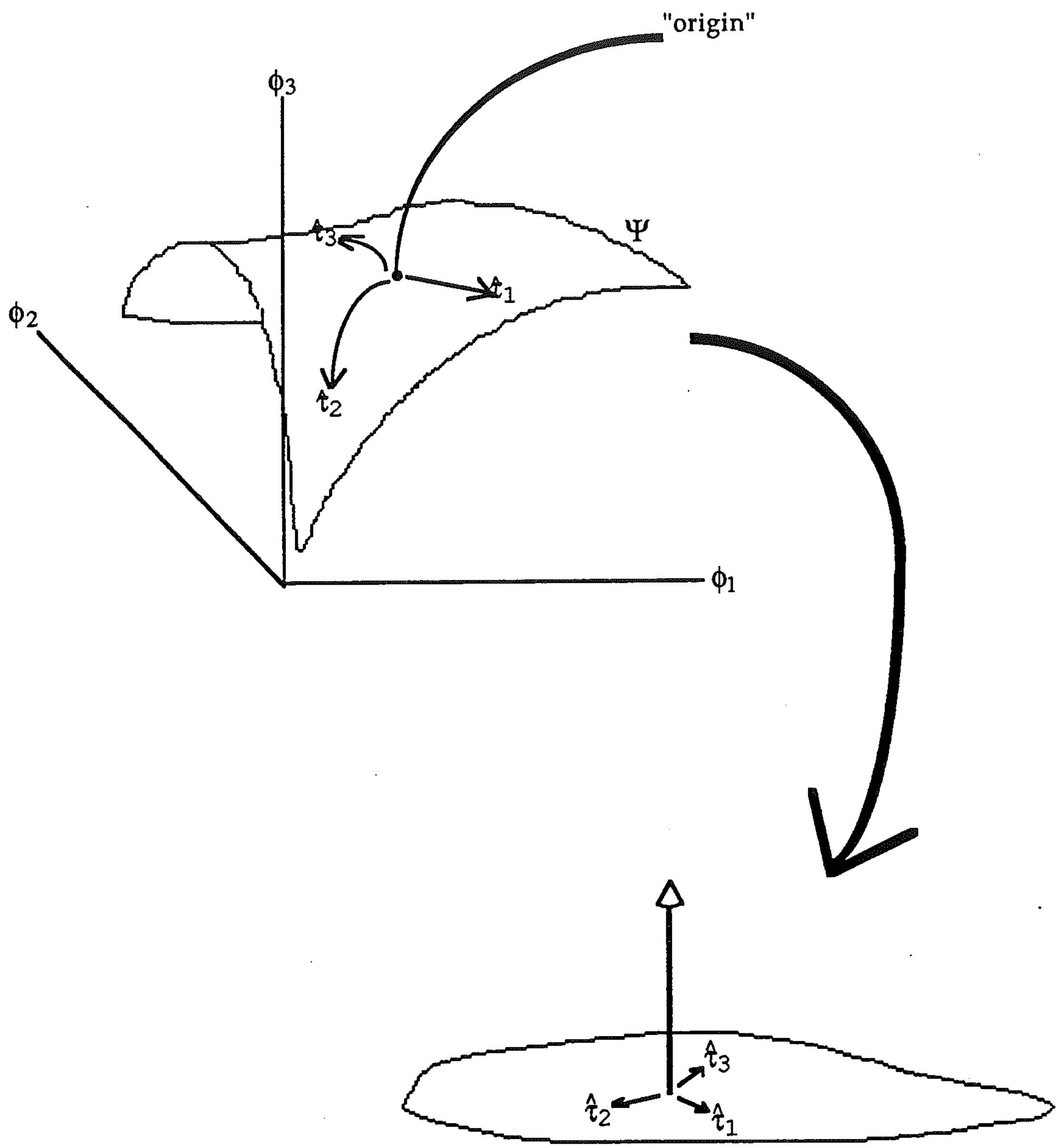


Fig. 2-8 A manifold Ψ of dimension k (here 2) has some set of mutually transverse intrinsic parameters $\{\hat{t}_1...\}$ lying within it. Any k of them serve to reparameterize the space; i.e., they "span" the manifold.

Say there is a manifold Ψ of dimension k. Any set $\{\hat{t}_1...\hat{t}_k\}$ that are all transverse to one another spans the new space: that is, any point on the manifold can be expressed as a weighted sum only of the $\hat{t}_i$. Hence any set $\{\hat{t}_1...\hat{t}_k\}$ of transverse translations coupled with an origin 0 defines a "model" of points in Ψ.

To help the intuition here, imagine that there is an ant on the surface, poised at the point 0. Say that it can only move in directions $\{\hat{t}_1 ... \hat{t}_k\}$—that is, in some set of straight directions, and no others. Then, if an only if $\{\hat{t}_1 ... \hat{t}_k\}$ are mutually transverse, it can move to any point on Ψ just by moving in those directions. With fewer directions to work with, or some of them are lined up with one another, there will be points on Ψ that the ant cannot reach.

In the example in the figure, then, there are three different models of the same new space: $<\hat{t}_1, \hat{t}_2>$, $<\hat{t}_1, \hat{t}_3>$, and $<\hat{t}_2, \hat{t}_3>$. Each of these suffices to represent the position of some point in the new space, i.e., some object in the category Ψ (Fig. 2-9).

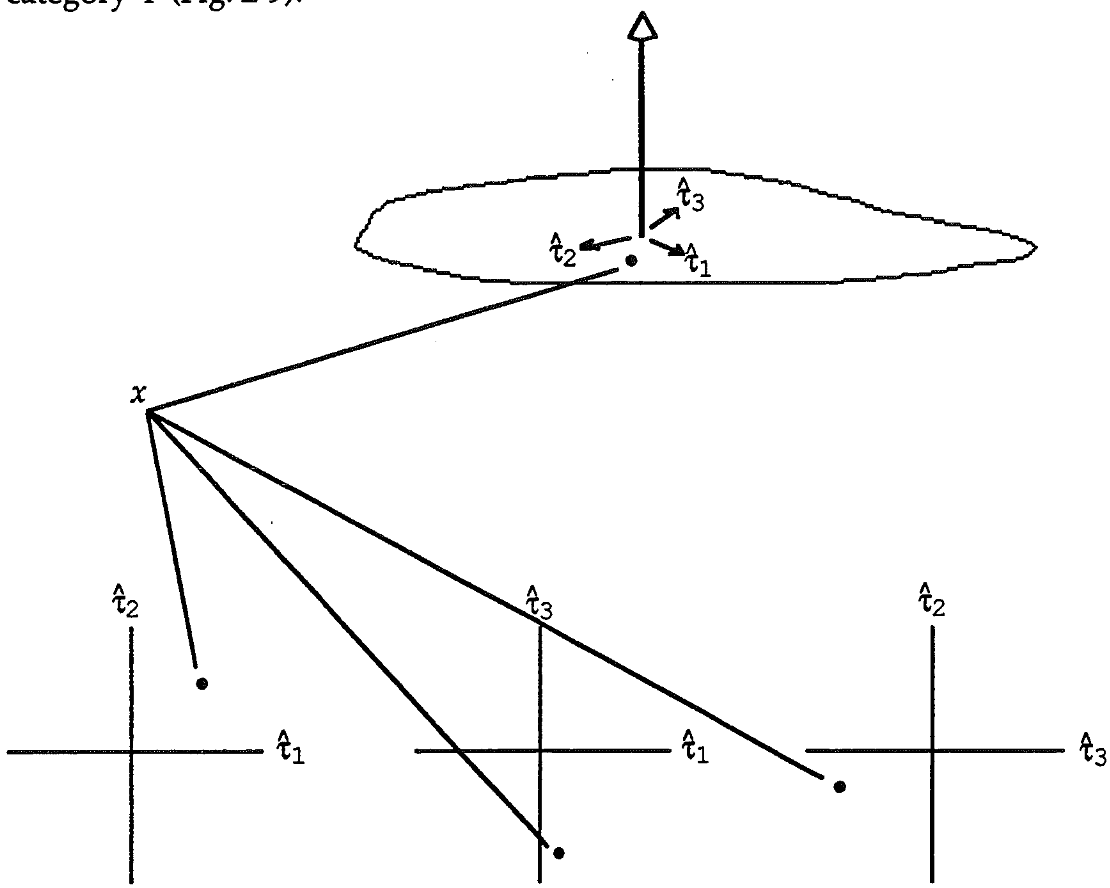


Fig. 2-9. Each of the three pairs of unit translations drawn on the manifold Ψ in the previous figure can act to parameterize the new space—the category—in a different way. Hence some point x on the manifold—an object in the category—can be expressed in a different way as a point in each of the three new spaces.

The existence of a direction simultaneously orthogonal to all of the model class directions (drawn as an arrow upwards in the Fig. 2-9) is a consequence of their transversality. This direction in effect corresponds to another "transformation" that *cannot* be applied to objects in the category without taking them out of the category—that is, a direction that takes them off of the category manifold. In the dot-on-line example from Ch. 1, this corresponds to movement off the line segment: any application of that operation to an object in the category produces a new object *not* in the category.

The selection of a generative model forces a choice of reparameterization. So far nothing has been said about the operations used to reparameterize the space in the above example, except that any transverse set of k of them suffices to make a new "model" of points in the category. It can now be observed simply that the choice of a "generative model" for objects in the category—a sequence of operations that we interpret as having brought each object into existence—forces a choice of reparameterization parameters.

Say there is a generative sequence of operations $\rho = \{g_i \cdot \ldots \cdot g_k\}$ acting on some primitive object x_0, from which each object in the category is conceptualized as being descended. Simply, the choice of the object x_0 picks out an origin point $0_\rho \in \Psi$ on the manifold; and each generative operation g_i picks out an independent direction $\hat{t}_{i_\rho}$ measured along surface of the manifold, a parameter along which objects on the manifold can vary. The generative model ρ, then, selects a reparameterization model from all the infinity that are possible. That is, the choice picks out a *basis* for the new configuration space.

Now, having selected a generative model ρ, each object x in the category can be represented in a natural way as a weighted sum of these defined unit translations in the new space:

$$x = 0_\rho + \sum_i \alpha_i \hat{t}_{i_\rho},$$

where the various α_i are simply the magnitudes to which the i-th operation $\hat{t}_i$ is carried out. For clarity the subscript ρ will be omitted whenever the intended choice of generative models is clear. The choice of origin and basis vectors for the newly parameterized space, $\rho = (0_\rho, \langle\hat{t}_1 \ldots \hat{t}_k\rangle)$, are called a *model* for the category, or *category model*. Putting it another way, the choice of ρ engenders a "language" in which the structure of objects in the category may be expressed.

Notice that choice of ρ to reparameterize a manifold actually carries with it a number of other category models, corresponding to the various subsets of $\langle \hat{t}_i \ldots \hat{t}_k \rangle$. Each of these, in other words, is a *subspace* of the category embodied in the manifold Ψ. Equivalently, each of these corresponds to one of the $\alpha_i = 0$ for some i, so that the i-th operation completely drops out. Each of these models, in turn, constitutes a non-generic special case of the general category. Without any constraint, there are 2^k such models, each corresponding to the sum of a different subset of the k generating operations. Thus for clarity the general choice of generative operations $\langle \hat{t}_1 \ldots \hat{t}_k \rangle$, which defines a miniature universe of objects, will be referred to as the *model class*:; and sums taken over subsets of these operations (e.g. $\hat{t}_1, \hat{t}_2, \hat{t}_1 + \hat{t}_2, 0$) will be referred to as individual *category models*. When no confusion is engendered, the model class (e.g., $\langle \hat{t}_1 \ldots \hat{t}_2 \rangle$) and its codimension 0 model ($\hat{t}_1 + \hat{t}_2$), the most generic model in the class, will be referred to interchangeably.[8] Note that a category model is instantiated into a particular object simply by attaching numeric weights to each of the terms in the model; for example, the object $1.6\hat{t}_1 + 2.7\hat{t}_2$ in some model class $\langle \hat{t}_1 \ldots \hat{t}_2 \rangle$ corresponds to subjecting the primitive origin object 0 to the $\hat{t}_1$ transformation to a magnitude of 1.6, and to the $\hat{t}_2$ operation to a magnitude of 2.7.

The critical roles played by genericity and non-genericity (i.e., the application of some transformation to a magnitude of 0) within category models can now be taken up. Briefly, objects in a given category model will be required to be mathematically *generic* in that model, meaning that all of their α_i's are non-zero. This allows us to interpret non-generic objects (objects in which some $\alpha_j = 0$) as in fact generic objects in a lower-dimensional category, (namely a category in which $\hat{t}_j$ does not appear at all in the generative model). This enables us to conceptualize the model class as a collection of discretely different generative models, each of which produces

[8]Notice though that later when we consider signed (+/-) parameters, these two are no longer interchangeable, as there may be multiple distinguishable codim-0 models within a single model class (i.e. the various "quadrants" at each dimensional level).

only a highly uniform class of objects, but between which structural differences are relatively sharp.

2.2.3. Genericity

A species of the ubiquitous *genericity* constraint will now be adapted for application to category models. This constraint simply dictates that each object is mathematically typical of the model class to which it belongs.

(Genericity)

Each object should be identified with a generative model in which it is generic..

This simply means that for each object $\sum_i \alpha_i \hat{t}_i$ all $\alpha_i \neq 0$, i.e. each weight is non-zero. Interpreting the parameters in terms of a generative model, what this means simply is that each object in a model class is a *typical* product of that class, in the sense that its production involved every individual generative transformation to some non-zero extent. In other words, by simply assuming that each object in a class is the product of the same causal history, their representations in parameter space can be expected to have non-zero weights in every direction.

The genericity constraint elsewhere. The deceptively simply but powerful genericity constraint makes an appearance in one form or another in a wide variety of perceptual theories, always providing the critical inferential leverage to disregard atypical cases. In all forms, the constraint dictates that we can safely assume that the world will take a typical configuration, since it does so in the vast majority of cases. Putting it another way, we should set up perceptual inferences so that they do not depend on atypical or non-generic configurations in order to operate. That way, false targets (in which the perceptual scheme is "fooled") occur in a vanishing, atypical subset of cases. This idea appeared originally in the context of inferring viewpoint-invariant 3-D structure from single 2-D views (see for example Koenderink & van Doorn, 1976; Lowe, 1985; and Koenderink, 1990), and in the interpretation of structure from motion (Ullman, 1979), in which the points in the successive views must be assumed to be in general (i.e., generic) position.

When attempting to infer the 3-D shape of an object from 2-D cues, these theorists have argued that we assume that the object appears in an orientation that is generic with respect to the line of sight. That way, the inference scheme only fails in non-generic views (such as when the line of sight falls in a plane parallel to one side of an object), which are a measure-zero subset of the sphere of orientations. The genericity constraint commonly appears as a foundational assumption in more generalized theories of perceptual inference (Bennett & Hoffman 1989, Jepson & Richards 1991; Clark & Yuille 1990). In a different but analogous vein Sacks (1990) constructs a scheme for automatic analysis of the qualitative behavior of dynamical systems, by critically assuming mathematically generic behavior in the governing differential equations. Sacks & Doyle (1991) argue that human experts, examining such systems and attempting to predict their behavior intuitively, make this assumption as well, restricting their attention to generic cases and ignoring atypical cases that rarely occur.

In the context of category models, the genericity constraint simply dictates the assumption that all objects that are the product of a particular generative model are generic in that model. Equivalently, each object can be expressed algebraically as a member of a model class in which it is generic—i.e., in which all of its parameters are non-zero. This simply enforces the algebraic cancellation of identity (zero-valued) transformations. Notice, however, that objects that are non-generic in one model can very well be generic in a higher-codimensional model: namely, a model containing just those parameters on which the given object *was* generic. For example,

$$\alpha_1\hat{t}_1 + 0\hat{t}_1 = \alpha_1\hat{t}_1,$$

$$0\hat{t}_1 = 0,$$

and so forth. That is, the genericity constraint has the effect of partitioning the configuration space into equivalence classes: sets of objects all of which are generic in the same generative model—i.e., the category models in the model class (see Fig. 2-10). These category models are thus treated as, in effect, equivalence classes of causal history. The model class $<\hat{t}_1,\hat{t}_2>$ thus has exactly four distinct category models in it: 0, $\hat{t}_1$, $\hat{t}_2$, and $\hat{t}_1 + \hat{t}_2$. Sec. 2.3 below will consider some concrete examples of this, in which a particular object-producing generative process is adduced.

$\hat{t}_2$ \ $\hat{t}_1$	$= 0$	$\neq 0$
$= 0$	1	2
$\neq 0$	3	4

1 generic in model 0

2 generic in model $\hat{t}_1$

3 generic in model $\hat{t}_2$

4 generic in model $\hat{t}_1 + \hat{t}_2$

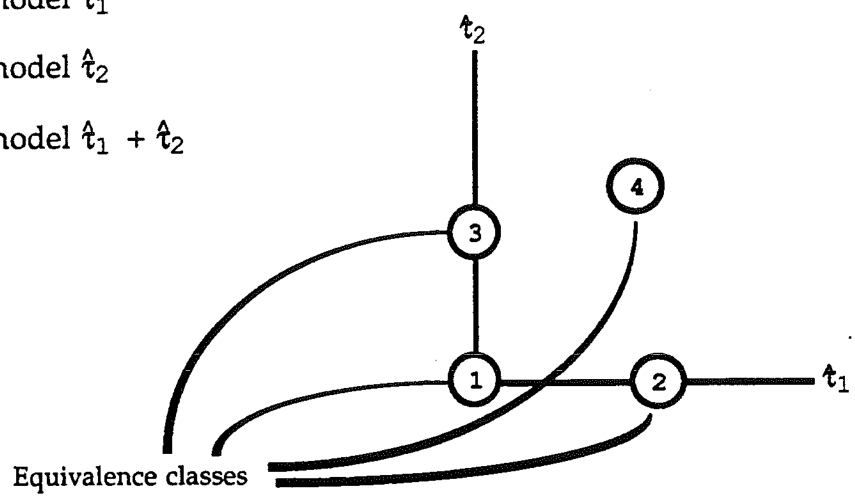


Fig. 2-10. The genericity constraint partitions a configuration space into a number of equivalence classes, each of which contains objects that are uniformly generic in the same model (subspace). On the upper left, a table of the four cases: $\hat{t}_1$ generic or non-generic $\times$ $\hat{t}_2$ generic or non-generic.

If a concept of "sign" (positive or negative) is added to each parameter (see discussion below), the result is quadrants (in general, n-ants) each constituting a different way of being generic in the given space (Fig. 2-10). In the Fig. 2-11, a schematic for a model class with two signed parameters $\hat{t}_1$ and $\hat{t}_2$, there are 4 different ways of being generic in the model $\hat{t}_1 + \hat{t}_2$, two different ways of being generic in the model $\hat{t}_1$ (i.e. the two half-space 2-ants of the axis $\hat{t}_1$), similarly two different ways of being generic in the model $\hat{t}_2$, and finally just the one way of being "generic" in the completely collapsed model 0, namely to have structure fixed at the origin of the space.

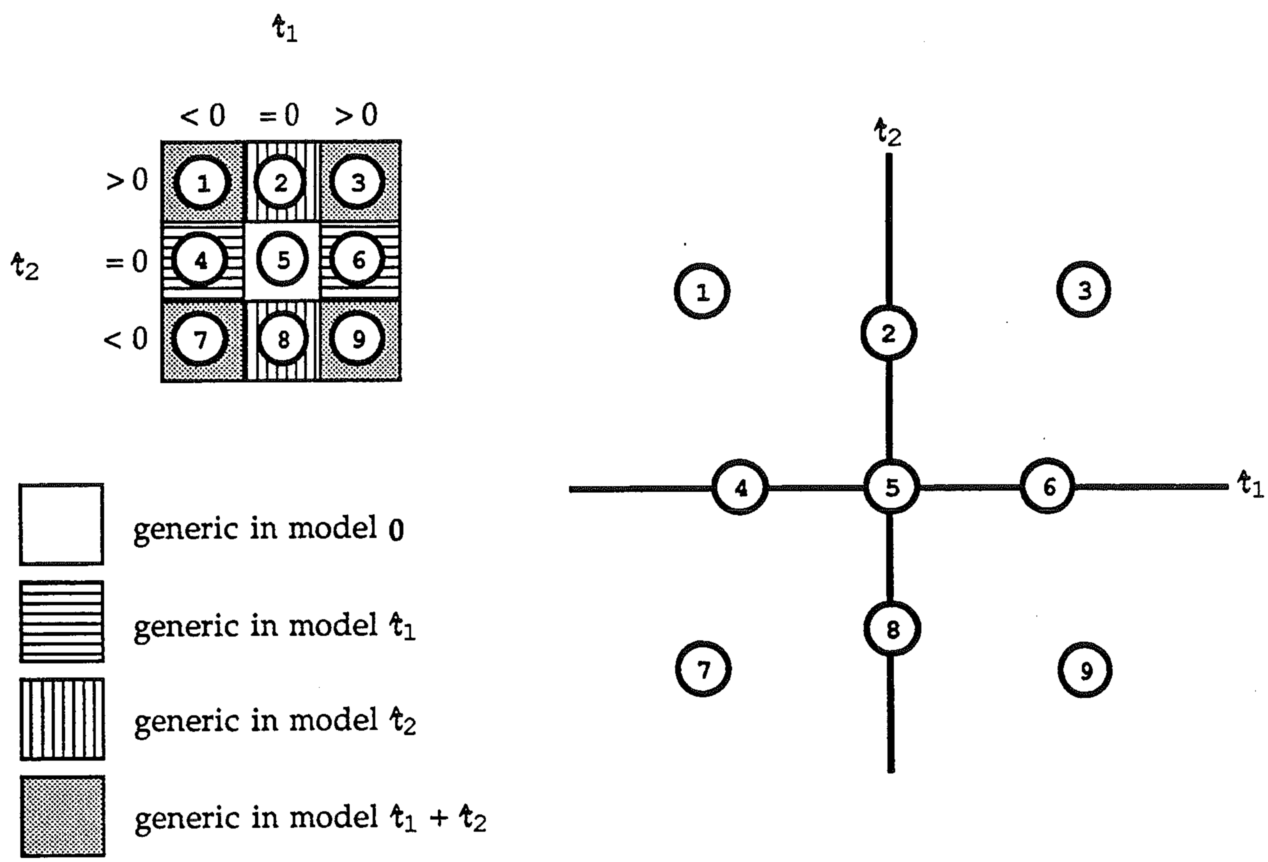


Fig. 2-11. Schematic of a space defined by two *signed* parameters. There now four different ways of being generic in the codim-0 model $\hat{t}_1 + \hat{t}_2$, two different ways of being generic in the model $\hat{t}_1$ and the model $\hat{t}_2$, though there is still only way of being "generic" in the completely collapsed model 0.

The parametric collapse of one category model to a lower-dimensional model is called a *modal* non-genericity. Here the term modal simply emphasizes that the non-genericity is one that is explicitly stipulated in the vector basis—that it is a is a non-genericity that is "expected" to happen, and to which we attach special meaning, namely that it connotes a distinct generative model. Ch. 5 will explicitly identify these modal non-genericities—the various categorical ways in which we expect structure to manifest itself in the constrained world under observation—with *statistical* modes in a probability distribution defined over the configuration space.

2.2.4. Choice of bases for the parameterization

This section takes up the question of which points along parameters should be treated as "special" values, i.e. the values at which parametric collapse to a distinct lower-dimensional subcategory occurs. In order that the parametric collapses occurs at the "special values," the parameters must always take the normalized form of translation through parameter space away from the special value, so that the zero of the parameter is exactly the modal non-generic point. Within a given context, each special value on some abstract dimension corresponds to a particular configuration of some concrete structural parameter. The set of configurations picked out as special by the observer will sometimes be referred to as the observer's set of "concepts": these are the configurations the observer presumptively regards as liable to require causal explanation.

As a possible criterion for special values, consider functions that are especially likely to be an extremum of some function that plays a role in the regularities in the given context—i.e., that they fall at extrema of functions that are likely to come up in scientific theories governing the domain in question. Plainly, *any* point along a parameter is a zero for some function. However, this criterion has the effect of promoting several characteristic types of points, namely those that are the locations of extrema in canonic forms of functions that appear frequently in mathematical descriptions of systems: sine, cosine, linear functions, monotonically increasing or decreasing funtions, and so forth. In other words, these values along parameters are special in the world because are featured in the regularities governing the world; their presence in our concept set is a recognition of their importance in the world, and hence mark an attempt to orient our categorizations around these regularities. Some of these characteristic types will now be enumerated.[9]

Consider first a parameter x with the topology of $[0,\infty)$, such as translation in one direction away from a point. Any function $f(x)$ that is monotonically increasing, plainly, will have a global minimum at $x = 0$.

[9]These can be compared to Luce's (1990) short list of discrete types of singularities (zeroes) along psychological parameters.

Hence zero is a good candidate for being a "special", i.e. non-generic value of this parameter. While there is no a priori reason to expect any particular function to be monotonic, real-world functions have a tendency to be simple in overall form, and of course monotonic functions come up everywhere. The point is not even that there is a high probability of a function being monotonic; in fact it is problematic even to assign a prior probability to such an event. Rather, note simply that monotonic functions constitute a coherent model is entertained, discretely, along with others that are also entertained.

Special value type **(i)**: $0 \in (0,\infty)$.

Consider next a parameter x with the topology [0,1], such as translation along a finite segment. Any *linear* map $f(x)$ from this parameter must have an extremum at either $x = 0$ or $x = 1$. Again, this is simply a nice, coherent model that may be considered,[10] not in any sense an outcome to which a high probability can be meaningfully attributed. A higher order function may of course have an extremum at a point somewhere in the middle $0 < m < 1$, so that the normalized parameter is now $(x\text{-}m)$. This creates the possibility of the parameter having a sign: positive in $x > m$, negative in $x < m$. Points that fall all on one side of such a division may be thought of as being all qualitatively equivalent to one another, but being different from points on the other side, in the sense that a point cannot pass from one sign to the other without passing through a different category, namely the origin.

It is possible to be more specific about where this point m may fall. The parameter $x \in [0,1]$ has a *symmetry*: the endpoints are not distinguishable. This symmetry is recognized if x and $1\text{-}x$ are treated equivalently. Any quadratic function that is symmetric in these two quantities, i.e. that is monotonic in $x(1\text{-}x)$, has an extremum at $m = 1/2$. Hence we have a third possible special value, creating the concept "bisection." This may of course be repeated recursively. Hence we can add to the list:

[10]More specifically, one may argue that due to the structure of the Taylor series, that when considering analytic functions, it is meaningless to consider higher-order functions without also considering lower-order functions—and hence, meaningless to consider *any* functions without consider linear ones. In effect, the mathematics dictates the primary importance of linear and other low-order functions.

Special value type **(ii):** $0 \in [0,1]$,
Special value type **(iii):** $1 \in [0,1]$,
Special value type **(iv):** $1/2 \in [0,1]$,

Some real-world parameters x are periodic in nature, i.e. will have the topology of the circle, $x \in \mathbf{C}$. Of course, such functions are conveniently expressed trigonometrically. Hence if $f(x)$ is monotonic in $\sin x$ $(x \geq 0)$ it will have a minimum at $x = 0$ and $x = \pi$. Similarly, if $f(x)$ is monotonic in $\cos x$, the minimum will fall at $x = \pi/2$. Thus, right angles are special, in essence, because the cosine function is salient in the physical theories that describe the regularities that give rise to world categories.

Special value type **(v):** $n\pi/2 \in \mathbf{C}, n = 1,2...$

The observer's conceptual scheme determines what points to treat as special; again, any alternative conceptualization of a zero point, that is, any alternative conceptualization of a special configuration, could be adopted. In any case the parameter would be normalized to distance away from the special value.

Parametric collapse at modal non-genericities. Critically, again, if some special value s along some parameter x is stipulated to be a modal non-genericity, then a subcategory springs into being at $x = s$. Using the normalized parameter x-s, there is a discrete modal category change, e.g.

$$\alpha_1\hat{t}_1 + (x\text{-}s)\hat{t}_2 \rightarrow \alpha_1\hat{t}_1 \quad \text{at } x = s,$$

that is, a parametric collapse from one category to a lower-dimensional subspace of it. The genericity constraint dictates a shift from one model to another one, down the ladder of codimension.

2.2.5. The "fundamental hypothesis"

The theory presented above can now be organized into a single, underlying principle of inference for categories. Human categorizers, it is assumed, seek in effect to organize objects into coherent groups, such that each group represents objects that have something fundamentally in common with respect to their causal origins, their uses, their behavior, and most generally their unknown properties (see Bobick, 1987). However, one cannot make infinite investigations into the deep, hidden relationships among observed objects, nor vacillate endlessly between alternative groupings of observed objects that are cannot be given an absolutely a priori preference ordering. Rather, one must take relatively obvious guesses about the apparent structure of objects, guessing that some properties are liable to be causally important, and classify first according to these. The notion of genericity and the selection of some set of "special values" of familiar parameters simply provide a scheme for going about accomplishing this.

Our "fundamental hypothesis," then, comparable to that of Marr (1970) and to the "Natural Modes Principle" of Bobick (1987) and Richards & Bobick (1988), is this: the regularities in the world—the patterns of covariation that tend to structure natural parameter spaces into lower-dimensional manifolds—can be profitably organized by postulating generative models that captures the extant variation in a more "meaningful" way. In this manner we as observers seek to pick up clues, albeit coarse ones, to the causal forces at work behind the objects. Treating generic examples of each given generative model interchangeably is a kind of cagey "cheat" on the part of the observer: we intentionally ignore some structural differences among observed objects—ones that we guess will *not* tell us anything reliable about other properties we might be interested in—in favor of a tiny handful of critical structural differences that, putatively, will.

2.3. Mode lattices

The genericity constraint forces a partition of a model class into various category models (subspaces), i.e. the various distinct ways of being generic in a given basis. This section develops an organizational scheme, the "mode lattice," to enumerate and interrelate the various models (the various modal non-genericities) in this collection. This section lays the scheme out conceptually and give two simple examples; in Ch. 4 the construction of lattices from a model class will be implemented computationally and examples discussed in more detail.

2.3.1. Basic machinery

Each model (subspace) is represented by a node in a directed graph, in which it is simply connected to all the nodes that are subspaces of it of dimension one lower. Thus we have $\begin{array}{c} a \\ | \\ b \end{array}$ whenever b is a subspace of a and dim a = dim b + 1. This relation forms a lattice (a kind of partial-order than can be pictorially depicted in a convenient way; see for example Gericke, 1963). Each line segment in the depiction (meaning, every lower neighbor relation) thus represents a single parametric collapse.

Fig. 2-8 gives an example of the schematic representation, the canonic form for a two-parameter space.

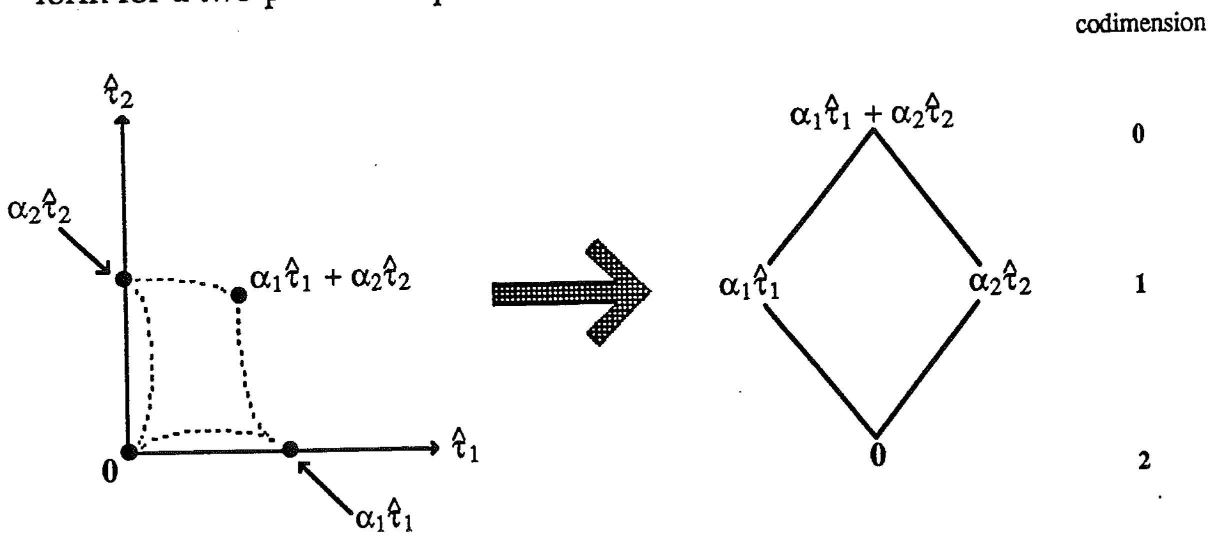


Fig. 2-8. The category models in a model class, seen (left) in the parameter space, and (right) in the form of a lattice.

In the left of the figure, the configuration space itself is depicted, with generic points in each of the 4 subspaces (including the improper subspace, i.e. the full space) marked with points and corresponding symbolic expressions. On the right the same four subspaces are arranged into a lattice, with the codimension of each row marked on the right. The codim-0 row represents the completely generic class—the class of objects that obey no recognized regularity with respect to the general definition for the model class (e.g., scalene triangles), though note that entire model class can represent a constrained subcategory of some even larger set (e.g., triplets of line segments). Each codim-1 model, similarly, contains objects that have one degree-of-freedom's worth of non-genericity (e.g. right triangles, isosceles triangles, etc.). (See Feldman, 1991b for a fuller discussion of triangle lattices, including alternative parameterizations.)

2.3.2. Example (i): V's

The 2-parameter lattice can now be fleshed out by adopting a particular generative model, i.e. a mapping from points in a parameter space to objects. Each object in our example model class consists of a pair of segments joined at one endpoint, the V's of the parameterization example. The "natural" parameter expressing the difference between the orientations of the legs, as was found in Sec. 2.2.2 was "angle," or arclength along the equal-leg-length constraint manifold. Similarly, if the angle is held constant, the natural parameter expressing the difference in shape between two V's is the difference between the length of the legs. These two parameters may, in turn, be thought of as standing in for generative processes (opening and closing the V, like a pair of scissors, and stretching one leg with respect to the other, respectively) by whose action the shape of the V is conceived to have been determined. Now drawing a sample object at each node, the resulting lattice appears in Fig. 2-9.

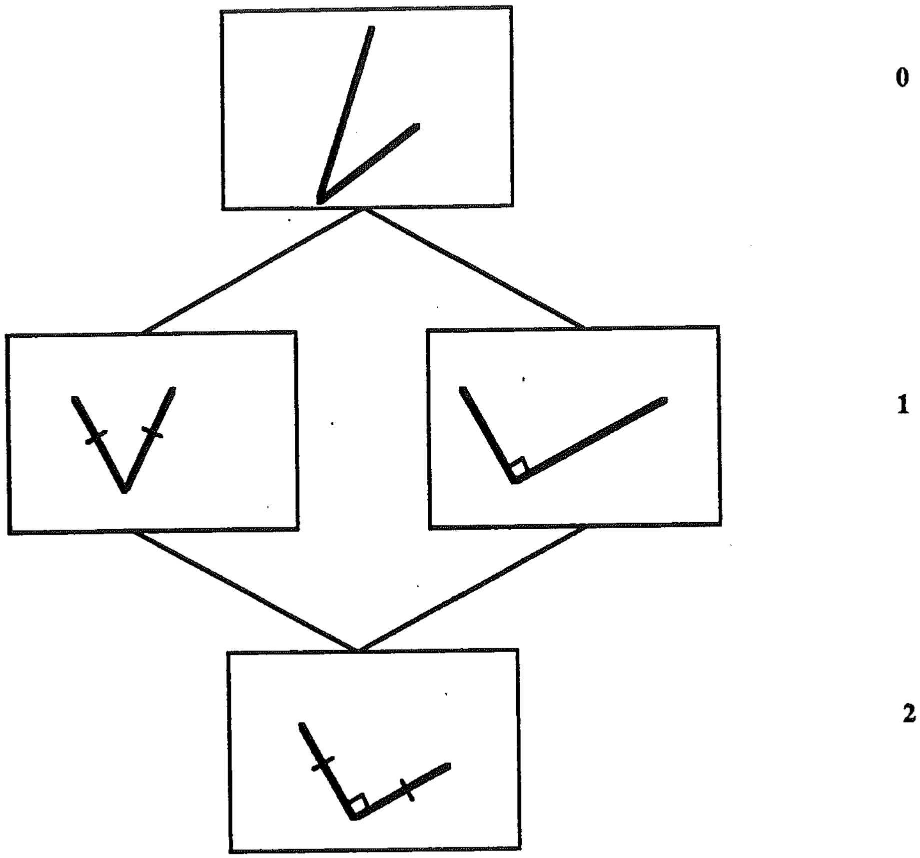


Fig. 2-9. Mode lattice for the V's model class.

At the top of the lattice is the codimension 0 category, the one in which a completely unconstrained V is generic. As discussed above, non-generic V's may be seen as generic in a lower-dimensional category; since it is a 2-parameter category, and each parameter has one zero, there are two such subspaces: right-angle V's and equal-leg-length V's. That is, because of the basis chosen for the parameter space, there are exactly two independent ways for a V to be *modally* non-generic: namely, the two codim-1 categories. (By contrast, some hypothetical atypical V such as one with one leg exactly twice the length of the other may be non-generic in some sense, but it is not modally non-generic in the chosen parameterization.) At the bottom of the lattice is the codimension 2 category, in which there are no degrees of shape freedom left to vary.

2.3.3. Example (ii): line-dot

Now consider an example that does not take the canonic form of the subspace: a line segment and a dot in the plane (Fig. 2-10).

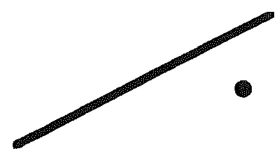


Fig. 2-10. A line and a dot. What are the categories of this model class?

First a parameterization needs to be chosen for the relationship between the dot and the line. Consider that the observer has a limited inventory or repertoire of well-defined operations to work with. Since the relation between a segment and a dot can be expressed as the *position* of the dot in the plane with respect to the segment — the position of the dot being what is to be parameterized— the relevant operations are all the translations (i.e., all the operations that affect position) that can be defined unambiguously with reference to the line segment. First of all, there is the two-parameter free translation from the one (arbitrarily selected) endpoint of the segment $\hat{t}_{r,\theta}$ (see Fig. 2-11). The angle must be measured from the segment, since in this miniature universe there exists no other coordinate frame. (Of course if we superimpose another frame, such as a gravitational one, that will lead to a change in the categories, as desired.) Also available is the one-parameter translation *along* the segment, starting from each endpoint: $\hat{t}_{tangent}$

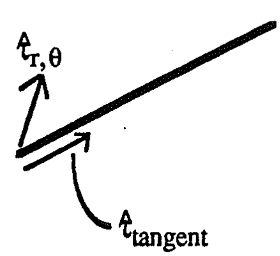


Fig. 2-11. Operations available to construct a generative model for the relation between line segment and dot.

The lattice for the line-dot model class is depicted in Fig. 2-12.

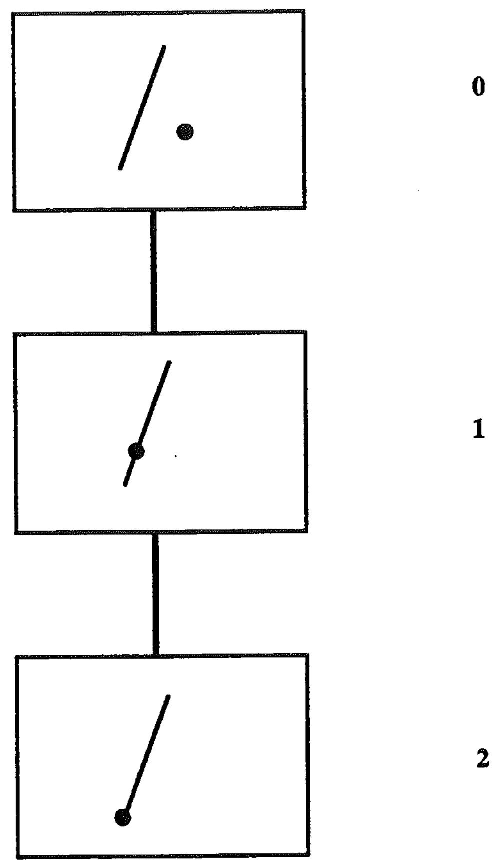


Fig. 2-12. The lattice for the line-dot model class.

Note here that the codim-1 category, "dot aligned with segment" really contains both the "dot on line" case and the "dot collinear with line but not touching" case, in which the line segment "points to" the dot but does not touch it; these two cases are equivalent with respect to the concepts used to construct this particular lattice, which do not include any representation of contact per se. Some concept making contact explicit presumably is a part of the full human concept contingent, in that it seems unnatural to generalize from a single object (dot on line) to a pair of objects (dot collinear with line but not touching). Thus the lattice shown here can be thought of as corresponding to a somewhat impoverished concept set. A more detailed exploration of this lattice will be taken up later when a computer program is presented that produces these lattices automatically.

2.4. Category inference with mode lattices: a minimum principle

The mode lattice collects together all the causally coherent categorical hypotheses that an observer attributes to the given model class. Confronted with a particular object, the observer must choose one of the categories on the lattice as the one the object is most like to genuinely belong to. The question here is: which one?

Clearly there are some categories on the lattice that the do not fit the observation at all, and these may be rejected at once. However, the problem is that there are almost always multiple categories that *are* consistent with the observed object. Very simply, this is because if an object fits into one category on the lattice, it also fits into every upper neighbor (superset) of that category, recursively on up to the codimension-0 (generic) case, which by definition contains all others.

The genericity constraint solves the problem by providing the inferential leverage to distinguish the true category from others. While the observed object is *consistent* with both its true generating category and all of its upper ancestors, it is only *generic* in the true category — it is modally non-generic in all the ancestors, by stipulation. Hence the natural selection rule is simply to pick the highest codimension (lowest dimension) category consistent with the observation.

Such a rule is of course quite familiar. The Gestalt law of "pragnanz" ("good form"; see Koffka, 1935; Köhler, 1947) dictates that among all descriptions consistent with a stimulus, the simplest be preferred. A wide variety of other kinds of perceptual simplicity metrics have been proposed; see for example Hochberg & McAlister (1953); Simon (1972); Barrow & Tenenbaum (1981); Buffart, Leeuwenberg, & Restle (1981); and Kanade (1981). A version of this idea called the "Minimum Description Length principle" due to Rissanen (1978) has been applied with wide success to a variety of domains of data interpretation (see Quinlan & Rivest, 1989), including perceptual tasks (Darrell, Sclaroff, & Pentland, 1990).

The minimum rule proposed here is clearly intimately connected to these prior ideas. Note however that the rule proposed here there are several advantages to the form in which it appears here, which are taken up in the next two sections.

2.4.1. No presumption of simplicity

One advantage is in the motivation behind the dimension minimization. The "grounding" of a simplicity criterion—justification of why it is reasonable to assume the world will tend to have a simple structure—has troubled a number of authors dating back to Hume (see Sober, 1975, Hatfield & Epstein, 1985 and Leeuwenberg & Boselie, 1988 for contemporary discussions). Some authors have assumed that simple processes are more *likely* in the world (a view Sober calls "probabilism") while others (e.g. Quine, 1963; Sober, 1975; Harman, Ranney, Salem, Doring, Epstein, & Jaworska, A., 1988) take pains to avoid such an assumption, feeling that it tends to attribute to simplicity an objective status that they find difficult to justify. In some cases the latter authors characterize the preference for a simple description subjective, or argue that it derives from concerns entirely exterior to achieving a veridical description (such as constructing a convenient, compact, or transparent description). At least one author (Perkins, 1976) argues that simplicity criteria tend to run counter to veridicality, arguing that we impose regularities in the world that are not actually present.

The current theory tends to avoid these complications, though, because no presumption of simplicity, on the part of either world processes or stimulus patterns, has had to be made at all. Rather, the criterion falls out naturally from the genericity constraint, a notion that is wholly less problematic to justify. Putatively, the argument here shows how the need to associate objects with their generating classes veridically and reliably has the effect of dictating a kind of minimum principle (see also Feldman, 1991b for a discussion of this point).

2.4.2. Constraint on dimension minimization is explicit

A second advantage to the lattice rule is that while description dimension is minimized, there is no attempt at *global* minimization of complexity, an extremely problematic concept. When the goal is simply to minimize descriptive complexity, regardless of description language, the global minimum may take the observer well out of the familiar territory of meaningful description languages. (Of course in the limit, if all languages are accepted, there is always a language in which the current object can be described compactly, but emptily, by a single symbol.)

Here, by contrast, the idea is to delimit the "meaningful" languages of description, namely to generative model descriptions that have bases constructed from a limited inventory of concepts, as described above. The minimal description sought is strictly within these very tight boundaries — namely, the lowest node within the lattice. Hence there is really a dual goal: minimize complexity while restricting hypotheses to coherent categories as enumerated on the lattice.

The lattice thus forms a concrete framework of constraint within which descriptive complexity (per genericity) can be minimized, but outside of which category hypotheses are constrained not to wander. The result is a categorical hypothesis which is both *coherent* (in that it marks a potentially causally meaningful category) and *reliable* (in that false targets are non-generic).

- Chapter 3 -

Patterns of Covariation and Constraint Among Discrete Properties

3.0 Chapter Preview

This chapter explores the notion of "constraint" and "regularity" in a world represented only in terms of discrete properties. Various canonic forms of constraint are distinguished, each interpretable as a different form of causal influence among properties in the world. We find that each form gives rise to lattices of distinct category models at various levels of inclusiveness, exactly as in Ch. 2.

Sec. 3.1 argues that discrete properties make a natural context in which to consider canonic forms of constraint, regularity and property covariation in the most general way possible.

Sec. 3.2 attempts to enumerate canonic forms of constraint among two discrete properties. A system of definitions in Prolog is introduced that produces lattices under various defined forms of constraint. These simple two-property lattices can themselves be arranged into a "superordinate" lattice, which makes explicit the fact that some forms of constraint—some "causal theories"—are more constrained than others, just as (in an ordinary lattice) some category models are more constrained and others more inclusive.

Sec. 3.4 shows how the superordinate lattice can be used to make explicit the varying degrees of structure inherent in different inductive situations, such as Bongard problems. Various ways of analyzing Bongard problems are developed, each making explicit the structure that problem in effect imposes on the miniature property world in which it operates. The highlight is the derivation of the superordinate codimension of the canonic Bongard problem.

Sec. 3.5 considers the question of how complex a world can be categorized correctly. We propose a regularization rule whereby observers can limit the complexity of the discrete lattices they infer in the world—in essence, limiting the complexity of the causal relationships they are willing to hypothesize, and regularizing more complex worlds to a simpler but more intelligible form. This rule, a kind of discrete Natural Modes notion, can be construed as a discrete version of the constraint manifold conception of natural categories from the Ch.2.

3.1. The idea of constraint among discrete properties

In order for an observer to validly find structure in the world, it is necessary that the world actually *contain* some structure—that it not be a completely haphazard place in which no regularities consistently apply. Such structure might take a number of forms: some properties tend to occur only in the presence of certain other properties, some only in the absence of other properties, some properties never apply, some properties always apply, and so forth. This chapter will consider how this list—of regular ways that properties might relate to one another—might terminate. In Ch. 2, we considered a smooth relationship between two parameters of structure in the world as a model of "structure." In this chapter we generalize this notion somewhat, considering arbitrary "regular" relationships between logically orthogonal properties. The natural way to do this is to model properties simply as discrete, yes/no questions about properties, and then try to describe how causal influence among sets of such properties gives rise to discrete category models defined over those properties.

Thus in this chapter we will simply exhibit and attempt to enumerate the various ways in which properties can regularly co-occur with one another—that is, to create a more general model of what distinct forms "structure in the world" can take. We will treat "properties" as binary-valued functions defined over entities; i.e., a property is simply a fact that is either true or not true of each thing we consider. Exactly as in Ch. 2, the central formal structure enumerating and interrelating category models is the lattice. Discrete lattices, analogous to those in Ch. 2, come about when we connect each category model—i.e., each collection of defining properties—with more restricted category models that contain additional properties (i.e. require additional properties of their members). That is, the resulting lattice depicts the *subset* relation among more and less inclusive category models, expressing the organization among the category models that are possible in the universe defined (spanned) by the given set of properties.

I have created a Prolog environment in which one can build lattices can be constructed and manipulated, and in which it is possible to explore the varieties of structure that such lattices exhibit when various modes of constraint and regularity are imposed upon them. In Prolog, the computer attempts to satisfy a given expression (i.e., find values for its unbound variables) consistent with all the facts and definitions it has been given. An environment is thus defined by the sum total of the facts and definitions the user has inserted. This makes it well-suited for the category construction problem, where Prolog will take a given model class and given definitions of category models, and propose some consistent category interpretations. Since we do not assume familiarity with Prolog on the part of the reader, the next section will begin in a tutorial style, building up the lattice constructs and definitions gradually with examples.

In a sense, the goal of this line of inquiry is to define as carefully as possible what is meant by an appropriate level of generality for a model of observations. Even though, as we will see, one can think of constraint as taking only a small number of canonic forms, the resulting lattices expressing the relations among models of various levels of generality can get quite complex quickly. The fact that humans can make confident generalizations with little data— inductions that, amazingly enough, future facts often validate as having been pitched at about the correct level of generality— depends on the fact that we make certain simplifying assumptions about world structure. At the end of this chapter, with concrete definitions of "structure" and "generality" in place, it will be possible to make a concrete proposal about the form that this underlying regularizing assumption takes.

3.2. Automatic construction of orders and lattices

I have written a network of interrelated Prolog definitions, that enables Prolog to draw the lattice for a given order, a given property set, or (in Ch. 4) for a given structural object model. Prolog is designed to mimic the predicate calculus, yielding in effect a computational analog to the world of pure logical definitions. While Prolog's characteristic use of queries and proof attempts will pervade the exposition, and Prolog makes an extremely convenient choice of environments for our purposes, it should be emphasized that the use of Prolog per se is not critical in any way.

Why Prolog? A Prolog "program" consists of a collection of facts and definitions expressed in essentially pure logical form. The user can then query Prolog about new facts, and Prolog—using a version of the so-called "unification" algorithm—responds by attempting to "prove" the new fact using the the facts and definitions it knows. Hence each "query" consists of a single fact about some set of atomic terms and unbound terms. To each query, Prolog responds by attempting to find a proof of the queried fact, filling in unbound terms with whatever values are needed to make the fact turn out to be true. If Prolog can find no combination of values for unbound terms that make the query turn out true, it responds "no"; otherwise it responds yes and reports the values with which it bound the unbound terms. In choosing a set of values, Prolog in effect momentarily adopts a particular "interpretation" of the world for each attempt at proof, making it a convenient environment in which to compute lattices of category models under, say, different parameterizations of a given object, different concept sets, and so forth.

To make the use of Prolog as clear as possible, we begin by defining an order relation by itself before moving onto lattices.

3.2.1. Prolog defines a relation: `bigger`

Assume that Prolog has just one fact in its database:

```
bigger(elephant,zebra)
```
(Eq.3-1)

which we know to refer the relative sizes of two types of animals (though Prolog knows nothing of the sort, of course!). We can now ask Prolog to attempt to derive new facts from its database: such requests are called "queries". For example, we query Prolog with a fact

```
?- bigger(elephant,zebra)
```
(Eq.3-2)

(The special symbol `?-` is the Prolog query prompt.) Given what it knows, Prolog responds to the query by attempting to prove the queried fact using the facts and definitions that it knows. It answers `yes` if it can prove it, `no` otherwise:

```
?- bigger(elephant,zebra).
     yes.
```
(Eq.3-3)

More interestingly, queries can be written with variables (unbound terms), which always have names that begin with uppercase letters. If a query contains a variable, then Prolog furnishes the user with the values that it had to plug into that variable in order to prove the queried fact. Hence:

```
?- bigger(elephant,X).                                   (Eq.3-4)
      X = zebra
```

The variable X is bound to the value mouse, an atomic term; hence this query acted as an open ended question. After more facts are included in the data base

```
bigger(zebra,cat),                                       (Eq.3-5)
bigger(cat,mouse),
bigger(horse,cat),
bigger(elephant,horse),
```

then Prolog can return as many solutions as it finds proofs, along with the corresponding values for variables:

```
?- bigger(X,cat).                                        (Eq.3-6)
      X = zebra ;
      X = horse ;
      no
```

On the final request for another solution (semicolon), Prolog responds no, meaning that the query cannot be proved in any other ways. Notice that it does not find report X = elephant, because the relation bigger has not been defined as transitive— only as a collection of disjointed facts, among which cannot be found the transitive implication bigger(elephant,cat).

In Eq.3-4 the unbound argument appeared second; in Eq.3-6 it appeared first. There is in fact no restrictions on which arguments can be unbound; if they are all unbound, Prolog simply tries to prove the completely open statement in any way it can:

```
?- bigger(X,Y).                                          (Eq.3-7)
```

```
X = elephant,Y = zebra;
X = zebra, Y = cat ;
X = cat , Y = mouse ;
X = horse, Y = cat ;
X = elephant, Y = horse ;
no
```

Notice that with both arguments unbound, Prolog simply unloads the entire database defining `bigger`.

3.2.2. Automatic construction of lattices

Just as we can assert facts into the Prolog database, we can assert definitions: facts that are true contingent on some set of clauses being satisfied. Thus, just as we defined the two-argument relation `bigger` by a collection of facts, we can define the relation two-argument `subset` by a recursive contingent expression:

(Eq.3-8)

```
/* definition: subset */
     subset([],[]).
     subset(Sub,[L|List]) :-
          (subset(X,List),Sub = [L|X]);
          (subset(Sub,List)).
```

Without delving into the syntax, it suffices to understand that a "list" is an ordered set enclosed by square brackets, which we will abuse by considering as simply a set. Now, when we give Prolog queries involving `subset`, it attempts to prove them, filling in variables as necessary, in much the same way it did so with `bigger` . Hence we have

(Eq.3-9)

```
?- subset([a],[a,b]).
     yes
```

(Eq.3-10)

```
?- subset(X,[a,b]).
     X = [a,b] ;
     X = [a] ;
     X = [b] ;
     X = [] ;
```

no

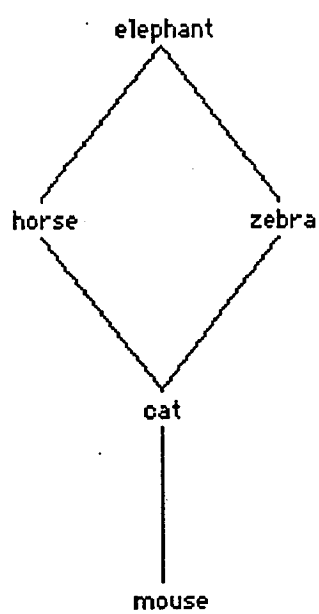
Going back to the relation `bigger`, recall that when we queried Prolog about the it with both arguments open Eq. 3-7, it responded by reproducing the relation in its fundamental set-theoretic form as a set of ordered pairs. The relation `bigger`, as defined, has a particular instrinsic structure—i.e., a particular topology of the connections among the elements in its domain—that can be depicted graphically; this graphic depiction take the form of a "lattice".

We now begin to introduce new definitions pertaining to lattices, some of which draw pictures as side-effects. We can draw the lattice for the relation `bigger` using the single-argument functor called `lattice_of_relation`:

```
?- lattice_of_relation(bigger).
     yes
```
(Eq.3-11)

Prolog's response `yes` in this case is not interesting; the purpose of `lattice_of_relation` lies rather in the picture it draws (which from a purely logical point of view is just a computational side-effect produced en route to satisfying its definition):

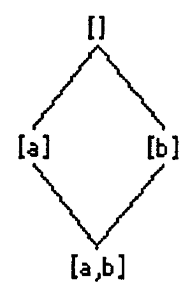
(Eq.3-12)



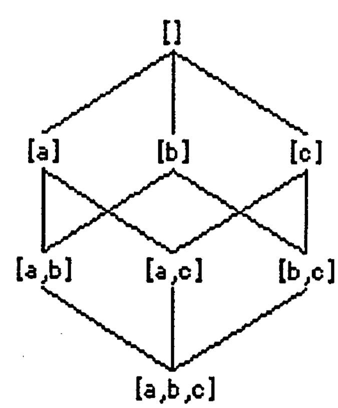
Just as we draw the lattice for the relation `bigger`, we can draw the lattice for the rather more interesting relation `subset` defined over (the power set of) some given set of atomic elements. As discussed above, this lattice become quite interesting if we conceptualize these elements as "properties" or binary facts about entities: then each object in the lattice (that is, each subset of the full property set) is a category model of a larger or smaller degree of specificity or inclusiveness.

Hence we have the lattices produced for the sets `[a]`, `[a,b]`, and `[a,b,c]`:

(Eq.3-13)



(Eq.3-14)

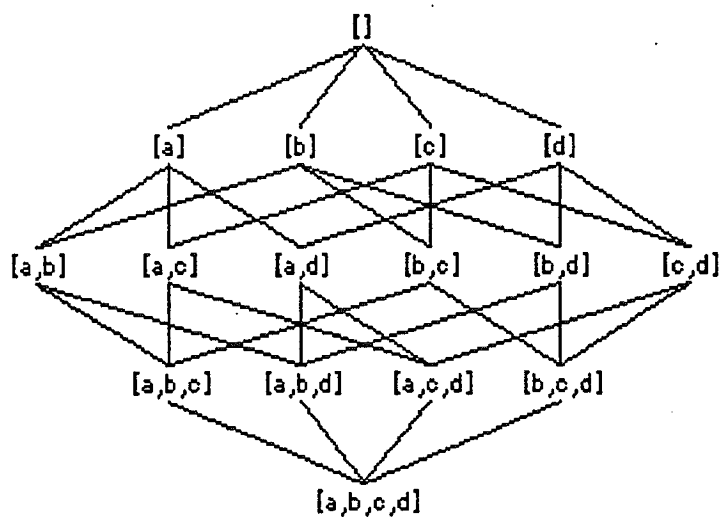


(Eq.3-15)

3.2.3. Geometric interpretation of lattices

Notice how each of these lattices has a neat visual interpretation: as a line, square, and cube respectively; or, putting it another way, as an edge, face, or full projection of a 3-cube. This is of course not a coincidence. Rather, it is because each successive property added to an existing set can be though of as extruding the existing lattice out into a new dimension in a completely regular way. For example, notice in the 2-property lattice, the 1-property lattice appears as the upper left edge, and is then extruded down and to the right into b-space to form the square. Similarly, in the 3-property lattice, the 2-property lattice of [a,b] appears as one face of the cube, which is then extruded out into c-space to form the cube.

This phenomenon extends into 4 dimensions,



(Eq.3-16)

as well as to higher dimensions, but of course the pictures no longer represent their shapes so transparently whenever more than 3 dimensions are being projected into the two dimensions of the page.

This geometric interpretation of lattices is a powerful source of intuitions about them. For instance it makes instantly clear that in any complete (fully crossed) lattice, all the nodes are perfectly interchangeable—just as the vertices of a cube are all symmetric and hence indistinguishable. The implication is that all the category models spanned by a given property set are exactly equivalent under relabelings, a version of the highly counterintuitive "Ugly Duckling Theorem" of Watanabe (1985), which shows that objects in a given set are all exactly equally similar to one another, just so long as their properties are construed in a completely unconstrained way (i.e. just so long as a property is defined as the set of objects bearing that property).

In another way, it is intuitively helpful to think of a 2-property lattice as isomorphic to a square, that is, to a 2-cube or plane; it underlines the fundamental isomorphism between discrete properties and continuous motions, say transverse unit vector translations, any two of which define a plane uniquely. This is of course the essential bridge that connects our analysis of discrete properties with the more general case of continuous generative processes modeled by continuous operations moving along smooth manifolds, as motivated in the previous chapter.

However, there is an important subtlety here. This isomorphism strictly requires that the absence of a property a from a node (category model) is interpreted as meaning that a *may or may not* apply to objects in the category; or, alternatively, that we do not know that a applies to a given member (i.e., a means "we know positively that property a applies"). In this case, for example, given the property set [a,b], the category model [b] denotes a class of objects that have property b and *may or may not* have property a, i.e.

Epistemic interpretation of properties:

$$[b] \Leftrightarrow \{[a,b],[\bar{a},b]\} \qquad \text{(Eq.3-17)}$$

We call this the "epistemic" interpretation, because the presence of a property in a symbolic model refers only to a state of knowledge or certainty about that property—namely, that we know it to be present in objects in the category. Conversely, in models that lack property a epistemically, we simply lack knowledge about their status with respect to a; i.e., they may or may not have it.

It is quite reasonable, alternatively, to interpret the presence of a property "ontically," meaning that it refers to the actual state of the world, rather than to our knowledge of it. On this ontic interpretation, the category [b] refers only to those entities that have property b but definitely lack a:

Ontic interpretation of properties:

$$[b] \Leftrightarrow \{[\bar{a},b]\} \qquad \text{(Eq.3-18)}$$

Some contexts require this latter interpretation. For example, the geometric mapping to the Ugly Duckling Theorem mentioned above only goes through under the ontic interpretation, since two sets with different *numbers* of elements can never be equivalent under any relabeling. Of course, since only the interpretations of the category models, not their symbolic structure, is affected by this distinction, it has no bearing on any of the mechanics; it must be simply be born in mind whenever formal results are interpreted.

3.3. Modes of constraint in lattices

We now consider what lattice forms ensue when constraint is imposed on the relationships among the various category models. Each constrained lattice, in which a certain pattern of causal influence is imposed on a collection of interacting properties, can be though of as a sort of miniature "causal theory" of the miniature world in which it operates. Hence we enumerate lattices in an attempt to guide the interpretation of causal structure, in much the same sense as in Pearl & Verma (1991). There, though, the emphasis is on inferring the true causal sequence—discovering which of two correlated variables is cause and which effect. Here we neglect this issue completely and attempt to infer the purely "logical" causal interaction present in observed predicates.

When we say that the relations among a set of properties are "constrained," we simply mean that some nodes are excluded from appearing in the resulting lattice.

Consider, for example, the lattice

$$\begin{array}{c} [\,] \\ | \\ [b] \\ | \\ [a,b] \end{array} \qquad \text{(Eq.3-19)}$$

from which the single case [a] has been excluded. This simply means that in the model class associated with the property set {a, b}, in this universe, entities never have property a without also having property b. Another way of putting this, of course is that in this universe a implies b, or a → b. That is, a logical constraint relation between properties is associated directly with a lattice. One must be careful, however, to keep straight the distinction between analytic logical implication (e.g., is_a_bachelor → is_male, attaching these properties to their conventional extensions) versus the contingent, synthetic sense meant here (e.g. is_a_bachelor → has_a_messy_kitchen), in which the implication is predicated strictly on a fixed world with certain non-necessary properties. The whole point, in other words, is to represent actually-occurring implications (and so forth) among properties that, from a purely logical point of view , might just as well behave with complete independence.

We can get Prolog to impose this sort of property implication among a given class of properties, and construct the resultant lattice, using the two-argument functor do_lattice. The second argument is a list of "constraints", which can take any of a small number of forms detailed below. The above lattice, for example, was produced on the command

```
?- Set = [a,b], Constraints = [implies(a,b)],          (Eq.3-20)
      do_lattice(Set,Constraints).
```

Prolog first computes the full lattice for the property set {a,b}. Then, it prunes this lattice of those nodes that violate any of the given constraints, and then patches upper and lower nodes together to make the lattice coherent. In this case, the node [a], the only subset of [a,b] that violates implies(a,b), was pruned.

When we add another constraint, a smaller lattice is produced:

```
?- Set = [a,b],
      Constraints = [implies(a,b),implies(b,a)],     (Eq.3-21)
      do_lattice(Set,Constraints).
```

It is natural to think of this fully correlated case, in which the two properties only occur in tandem, as more constrained than Eq.3-19 in the literal sense that it contains fewer cases. This raises the possibility of ranking all possible forms of constraint by their *degree* of constraint—a kind of superordinate codimension. This idea will be taken up below. First we must consider how we to enumerate "all possible forms of constraint."

3.3.1. Enumeration of constraint types for two properties

Now, for a fixed property set, how many different types of logical constraint—i.e., distinct lattices—are there? For the fixed property set {a,b}, clearly any distinct combination of category models forms a distinct lattice, so long as the set of implied orders is nonempty. That is, each subset of the power set of [a,b], forms a unique non-empty lattice under the ordinary subset order. Since there are four elements, [], [a], [b], and [a,b], there are 2^4 or 16 different lattices. Of these, we exclude the empty set {}, the 4 singletons {[]}, {[a]}, {[b]}, and {[a,b]}, as well as the set {[a], [b]}, because in each of these cases the set of *orders* is empty, so there is nothing to draw. This leaves just 7 cases, which are displayed in Fig. 3-1. Also displayed in the figure are the values for `Set` and `Constraints` used to generate the lattices, and an expression for each lattice in a new notation explained below.

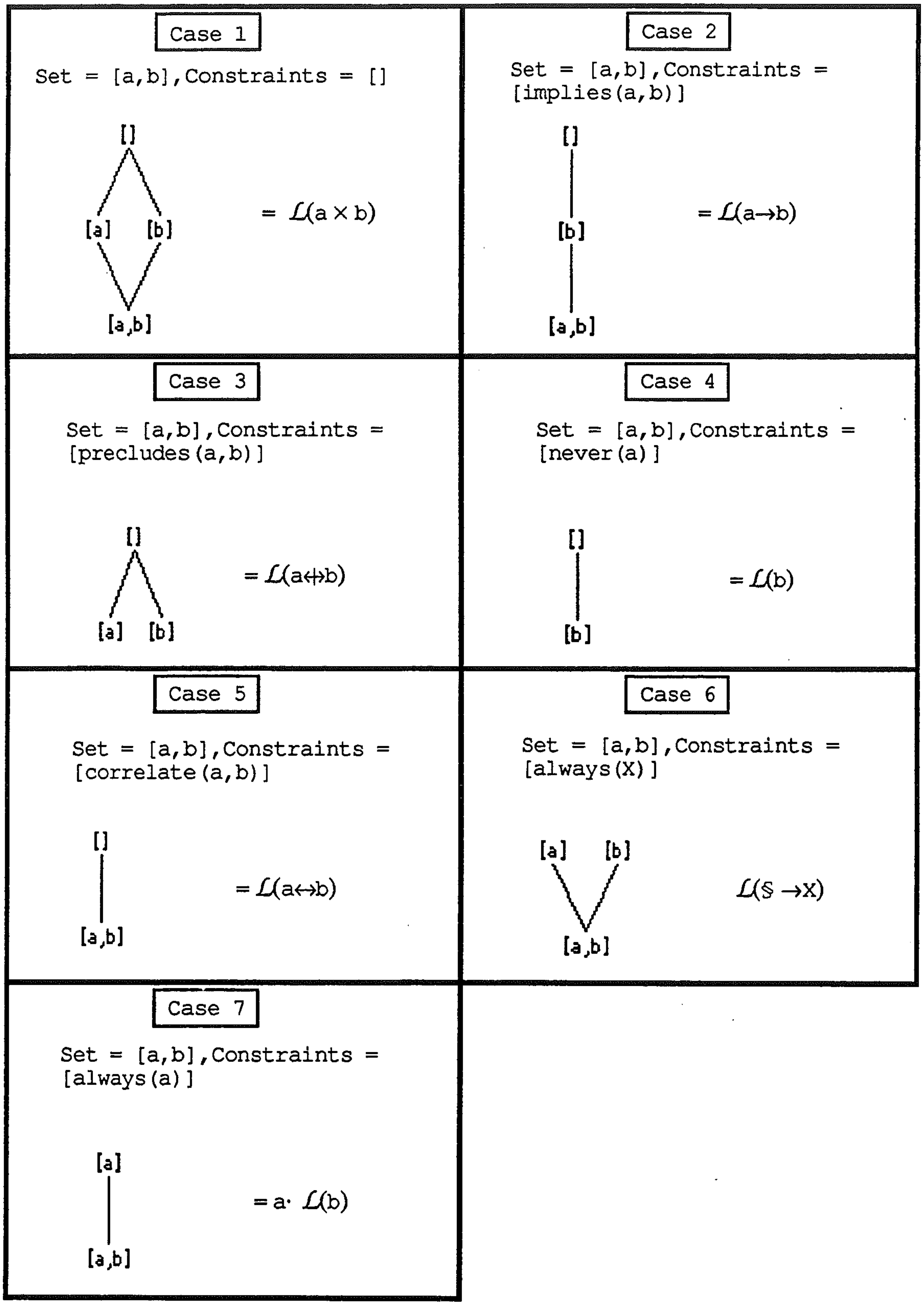


Fig. 3-1 The seven lattices in the two-property family

The forms of constraint embodied in this small set of canonic forms has been operationalized in five Prolog functors:

```
(i)     implies(a,b)
(ii)    correlate(a,b)
(iii)   precludes(a,b)
(iv)    always(a)
(v)     never(a)
```
(Eq.3-22)

which may be found with the lattices they produce in Fig. 3-1. In case 6, `always(X)` (where uppercase `X` is an unbound term) makes a case different from either `always(a)` or `always(b)`; Prolog must bind the term to some available property, with the result that `always(X)` means "always something", so that the corresponding lattice excludes only the empty category model `[]`.

Each of these functors has an intuitive meaning in terms of the relations among category models it allows. `never(a)` simply means that the property `a` never occurs in a category model; i.e. lattices produced under this constraint will contain only nodes lacking `a`. Conversely, `always(a)` requires that property `a` occur in each category model, i.e. that `a` is part of the consistent, underlying structure of entities in the domain, the background against which the other properties may vary. Note that in a larger lattice, the constraint `always(a)` effectively selects the sublattice "hanging" from the node `[a]`; this functor in effect restricts attention to an articulated sub-world of the full world.

`precludes(a,b)` refers to two properties each of which only occurs in when the other is absent, behaving effectively like positive and negative values of some single property. Note, though, that here we have no automatic connotation that the two properties are opposite in meaning. For example, `male` and `female`, in terms of their extensions over animals, are not in any well-defined sense "opposite;" yet simply in terms of the pattern of cases that occur—never both applying to the same animal—they behave as opposite values of one property, namely "gender". Similarly, with the properties "Indian" and "African," as applied to elephants, there is no meaningful semantic connotation of opposition, but the terms act in effect as opposite values of one parameter, namely "species of elephant".

`implies(a,b)` constrains one property, `a`, to occur only when `b` does as well. As such it embodies the idea that some properties are only meaningful when some underlying structure can be assumed: for instance we do not think of an animal as having a long tail unless it has a tail. The formal analog is to measurement along a manifold, discussed in the last chapter, in which we refer to a measurement (say a metric along a surface) that is naturally well-defined only for points that lie along the manifold, i.e. for points that obey some category constraint (or analogously, have some discrete property).

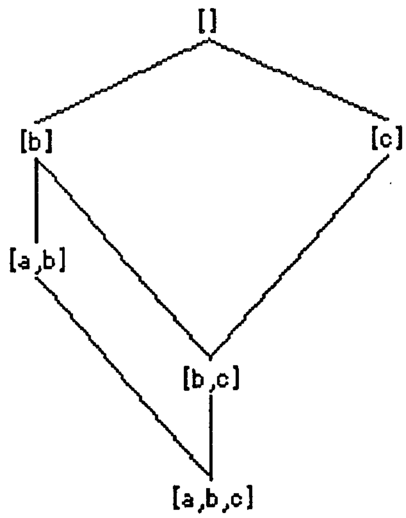
`correlates(a,b)`, in which two properties are constrained to only occur together, is in a sense the case of the greatest interest. It is the natural agent of our central model of a natural category as a collection of mutually co-occurring properties. Its role in allowing the observer to classify observed objects will be elaborated below.

3.3.2. Application of constraint to larger property sets

While the five Prolog constraint functors described above have definitions that come out of the exhaustive enumeration of constraint the two-property universe, once defined they may be applied to larger property sets. We present a few examples here just to give the flavor; obviously as the size of the property set increase it becomes impossible to conveniently catalog the possibilities. However, the critical theoretical question of how well these limited types of constraint do in cataloging the possibilities in higher dimensions will be taken up later (Sec. 3.5).

When one-way implication is imposed on just one pair of properties, we get:

```
Set = [a,b,c],Constraints = [implies(a,b)]
```



(Eq.3-23)

An entire property set can put into a "chain" of consecutive implication:

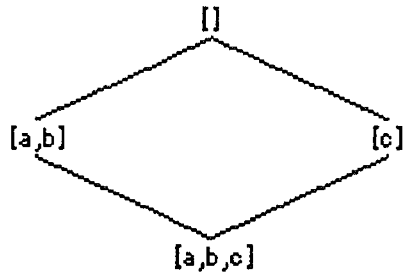
```
Set = [a,b,c],Constraints = [implies(a,b),implies(b,c)]
```

(Eq.3-24)

This particular arrangement can be thought of as imposing a fixed historical sequence on the category models, so that the ladder of codimension and the arrow of time become locally equivalent. That is, each model in an `implies` chain marks a uniform progression from more constrained (bottom) to less constrained (top) models. Then, we might imagine, the models would correspond to the stages of some fixed historical sequence in the development of an object class, in which each stage is characterized by the imposition of one additional property. This idea will be revisited in a more concrete context in Sec. 4.3.3.

Continuing with possible arrangements of three properties, we can have two properties constrained to only co-occur, while another property varies freely:

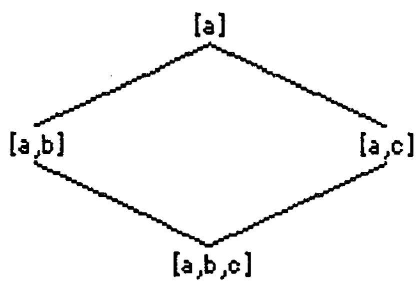
```
Set = [a,b,c],Constraints = [correlates(a,b)]
```



(Eq.3-25)

Here the shape of the lattice makes it pictorially obvious that the correlated pair is now acting exactly as one property, exactly as they should, given that in this world they are not free to vary independently. Similarly, when one property is always present, the resulting lattice looks like a fully crossed lattice of a lower dimension:

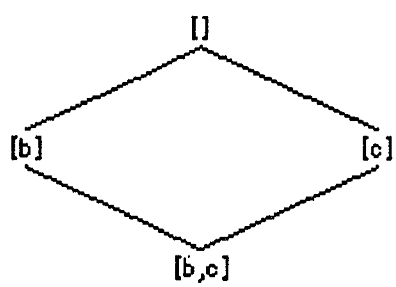
```
Set = [a,b,c],Constraints = [always(a)]
```



(Eq.3-26)

Similarly, when one property is never present, the lattice is literally reduced to a lower dimensional form:

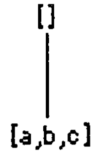
```
Set = [a,b,c],Constraints = [never(a)]
```



(3Eq.3-27)

Finally, and perhaps most interestingly, we have the fully correlated case:

```
Set = [a,b,c],Constraints = [correlates([a,b,c])]
```



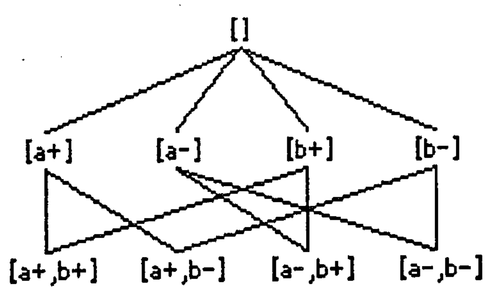
(Eq.3-28)

Notice, critically, that this lattice is *exactly* isomorphic to the fully correlated lattice (Eq. 3-21) of only two properties. That is, one category model in which all properties always occur is just like any other such model, *regardless* of how many properties there are. This simply mirrors the notion that when such a category exists, the actual number of properties that are "really" present is entirely a question of how some observer chooses to express them. In a suitably (and perfectly) constrained world, the three-property entity `[has_fur, bears_live_young, warm_blooded]`, or similarly some much larger property set that a knowledgeable biologist could produce, is no different from the one-property category model `[is_a_mammal]`.

One particularly interesting case of constraint on a property set is the case where four properties are segregated by the `precludes` functor into two signed pairs; for maximum transparency we give the four the names {`a+, a-, b+, b-`}. The resulting lattice,

```
Set = ['a+','a-','b+','b-'],
Constraints = [precludes('a+','a-'),precludes('b+','b-')]
```

(Eq.3-29)



serves as a map of the quadrants in the familiar two-parameter Cartesian space. On top is the "quadrant" comprising the entire space; in the middle, codim-1 row are the four half-space 2-ants; and at the bottom are the four quarter-spaces quadrants. The lattice relation is simply intersection in its usual sense: e.g., the intersection of the half-spaces `[a+]` and `[b+]` is the quadrant `[a+,b+]`, etc.

3.3.3. Minimal representation as implication and the lattice notation

Clearly, the five Prolog constraint functors do not constitute a minimal set, but rather a computationally convenient and expressively transparent implementation that is powerful enough to produce all cases for two properties. In fact, it is possible to conceptualize all five in terms of the single concept "implies", if we add negations and a special new symbol § denoting the proposition that is definitionally always true. Then, taking the symbol "$\rightarrow$" to denote implication, and for convenience adding the symbol "$\leftrightarrow$" to denote mutual implication (i.e. $a \leftrightarrow b$ iff $a \rightarrow b$ and $b \rightarrow a$), we can express the competence of these five functors as follows:

(i)	`implies(a,b)`	$\Leftrightarrow$	$a \rightarrow b$	(Eq.3-30)
(ii)	`correlate(a,b)`	$\Leftrightarrow$	$a \leftrightarrow b$	
(iii)	`precludes(a,b)`	$\Leftrightarrow$	$\bar{a} \leftrightarrow b$ (= $a \leftrightarrow \bar{b}$)	
(iv)	`always(a)`	$\Leftrightarrow$	$\S \rightarrow a$	
(v)	`never(a)`	$\Leftrightarrow$	$\S \rightarrow \bar{a}$	

Critically, this allows each constraint to be identified with a well-defined number, the number of primitive implications it contains. For instance, `never(a)` is really one implication; `correlate(a,b)` is really two. In the syntax of Prolog, the complex construct `correlate([a,b,c])`, in which the three properties `a`, `b` and `c` are each constrained to correlate with one another, is really $2\binom{3}{2}$ or 6 primitive implications.

Lattice notation. This table allows us to introduce a very convenient notation for lattices, so that we can refer to a lattice without drawing it completely. Since each combination of implied properties makes a unique lattice, we simply denote the corresponding lattices $\mathcal{L}(a \rightarrow b)$, $\mathcal{L}(a \leftrightarrow b)$, $\mathcal{L}(\bar{a}$ $)) \leftrightarrow b)$, $\mathcal{L}(\S \rightarrow a)$ and $\mathcal{L}(\S \rightarrow \bar{a})$, as is done in the appropriate places in Fig. 3-1. We add to this list the unconstrained lattice $\mathcal{L}(a \times b)$, the notation suggesting that the two properties are fully crossed.

Finally, to make this notation more expressive, we invent a kind of "lattice multiplication": take $c \cdot \mathcal{L}(a \times b)$, for example, to denote the lattice created when the property c is inserted into each node of the lattice $\mathcal{L}(a \times b)$; that is, property c is distributed throughout the lattice.

3.3.4. The superordinate lattice

As discussed above, each lattice for a fixed set of properties, that is, each kind of constraint, corresponds to a world in which properties are bound to occur only in some combinations. One may think of this constraint as the result of physical, biological, mechanical (etc.) laws operating on the entities involved forcing some properties to imply others due to causal laws described by the appropriate domain-specific scientific theory. Similarly, one may think of this constraint as deriving from the "meaning" of the properties themselves, so that the logical implications among a set of properties are definitionally part of their semantics under some interpretation. Below, we will think of each form of constraint as a kind of working hypothesis an observer uses in order to generate theories about the way entities it observes should best be categorized. In order to make this notion explicit, we need to introduce an ordering on the various modes of constraint, and ordering that makes explicit which of them correspond to more or less "flexible" theories about the model class: that is, more or less constrained kinds of constraint. Perhaps unsurprisingly, this ordering also turns out to be a lattice, which we call the "superordinate lattice" because its nodes are themselves lattices.

Recall that each case in Fig.3-1, the lattice is completely determined by the set of cases that appear on it (i.e., the orders follow from the cases via the subset ordering). Hence we can put all the cases into a larger lattice, in which each each lattice is the lower neighbor of another lattice that is minimally more inclusive, that is, contains exactly one more allowable category model. Fig. 3-2 shows this lattice for the two-property family or model class. For greatest clarity, we now include the three cases that are redundant under relabeling of a and b (There are of course still no lattices associated with the empty set, the four singleton sets, or the set {[a],[b]}, in all of which cases there are no orders to depict.) At the top of the superordinate lattice appears the fully crossed case, in which the properties are allowed to occur in any combination. At the bottom in the middle appears the fully correlated case, in which each property appears only if accompanied by the other: the two-property analog of a category prototype in which some natural kind corresponds to a set of properties that normally all occur in unison in generic members. It is worth examining the lattice for a moment to appreciate its extreme symmetry.

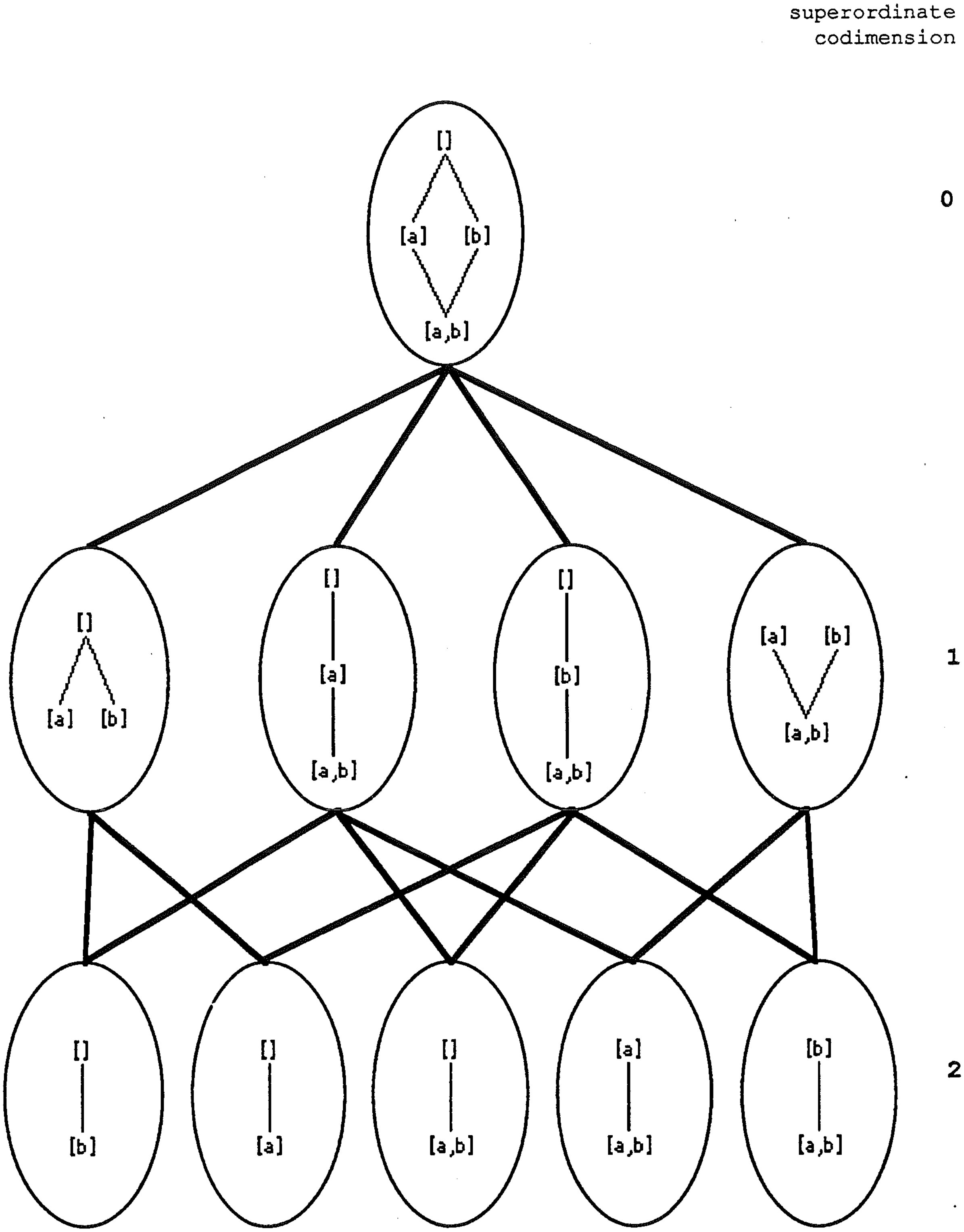


Fig. 3-2 The superordinate lattice for the two-property family

The same lattice is shown in the compact notation in Fig. 3-3. While the larger form is somewhat large and unwieldly, it is relatively transparent to read because one can comprehend the subset relations simply by inspection. The compact version is, by contrast, somewhat opaque.

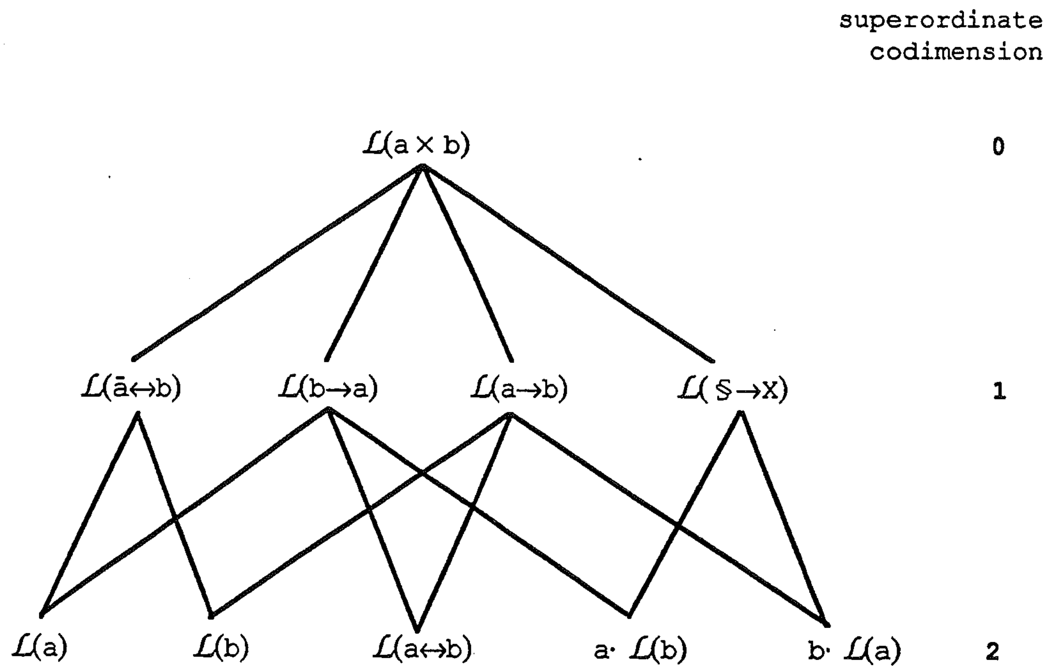


Fig. 3-3 superordinate lattice for the two-property family, in compact notation.

The superordinate codimension is clearly well defined in these pictures: all the lattices in the middle row, for example, have superordinate codimension 1, meaning that they are constructed conforming to exactly one more implication constraint than the unconstrained case (that is, conforming to exactly one implication constraint, since the unconstrained case by definition conforms to none). Consider a lattice $\mathcal{L}(\mathbf{C})$ that obeys a set **C** of pure implication constraints, containing a number $|\mathbf{C}|$ of pure implications. For instance with **C** = {`implies(a,b),implies(b,a)`}, we have $|\mathbf{C}| = 2$. In general then we have

$$\text{superordinate codim } \mathcal{L}(\mathbf{C}) = |\mathbf{C}|. \qquad \text{(Eq.3-31)}$$

Finally, we note that superordinate lattices can be constructed for higher numbers of properties in exactly the same manner as with two. So for instance in the superordinate lattice for the property set {a,b,c}, the lattice $\mathcal{L}(a \times b \times c)$ (Eq. 3-15) would be the top node, and the more constrained world of the lattice in Eq. 3-23 would appear as a node with superordinate codimension 1, and the maximally constrained world of Eq. 3-28 would appear on the bottom row as a node with the maximum superordinate codimension of 2 * 3 = 6.

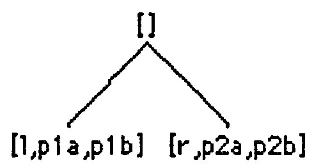
3.4. Category inference using discrete lattices

With all the machinery of lattices and superordinate lattices in place, we are now in a position to express a hypothesis about how categories are induced from observations. Each lattice, that is, each node on the superordinate lattice of a given property set, amounts to a hypothesis about what the constraint relations are among the properties in the current world, with the resulting set of possible category models in the model class. Basically, the idea is that the observer adopts a conservative (in the sense of Osherson, Stob & Weinstein 1986) strategy of starting at the bottom of the superordinate lattice—i.e., starting with the most constrained hypothesis—and only climbing the lattice as observation requires. This strategy will have to be combined, we will see, with domain-specific knowledge of and hypotheses about the meaning of the various properties, in order to enable the observer to arrive at reasonable conclusions that are pitched at the "correct" level of generality.

3.4.1. The canonic Bongard problem

We start by considering the classic induction problem, the Bongard problem. In its simplest, schematic form, a Bongard problem consists of a collection of pictures on the left satisfying some property p1 on the left, and another collection on the right satisfying some disjoint property p2 (Fig. 3-4).

```
Set = [l,r,p1a,p1b,p2a,p2b],
Constraints =   [precludes(l,r),
                 correlate([l,p1a,p1b]),
                 correlate([r,p2a,p2b])]
```



(Eq.3-39)

The lattice, notice, is again exactly isomorphic to the case with only one property on each side: the change in dimension does not change the essence of the situation at all, which is that there are two cluster categories, one on each side. However, the superordinate codimension does increase—from 4 as discussed for the case above, to 6 here. The increase is best thought of not as an increase in the structure of the problem model (since in fact it is exactly as structured as before) so much as an increase in the naïveté of the prior model embodied by the choice of properties: the larger property set stands in for a flexible, unstructured prior model of a world that turns out to be quite structured.

3.4.2. The Bongard problem without labels

In a classic Bongard problem, the classes are given, and only the correct descriptions are to be discovered. In the world, everything must be inferred. However, the distinction between given and inferred information blurs when both are viewed simply as aspects of world structure. Consider, for example, the collection of objects depicted in Fig. 3-6.

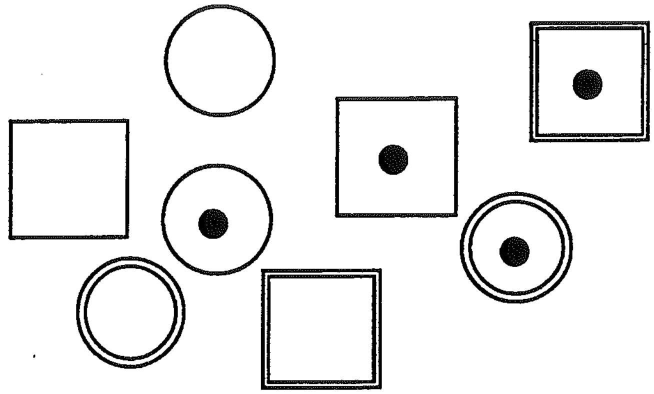


Fig. 3-6. What are the categories?

Coding this world (arbitrarily) as:

p = "square", $\bar{p}$ = "circle"
q = "single border", $\bar{q}$ = "double border"
r = "dot", $\bar{r}$ = "no dot"

we can present it symbolically as in Fig. 3-7.

$\bar{p}q\bar{r}$ $pq\bar{r}$ ...

$\bar{p}q\bar{r}$

$p\bar{q}r$

pqr

$pq\bar{r}$

$\bar{p}qr$

$\bar{p}\bar{q}r$

$\bar{p}\bar{q}\bar{r}$

$p\bar{q}\bar{r}$

Fig. 3-7. A schematic cluster of objects with properties

What are the categories would it make sense to extract from this world? While one one might be tempted to simply carve up the objects according to one of the available properties—e.g., all the objects that have property p versus all those that don't—which property to choose? None of the three is inherently more compelling than the other two. The problem, really, is that there are no natural categories in this world—or, more precisely, there are no natural categories in the {p, q, r} representation of this world, or in the projection of this world into {p, q, r}-space.

By contrast, consider a more structured world— one that is defined by the same three properties, but where some structured covariation holds sway among some of them (Fig. 3-8).

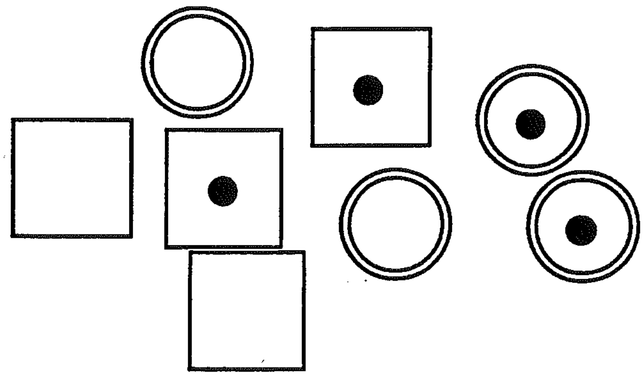


Fig. 3-8. A more structured world.

Using the same coding scheme, we have Fig. 3-9.

$$\bar{p}\bar{q}\bar{r} \qquad pqr \qquad \bar{p}\bar{q}r$$

$$pq\bar{r} \qquad pqr \qquad \bar{p}\bar{q}r$$

$$\bar{p}\bar{q}\bar{r} \qquad pq\bar{r}$$

Fig. 3-9. A more structured world, coded.

The structure that is now present provides the inferential leverage an observer needs in order to hypothesize a decomposition. By contrast with the unstructured world in which no compelling natural categories could be detected, this world seems to have two distinct categories (square single-bordered things and circular double-bordered things) with a third property (presence of a dot) contributing nothing (i.e. interacting orthogonally with the other two properties). While the unstructured world corresponds to the canonic crossed lattice $\mathcal{L}(p \times r \times q)$ (Eq.3-15), the more structured world has the form $\mathcal{L}((p \leftrightarrow q) \times r)$:

$$\begin{tikzcd}[column sep=small] & {[\,]} \arrow[dl, dash] \arrow[dr, dash] & \\ {[p,q]} \arrow[dr, dash] & & {[r]} \arrow[dl, dash] \\ & {[p,q,r]} & \end{tikzcd} \quad = \quad \begin{tikzcd}[column sep=small] & {[\bar{p},\bar{q},\bar{r}]} \arrow[dl, dash] \arrow[dr, dash] & \\ {[p,q,\bar{r}]} \arrow[dr, dash] & & {[\bar{p},\bar{q},r]} \arrow[dl, dash] \\ & {[p,q,r]} & \end{tikzcd} \qquad \text{(Eq.3-40)}$$

containing only those cases that actually appear in this particular constrained world.

The concept of superordinate codimension makes the distinction between a less structured and a more structured world explicit. While the world of Fig. 3-6, in which the three properties were fully crossed, had superordinate codimension 0 in the three-property superordinate lattice—that is, it was generic and hence unstructured with respect to this property set—the world of Fig. 3-8 has superordinate codimension 2. Notice that this codimension corresponds to the 2 independent implications that are in force in this structured world ($p \rightarrow q$ and $q \rightarrow p$), in exactly the same way that the continuous codimension in Ch. 2 corresponds to the number of independent constraints on (degrees of freedom removed from) the constraint manifold.

This analysis brings to the surface the deep isomorphism between this problem and an ordinary Bongard problem, in which the categories are given in spatially segregated form, one categorie on the left and the other on the right. Here, the additional property (q) correlated with p, and thus adding structure to the otherwise random property mix, acts exactly like the "side" (right or left) variable in a Bongard problem. In both cases the extra structure acts to sort the objects into their natural classes. The more general point is that there is no real distinction between a labeled and an unlabeled problem: in both cases, some correlative structure provides the inferential leverage needed for the category structure of the world to make itself known.

3.5. Modality in the world vs. modality in our heads

We now take up the critical question of whether the five constraint functors introduced in the last section can accommodate all possible forms of constraint in higher dimensions. Previewing, the answer will that while there are many possible arrangements of constraint that these five functors alone *cannot* produce, we will propose that human categorizations—the hypotheses that people are willing to generate and believe about what kind of constraint is governing the world they observe—are restricted to these five. The reason, as with all forms of constraint underlying human perceptual competences, is that restricting hypotheses in this manner amounts to adopting a model of the world as a deeply regular and constrained place: another version of the natural modes hypothesis.

3.5.1. Five forms of constraint don't suffice

First, we note that with more than two properties, some lattices cannot be constructed simply by implication, and hence cannot be constructed using the five Prolog functors, and equivalently do not appear anywhere on their respective superordinate lattices. In fact, as the number of properties increase, the lattices that can be expressed by implication alone—the "pure implication lattices," which for N properties we call $\{\mathcal{L}(C_N)\}$—become an ever-smaller minority. A simple counting argument suffices to see this. For N properties, $\{\mathcal{L}(C_N)\}$ numbers no more than the number of different sets of primitive implications. A single implication takes two distinct arguments, and there are $N + 1$ possible arguments (the N properties plus the certain property §). The number of primitive implications is no more than $N(N + 1)$. Hence the number of sets of such things, i.e. of distinct pure-implication lattices is

$$|\{\mathcal{L}(C_N)\}| \leq 2^{N(N+1)}. \qquad \text{(Eq.3-40)}$$

On the other hand, to count the number of lattices for N properties, consider that every non-empty set of non-empty subsets of the N properties, when combined with the empty set `[]`, forms a lattice, because the empty node will be a subset (and hence an upper neighbor) of every other node. There are 2^N - 1 such nodes to choose from, and any set of them is a lattice. Hence the total number of lattices for N properties satisfies the inequality

$$|\{\mathcal{L}_i\}| > 2^{(2^N - 1)}. \qquad \text{(Eq.3-41)}$$

Since $2^{(2^N - 1)} > 2^{N(N+1)}$, we can conclude that the total number of lattices with N properties is absolutely larger than the number of pure-implication ones,

$$|\{\mathcal{L}_i\}| \quad > \quad |\{\mathcal{L}(C_N)\}| \qquad \text{(Eq.3-42)}$$

and this discrepency increases with N, as the difference between 2^N - 1 and $N(N$-1) increases.

An example of a lattice that is not a pure-implication lattice can be easily constructed. Consider a fully crossed lattice with four or more properties, {p,q,r,s...}. Now, we puncture this lattice by removing just one node in the codimension-3 row, say the node [p,q,r]. The resulting lattice is non-pure—that is, it could not have been produced by any constraint set. To see why, consider what its constraint set would have had to contain. Since the lattice contains [], we know it has no instances of always. Since its bottom node is the set of all properties, we know it has no instances of never. Since it contains all the singleton nodes in the codimension 1 row [p], [q], etc., the constraint set cannot contain any instance of implies. Since it contains all the two-element codim 2 nodes, the constraint set cannot contain any instances of precludes. Hence its constraint set is empty. Yet, it is not the lattice produced by the empty constraint set—i.e. it is not the complete fully crossed lattice—because by stipulation it is missing the node [p,q,r]. Pure-implication lattices—worlds in which the categories obey this simple kind of causal law governing their properties—are thus a specialized subset of all worlds.

Projection of lattices into "planes." It should be immediately clear that a world in which all the natural model classes were pure-implication lattices would be a very "kind" world indeed. One way to see this is to note the extreme computational transparency of the implication networks. A pure-implication lattice of N properties, since it is fully defined by the implications among pairs of its properties, can be completely recovered from its $\binom{N}{2}$ projections into two-space, i.e. into the various "planes" <a-b>,<a-c>,<b-c>, etc. For example, a lattice in a universe {a,b,c} has a projection into <a-b> space, constructed by simply collapsing over c in all nodes in the lattice. This projection, itself a lattice in a-b, reveals whatever form of implication existed between a and b in the original lattice (e.g., a→b, b→a), since it has to take one of the canonic forms; similarly for the other two projections. Then the lattice for {a,b,c}, stipulated to be pure-implication, can be constructed simply by taking the implications found in the three projections, and bringing them to bear on the full lattice of {a, b, c}.

Note, critically, that if this procedure is executed on a lattice that is *not* pure-implication—e.g. that has missing cases—the result will nevertheless necessarily be a pure-implication lattice. The procedure thus "regularizes" non-pure lattices to pure ones. Another way of putting this is that if an observer adopts the constraint that natural categories must be pure-implication lattices, as a regularizing constraint on its interpretations of the world (which is exactly what we propose in the next section) then this inferential procedure will have the effect of "regularizing" the categorical structure of the world.

3.5.2. Discrete natural modes

Without any further fanfare, then, it is now possible to propose a discrete-properties version of a Natural Modes hypothesis (again, cf. Marr's 1970 "Fundamental Hypothesis").

Natural Modes, discrete properties version

- Categories in the world are interpreted as pure-implication lattices.

The intuition behind this is as follows: natural categories can be meaningfully assumed to arise out of constraints from the physical (etc.) worlds that act in a causally manner; causal effects take the form of implication; and pure-implication networks (by definition) are the result when implication only governs the relations among properties. That is, wherever coherent causality reigns, pure-implication lattices occur.

Of course, there can be cases missing in the real world, due to hidden constraints or constraints involving hidden variables. In such cases, the Natural Modes constraint will lead an observer astray, and the observer's interpretations will be more regular than reality. But such cases are the exception exactly to the extent that the world is coherent in the intended sense. One may also view this constraint as creating an expectation of hidden variables in operation, whenever observable variables seem to be behaving incoherently, an inference process that is in fact universal in scientific reasoning. Without a constraint of modal causality—e.g., if arbitrary nonimplication lattices were admitted into the theoretical descriptive language—there would of course be no reason to postulate hidden variables, since the behavior of the observed variables would fall within the range of the accepted model.

Finally, we note that pure-implication lattices have a deep isomorphism with the concept of smoothness of constraints governing world categories. In both cases, we observe that categories arise from a consistent model of "coherence" are a special subset all possible world arrangements. Moreover, the two models of coherent world regularity (constraint manifold and discrete implication) are analogous to one another: in both cases, knowledge of the value of one property is dispositive to knowledge of the value of another property. This correspondence can in fact be made more explicit. A sketch of a formal classification of the correspondence, based on the differential geometry of possible manifold shapes, is presented in Appendix A: the aim there is the construction of a detailed correspondence between discrete superordinate lattices on the one hand and smooth Morse forms of constraint manifolds on the other.

- Chapter 4 -

Lattices of Structural Category Models

4.0. Chapter preview

This chapter will explore the category structure that may be found within some simple families of planar figures—triangles, V's, and lines with dots, 4-gons, pairs of lines, etc. First we extend the computer system from the Ch. 3 to construct visual objects defined by parameter sets, and then we explore the application of constraint to the relations among these parameters.

Sec. 4.1 describes the extension of the Prolog definitions described in Ch. 3 to model classes of objects, defined structurally. That is, we define a model class as a generic object: e.g. a line and a dot, without any relationship between them specified. Then the program constructs the lattice of category models for this class, using "special values" of structural parameters as discussed in Ch. 2 (right angles, equal length of line segments, etc.) as the modal non-genericities marking the transitions between category models of different codimensions.

Sec. 4.2 develops a notion of psychological *typicality* that can be incorporated into the automatic generation of examples of object models. A typical example of a model class is one picked in a biased way to avoid confusion with lower-dimensional cases, and hence unlike a random example is guaranteed to appear "genuinely" generic. In Sec. 4.2.1, using this definition, the program generates examples which can be readily identified with the intended generating class, allowing for the construction of "synthetic Bongard problems." This definition of typicality lays the groundwork for the probabilistic notion developed in Ch. 5.

Sec. 4.3 describes how constraint can be imposed on these lattices of objects models, in exactly the same way as was investigated in Ch. 3, with the effect of imposing some organizion on the full lattice and reducing its size. We consider three particular interesting cases of constraint application. First, we show how the functors may be used to eliminate degenerate cases from the lattice. Second, we show how the functor `always` has the effect of defining a "context" for a class of objects models (i.e., an aspect of structure that is always present, and hence becomes "generic"). Finally, we show how an implication chain, constructed using the `implies` functor, can serve to model a fixed historical sequence of stages in the development of objects in the class.

4.1. Lattices for object model classes

In Ch. 3 we introduced a Prolog system for constructing lattices automatically for object models defined only in terms of blank properties. We now extend that system to simple structural category models defined in terms of the spatial relations among constituent parts.

4.1.1. Motivation for structural model classes

We will use the same notion of varieties of constraint developed for discrete properties in the last chapter, calling upon Prolog to construct lattices for each family and to draw examples of each category model that appears in each lattice. Then it will be possible to judge by eye whether each category model corresponds to a conceivably "natural" category of shape, to the extent that just a small number of examples allows the observer to guess correctly the generating category of each example. The structure of the each lattice, we will see, hinges principally on two choices: 1) a choice of a "generating process" for each family of shapes (i.e., each model class; see Sec. 2.2), which leads directly to a choice of parameterization; and 2) a set of "concepts," meaning a prior selection of special values for each type of parameter; right and parallel for angles, equal length and zero length for line segments, etc. (see Sec. 2.2.5). Note that the selection of "concepts" is a one-time choice for each model class: after it is made, a multitude of category models appear epiphenomenally on the lattice, categories of shapes that were not put into the lattice by hand in any way.

Furthermore, we can impose on the resulting lattice additional constraints of the types proposed in the previous chapter. For instance, just to give the flavor, the functor `always`, which picks out some otherwise special property and constrains the entire model class to satisfy it, can here be used to create specialized sub-worlds or "contexts." In the general 4-gons lattice, for instance, having two parallel lines is a special (codimension 1) category; when we apply `always` to this property , on the other hand, we effectively create a new model class and lattice in which having two parallel lines is generic, since it is true of the top node as well as all other nodes.

One intriguing possibility raised in this chapter—made possible by Prolog's construction of a lattice of 4-gons in which the category model concepts seem to mirror the category construction schemes human observers use—is the possibility of automatically generating an entire catalog of "synthetic Bongard problems." This is more of a thought experiment that a practical proposal, since the prospect of examining the resulting problems is a tedious one. A few examples suffice to demonstrate the significance of the idea, though, that these problems, generated in a completely automatic way from the lattice by juxtaposing nodes (and other modal groupings) from the lattice, might be generally solvable by human observers. This suggests that the construction of categories by composition of concepts in the manner built into the lattice-building is fundamentally on the right track.

The critical mapping connecting the last chapter with this one is the analogy between the discrete properties of the last chapter and the "special values" along continuous parameters of this one. Everything else, in essence, follows from this idea. In Ch. 3, one moved down a lattice from more inclusive models to more specific ones by adding properties, or (epistemically) by gaining more specific knowledge about properties. Here, one category model collapses modally to a more constrained one when some meaningful parameter of the structure of an object—say, the angle between two segments—takes on a special value—say, 90°. Hence "has a right angle" plays exactly the role of one of the discrete properties as treated in the last chapter. If (critically) the parameterization of the object's structure has been well chosen, then properties taken as special values gain, potentially, two sources of "meaningfulness": first, that the "collapsing" parameter is a meaningful parameter of structure, one that is liable to interact causally with the identity of the object, and second, that the special value is a potentially causally meaningful special case, such as a right angle.

A right angle might be physically meaningful, for example, in the relationship between the directions of two forces, which fail to mix (add) only when they are orthogonal; similarly, "parallel" is physically the case when they mix in a pure fashion. When the chosen parameterizations and the special values mirror physical (and other) constraints and regularities in the world, the categories constructed by composition of these special values along these parameterizations—the classes of objects liable to behave qualitatively the same with respect to reasonable hypotheses about constraints to be encountered in the world—will be wise choices. Hence moving from arbitrary discrete properties to special values in object parameterizations moves us a step closer to the goal of inferentially reasonable categorizations that match world regularities.

In the next section the basic pieces of the program will be described, with more and more complex pictorial examples being described along the way.

4.1.2. Object definitions as generative models

We first define a "model class" or generic object, simply as a group of points and defined connections (drawn-in segments) among them. The program will then pick parameters that fully define this set of points and, using a set of special value concepts, build a lattice.

We use V's as a first example. A V is defined by the assignments:

```
Object = [p1,p2,p3],Connections = [(p1,p2),(p1,p3)]
```
(Eq. 4-1)

defining three points with the names `p1`, `p2` and `p3`, with the segments `p1p2` and `p1p3` drawn in, forming a V with fulcrum at `p1`. The program then compiles a list of parameters—angles between segments and ratios of lengths of segments—that fully defines all segments, recursing backwards to the primitive first segment. That is, the first segment (here `(p1,p2)`) is defined; any other segment is defined if both its length and its orientation are referenced back to a defined segment. For V's, then, since there are only two segments, the second segment has both length and orientation defined with respect to the first segment. Hence the final set of parameters includes, naturally enough, the orientation difference between the two segments (i.e., the angle of the V) and the length ratio between the two segments. Since values of these two parameters completely define the structure of each V (up to scale, position and overall orientation), it is natural to think of them as a crude model of a "generative process": that is, a list of the parameters governing the generation of the object. Here, there was only one choice: with more points, there will be a number of alternatives. The program now sets about trying to plug various special values into these parameters, and constructing a lattice accordingly.

4.1.3. Special values as concepts

The program is "born with" a short list of special values for each type of parameter (orientation difference or length ratio); we call these the concepts (see Sec. 2.2.4). Concepts can be added or dropped at will, making it possible to test instantly whether different combinations lead to more or less psychologically correct categories. A typical initial state might be:

```
current_concepts(X).                                        (Eq. 4-2)
   X = [right,parallel,equal_length,coincident]
```

Each of these concepts literally corresponds to a value plugged into one of the parameters. `right` and `parallel` refer to angular differences, `right` meaning 90°, and `parallel` meaning either 0° or 180° (depending on the sequence of the points). `equal_length` and `coincident` refer to ratios between segment lengths, `equal_length` meaning 1 and `coincident` meaning 0.

Each node—category model—in the resulting lattice is a full set of defining parameters bound to values. Each parameter takes either a special value or else a reserved non-numeric value `gen`. When each category model is drawn, parameters with the value `gen` are given random numeric values: uniformly around the circle for angles, and drawn from a Gaussian distribution about a standard length for segment lengths.

(Note though that strictly speaking these values are *random* rather than psychologically *generic*. That is, they are not prevented in any way from accidently taking on special values or values so near special values as to appear special, and in fact some examples will show that this leads to some figures that are peculiarly difficult to interpret. In Sec. 4.2 the critical issue of assuring psychogical genericity for non-special values will be elaborated more fully.)

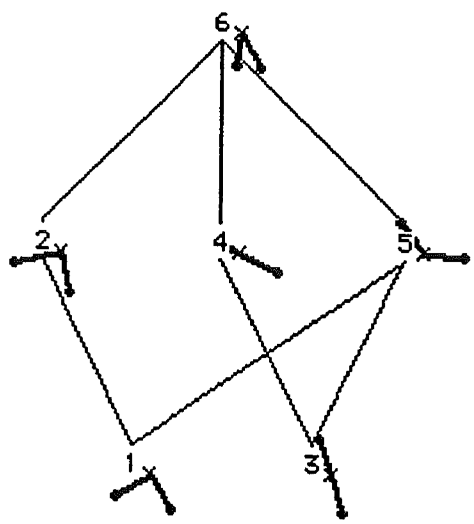
In a sense the V family is the canonic case to illustrate the special values, since it has just one angular difference and one length ratio. A generic V, as we would expect to encounter at the top of the V's lattice, would be the parameter list

```
[(a(p1,p2) - a(p1,p3), gen),                                (Eq.4-3)
  d(p1,p2) / d(p1,p3), gen)]
```

while a right-angled V would be

```
[(a(p1,p2) - a(p1,p3), 1.570795),
  d(p1,p2) / d(p1,p3), gen)]
```
(Eq.4-4)

Because this right-angled V is a specialization—in exactly one parameter—of the generic V, we would expect to find it attached as a lower neighbor, that is, as a codimension 1 node in the complete Vs lattice. This is confirmed when the complete lattice is drawn (Fig. 4-1).




```
Object = [p1,p2,p3],Connections = [(p1,p2),(p2,p3)],
              categories(Object,Connections,Tree,Object_list).
    Object  = [p1,p2,p3],
    Connections  = [ (p1, p2), (p2, p3)],
    Tree = [[p1,p2],[p2,p3]],
    Object_list =
         [ (1, [ (a(p1,p2) - a(p2,p3), 1.570795),
                (d(p1,p2) / d(p2,p3), 1)]),
           (2,
             [ (a(p1,p2) - a(p2,p3), 0), (d(p1,p2) / d(p2,p3), gen)]),
           (3,
             [ (a(p1,p2) - a(p2,p3), 0),(d(p1,p2) / d(p2,p3), 1)]),
           (4,
             [ (a(p1,p2) - a(p2,p3), 1.570795),
               (d(p1,p2) / d(p2,p3), gen)]),
           (5,
             [ (a(p1,p2) - a(p2,p3), gen), (d(p1,p2) / d(p2,p3), 1)]),
           (6,
             [ (a(p1,p2) - a(p2,p3), gen),
               (d(p1,p2) / d(p2,p3), gen)])]
```

Fig. 4-1. V's lattice, and object list.

Prolog numbers the nodes arbitrarily and (as discussed above) draws a random example of each model nearby, with the point `p1` always drawn in as a light X-mark rather than a dot, for reference. Here we find, at the top, the generic V, with generic angle and generic ratio of line lengths. In the codimension-1 row, we find two different specializations of the angle: `right` and `parallel` (see nodes 2 and 4 in the object list). Notice that here `parallel` amounts to "collinear" only because of the structural definition of a V, in which these two segments are coterminous by definition. Naturally, `right` and `parallel` , being disjoint special values of the same parameter, do not have an intersection. However, each of them has an intersection with the other codimension-1 node, "isosceles." The result in the codimension-2 row are two distinct "origins," or maximally constrained category models: a right isosceles V (node 1) and a bisected straight line (node 3). Notice that some of the concepts here—all those with codimensions larger than one—are in no sense "given" to Prolog, but emerge from legal composition of given concepts.

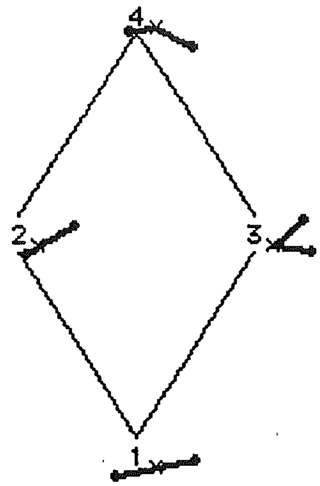
As mentioned above, we can change the concept list at will. The command

```
?- drop(right).                                                    (4-5)
        yes.
```

reduces the concept list:

```
?- current_concepts(X).                                            (4-6)
        X = [parallel,equal_length,coincident]
```

Now we rerun the V's definition again (Fig. 4-2).




```
Object = [p1,p2,p3],
Connections = [ (p1, p2), (p2, p3)],
Tree = [[p1,p2],[p2,p3]],
Object_list = [
        (1, [ (a(p1,p2) - a(p2,p3), 0), (d(p1,p2) / d(p2,p3), 1)]),
        (2, [ (a(p1,p2) - a(p2,p3), 0), (d(p1,p2) / d(p2,p3), gen)]),
        (3, [ (a(p1,p2) - a(p2,p3), gen), (d(p1,p2) / d(p2,p3), 1)]),
        (4, [ (a(p1,p2) - a(p2,p3), gen), (d(p1,p2) / d(p2,p3), gen)])]
```

Fig. 4-2. Another V's lattice, without "right."

Now of course not only does the right-angled codimension 1 case disappear, but so do all its lower neighbors, in this case eliminating the second origin.

4.1.4. Concepts are keyed to connections

In some cases, in order to give the concept definitions a full measure of psychological correctness, they are keyed to the drawn-in line segments. The special value "right," for example, only applies to two drawn line segments, while the special value "parallel" applies regardless. The result is that adding or subtracting one line segment—while leaving the essential structure of the object the same—leads to a completely different lattice.

For example, let us construct the lattice for an object defined as three dots but with only the first two connected. The result is a segment and a dot. With `right` restored to the concept list, we run the same three-dot definition with one connection omitted (Fig. 4-3).

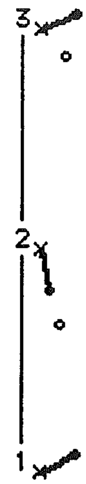


Fig. 4-3. Lattice for line and dot.

The result is just three categories: generic, collinear, and at the bottom, coincident (the dot is drawn over the reference point p1). Note that here "dot collinear with line" and "dot on line" are equivalent, since the natural concept "touching" is absent from the concept set we have used. This concept would have to be added to make the lattice more perfectly psychologically valid.

The need to hinge concepts to the connection structure of an object definition allows a fairly strong conclusion. Three dots are, after all, just three dots in terms of the information present: no information is gained or lost per se when a line segment is added or erased. However, the human observer apparently does draw some conclusions from these things, and as a consequence the formal structure of our categories and the topology of their relations to one another are altered. Despite the informational equivalence, though, such conclusions may easily be seen to be justified for an observer who seeks to make reasonable guesses about the relation of structures in the world to physical constraints. For instance, that drawn-in line segment might indicated a hard boundary on a solid object, as opposed to, say, a concavity. A myriad of other conclusions—say, about the center of gravity of the object, or its behavior when a stream of water strikes it—diverge depending on this categorical distinction. We use the presence (or absence) of various line segments as a key to a choice of a particular model class or "context" in the sense of Jepson & Richards (1991); the resultant categories are sensitively dependent on this choice.

4.2. Psychological typicality

As noted in the previous section, giving random examples of each category model is less than satisfactory, as it tends to lead to badly garbled and uncommunicative shapes. The problem, in essence, is that random orientations (say) are picked *without bias* around a circle that is actually a veritable mine-field of bias—that is, many entire regions, many classes of orientations, fall too close to a special value to appear psychologically generic. Thus many supposedly random angles end up appearing nearly right, nearly parallel, nearly straight, and so forth, with the effect of introducing upredictable pockets of unintended structure—in effect false target modal genericities. Even more seriously, random values end up getting selected willy-nilly on either sides of special values, producing effects such as the unpredictably crossing line segments in some of the examples in Fig. 4-7 below.

The result is that some objects that do not appear even to be examples of categories that they are supposed to epitomize. For example, when one imagines a generic V, one brings to mind a generic-looking angle. A random V constructed without bias, however, is somewhat likely to appear approximately straight (180°); just how likely depends on the size of the "sphere of influence" surrounding the 180° special value. This in turn is something that must be determined empirically (though one may bring some constraints to bear, about which see below).

In this sense the psychological notion of a random differs sharply from the normative definition. By psychologically random, we mean something more like "apparently typical"—which, as we have seen, is a rather complex property, not at all coincidental, but rather requiring an almost studied avoidance of lower-dimensional structure. Genuinely typical examples, the argument goes, avoid appearing non-generic by virtue of having arisen genuinely out of the generative model in question. Actually random examples, by contrast, will often appear non-generic unless carefully steered clear of non-genericities. This beings to approach the slippery but critical notion of the "mean shape[1]" of an object class (though in the light of the

[1] I thank Stan Dunn for suggesting this term.

underlying motivation of the current theory, we might prefer the term "modal shape").

Psychological typicality within a model, then, stands in sharp contrast to any normative definition of "random" selection. The properties of statistically random examples of a geometric class is subject of the statistical theory of shape (for a synopsis of which, see Kendall, 1989). Discovering these properties for even very simple shape-classes turns out to be a very difficult mathematical question (see for example Kendall & Kendall, 1980, and Small, 1982, on triangles) on which in fact progress was not made until quite recently. Intriguingly, modal points in distributions of such properties are can be peculiarly at odds with psychologically salient values: for example, the distribution of the largest angle in a random triangle has a mode (most frequent point) slightly *under* 90°.

Hence if we wish to construct typical-*looking* members of a category model automatically, we are well-advised not to choose points randomly from the appropriate manifold, but rather pick from a distribution biased away from every modal non-genericity, i.e. each special value along each parameter. One natural way of accomplishing this is to perturb a parameter a "substantial" distance away from some special value, where what we mean by "substantial" is, broadly, large enough to make the generic value noticeably distinct from the special value, but not so large as to bring the value close to the another special value along the same parameter.

Secondly, since we can perturb in either direction along each 1-D parameter, we must give a sign to each perturbance. As we have discussed, the two generic cases falling in opposite directions from the same special value will form psychologically distinct categories (though of course related ones, typically as paired opposites), resulting in a segregation at each special value between two distinct groups of points, each of which is generic in a different way. Hence the need to steer clear of non-generic values has the effect of "balkanizing" the configuration space into distinct modal regions, within which structure is relatively stereotyped and restricted, and between which contrasts are relatively striking.

4.2.1. Automatic construction of typical cases

To make this more concrete, we return to the V's lattice (Fig. 4-1), and use as an example referencing an "isosceles V" to its lower neighbor the "right

V." We can perturb this special value on one side to produce a tightly constrained collection of "acute isosceles Vs" (Fig. 4-4, left), and similarly, on the other side to produce "obtuse V's."

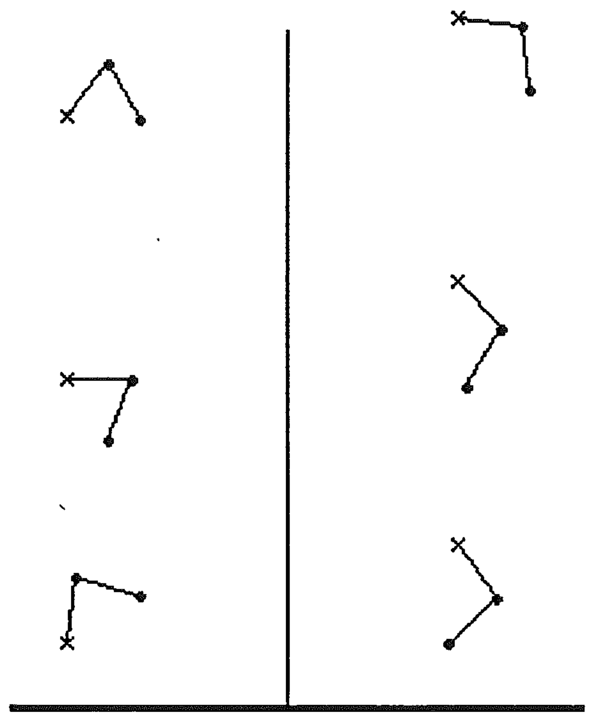


Fig. 4-4. Generically acute (left) vs. generically obtuse (right) isosceles V's. A proto-Bongard problem?

Now, both categories are readily identifiable for what they are; generic examples of a particular category. Here, then, we have the makings of a automatically constructed but nevertheless solvable Bongard problem.

Notice, critically, that these modes of typicality (i.e., distinct ways of being typical in a given space), rather than deriving from lack of constraint or structure, are tightly coupled to a particular choice of special values—i.e. those values that typical examples carefully avoid.

As an immediate consequence of this scheme for isolating inferential genericity, one cannot speak of an *absolutely* typical class of shapes in a given model (with the lone exception of an origin node), or, indeed, of a *single* class of typical examples; rather, typicality is always defined with respect to a chosen concept set, and the modes of typicality typically come in (at least) pairs. Hence it is not possible to redraw the lattices above replacing "random" examples with "generic" examples, because for each node there is a multiplicity of generic cases: two (one on each side) for *each* special-valued lower neighbor of the node. In fact, in order to make each structural parameter of a given node really typical, it is necessary to reference it all the way down to its own "origin," so that each of its parameters has independently been referenced down to some special value (the set of which special values constitute the chosen origin). A typical V, for example, might be referenced all the way down to (i.e. constructed as a perturbation of) the isosceles right case.

When we reference generic V's (not isosceles this time) all the way down their lattice to right isosceles V's. The result, naturally, is 4 quadrants of V's, all of which be readily placed by eye into their correct respective sides of the special values in question (Fig. 4-5).

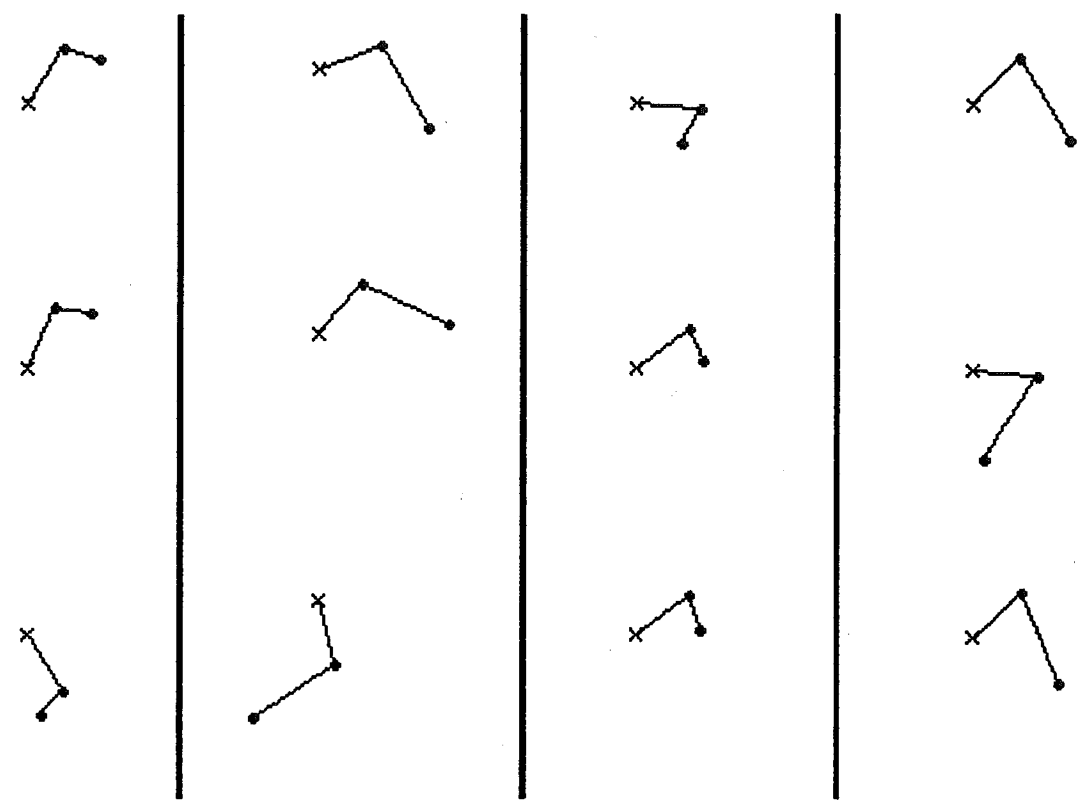


Fig. 4-5. Four putatively distinguishable categories of V's

Fig. 4-5 is reminiscent of Bongard's induction problems: the examples in each column bear clear, psychologically transparent contrast with those in each other column. The idea is an intriguing one, since the procedure for generating the examples is entirely automatic. Bongard's own "Bongard problems" have solutions that a human observer can immediately grasp as intuitively right, partly because Bongard made them up that way—using his own categorical intuition, he selected what he believed to be well-formed categories, putting structure into the problem by fiat. In that sense, they are solvable due to a particular kind of concept match between Bongard and us his readers; we share with him some of the same intuitions about what properties and parameters of the pictures are intuitively liable to be involved in a solution.

In much the same way, the capacity of the Prolog-produced lattices to produce solvable synthetic Bongard problems stands as prima facie evidence that here the theory is composing categories in about the right way using the about right concepts and the right parameters. Of course, this set of problems the program would produce is much more *boring* that Bongard's; while he selected what he thought were intriguing puzzles, the synthetic ones would represent a putatively exhaustive enumeration of the "Bongard potential" of some very small, closed model class made up of very simple figures.

Returning to the automatic generation of typical cases, while we cannot redraw each of the lattices above with "typical" cases replacing "random" cases—simply because there are as many generic cases for each node as the node has origins (collapsing over all the quadrants, e.g. taking the "first" (i.e., fully positive) quadrant in each case. Consequently, though, what we *can* do to create collections of genuinely typical-looking examples is to redraw the entire lattice once for each origin present. Then all the nodes that are connected to that origin can be exemplified by an example that is generic in the "first" quadrant with respect to that origin. (Of course the signs are all arbitrary anyway, so the "first" may be replaced by any quadrant.)

The V's lattice now splits into two origin-tagged "typical collections" (Fig. 4-6).

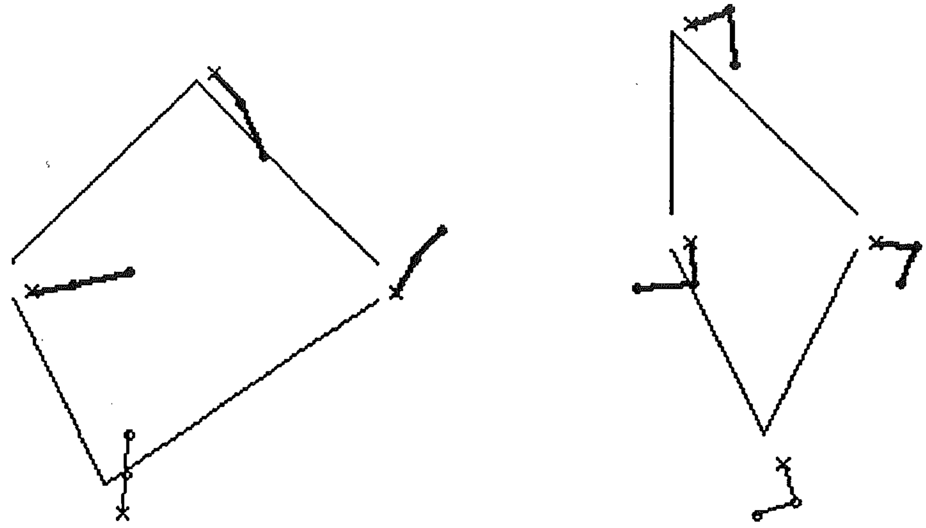


Fig. 4-6. Collections of "typical" V's. On the left, examples are generic with respect to the straight isosceles V; on the right, with respect to the right isosceles V.

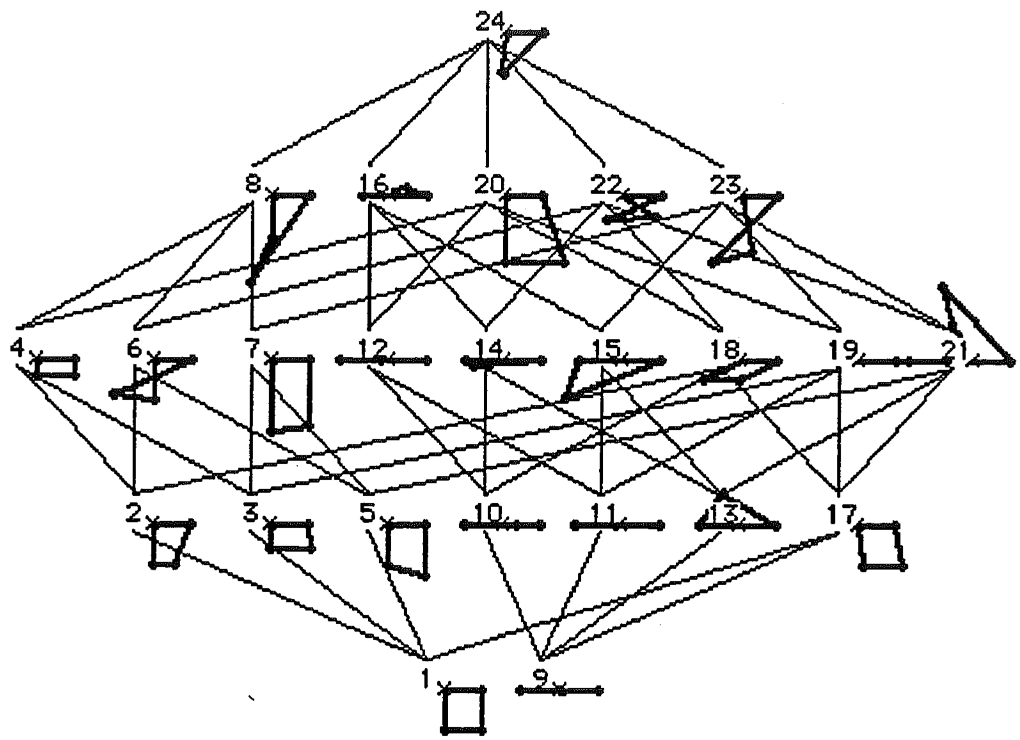
In the left collection, using the "straight isosceles V" origin, we see V's that are slightly bent from straight, or straight but with the middle dot slightly closer to one side, or both (top node). In the right collection, we see a completely different set of V's: slightly acute, slightly L-shaped, or both (top node). All 8 nodes in the two collections seem like psychologically typical examples of some kind of V, and each of a different kind of V. Again, here we see the makings of automatically generated but nevertheless solvable Bongard problems, in which examples of these psychologically transparent categories are pitted against each other.

4.3. Imposing constraint on object models

We now investigate the imposition of constraint, via the constraint functors introduced in Ch. 3, on the structural object models defined in this chapter. The resulting lattices have somewhat more complex form, due to geometric interactions among the structural parameters. For instance, some parameters tend to produce geometric singularities of various kinds when they are brought together, and so forth. We will simply explore this complicated issue using three particular interesting examples.

4.3.1. Pruning a large lattice

Now having set our concepts in place with V's and other 3-dot objects, we are in a position to follow through with 4-dots, and see if the same concepts continue to lead to interesting trees full of recognizable categories. Without further ado, then, the full "4-gon" lattice (Fig. 4-7)




```
?- Object = [p1,p2,p3,p4],Connections = [(p1,p2),(p1,p3),(p3,p4),(p2,p4)],
    categories(Object,Connections,Tree,Object_list).
    Object  = [p1,p2,p3,p4],
    Connections  = [ (p1, p2), (p1, p3), (p3, p4), (p2, p4)],
    Tree = [[p1,p2],[[p1,p3],[p3,p4]]],
    Object_list = [
          (1,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), 1)]),
          (2,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), gen)]),
          (3,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
              (d(p1,p2) / d(p3,p4), 1)]),
          (4,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
              (d(p1,p2) / d(p3,p4), gen)]),
          (5,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), 1)]),
          (6,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), gen)]),
          (7,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen),
              (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
          (8,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen),
```

```
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(9,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), 1)]),
(10,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), gen)]),
(11,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), 1)]),
(12,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(13,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), 1)]),
(14,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), gen)]),
(15,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(16,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(17,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), 1)]),
(18,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), gen)]),
(19,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(20,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(21,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), 1)]),
(22,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), gen)]),
(23,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(24,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen),
```

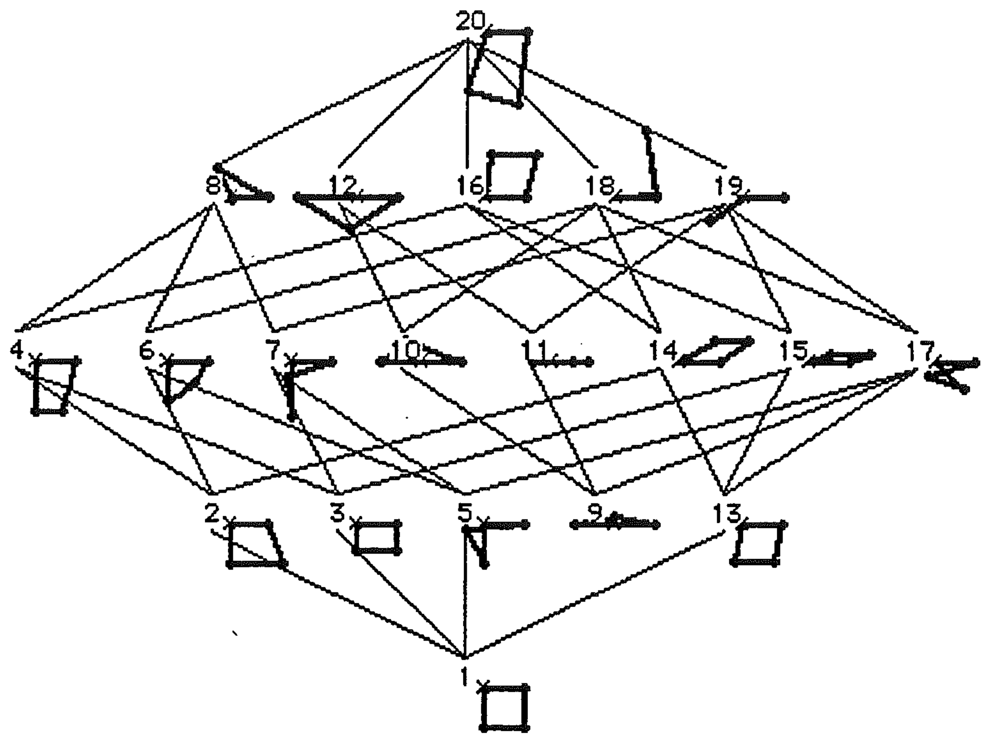
```
(d(p1,p2) / d(p1,p3), gen),
(d(p1,p2) / d(p3,p4), gen)])]
```

Fig. 4-7. The full 4-gons lattice. (Note that the sample objects drawn at some of the nodes are difficult to comprehend visually, because they have been selected randomly rather than "generically" from the model at that node: see text for a full explanation, or Fig. 4-9 for a more visually comprehensible depiction of the models.)

This lattice is quite difficult to interpret by eye; for the codimension-1 cases, for instance, it is hard to tell by looking at each example what special value concept is being invoked. Interestingly, it turns out that this is not a bug in the program, but actually a genuine result of how we have defined "random" examples of each category. Genuinely "typical" examples would seem much more representative, and would allow us to deduce each special value easily. Such typicality, is perfectly possible to define within the formal scheme, and will be taken up in the next section.

Clearly recognizable on this lattice is the *square* as a codimension 4 origin. Other examples, especially in the heavily-constrained codimension-3 case, appear to be highly structured. But something is clearly amiss with all the descendents of node 12. All of them, including the origin node 9, are degenerate—they have collapsed to a line segment. This is because of the unfortunate intersection of nodes 16 and 20, each of which has a segment (`p1p3` and `p3p4`, respectively) parallel to the base segment. Geometry dictates that three segments in a 4-gon cannot be parallel without the object collapsing, but Prolog has no knowledge of such fine points. So the intersection of these two nodes (node 12) as well as all its descendents, are degenerate.

We can remedy this situation by calling on the `precludes` functor defined in Ch. 3. We can redraw this lattice, that is, under the constraint `precludes(16,20)`, where special values (which are after all isomorphic to codimension 1 nodes) are being referenced by their second row representatives. The result, we would expect from the meaning of preclusion, is to prune the lattice of the offending sublattice (Fig. 4-8).




```
Object = [p1,p2,p3,p4],Connections = [(p1,p2),(p1,p3),(p3,p4),(p2,p4)],
    categories(Object,Categories,Tree,Object_list).
    Object = [p1,p2,p3,p4],
    Categories = [ (p1, p2), (p1, p3), (p3, p4), (p2, p4)],
    Tree = [[p1,p2],[[p1,p3],[p3,p4]]],
    Object_list =
    [ (1,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), 1)]),
          (2,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), gen)]),
          (3,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
              (d(p1,p2) / d(p3,p4), 1)]),
          (4,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
              (d(p1,p2) / d(p3,p4), gen)]),
          (5,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), 1)]),
          (6,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
              (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), gen)]),
          (7,
            [ (a(p1,p2) - a(p1,p3), 1.570795),
```

```
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(8,
  [ (a(p1,p2) - a(p1,p3), 1.570795),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(9,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), 1)]),
(10,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), gen)]),
(11,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(12,
  [ (a(p1,p2) - a(p1,p3), 3.141591),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(13,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), 1)]),
(14,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), gen)]),
(15,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(16,
  [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)]),
(17,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), 1)]),
(18,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen), (d(p1,p2) / d(p1,p3), 1),
    (d(p1,p2) / d(p3,p4), gen)]),
(19,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
(20,
  [ (a(p1,p2) - a(p1,p3), gen),
    (a(p1,p2) - a(p3,p4), gen),
    (d(p1,p2) / d(p1,p3), gen),
    (d(p1,p2) / d(p3,p4), gen)])]
```

Fig. 4.8. A constrained version of the 4-gons lattice. (Note that the sample objects drawn at some of the nodes are difficult to comprehend visually, because they have been selected randomly rather than "generically" from the model at that node: see text for a full explanation, or Fig. 4-9 for a more visually comprehensible depiction of the models.)

Now we have a lattice full of non-degenerate 4-gons as desired. Notice that the second origin, node 9, still remains an origin, though it is no longer on the bottom row.

In order to inspect the objects on this lattice more clearly, we can redraw the lattice substituting psychologically typical cases, as defined above, for the random examples in the figures above. Using the version with only one zero-DOF origin, we arrive at the unique enumeration of the possible categories of 4-gon, each category now exemplified in a psychologically typical and hence categorically transparent manner (Fig. 4-9).

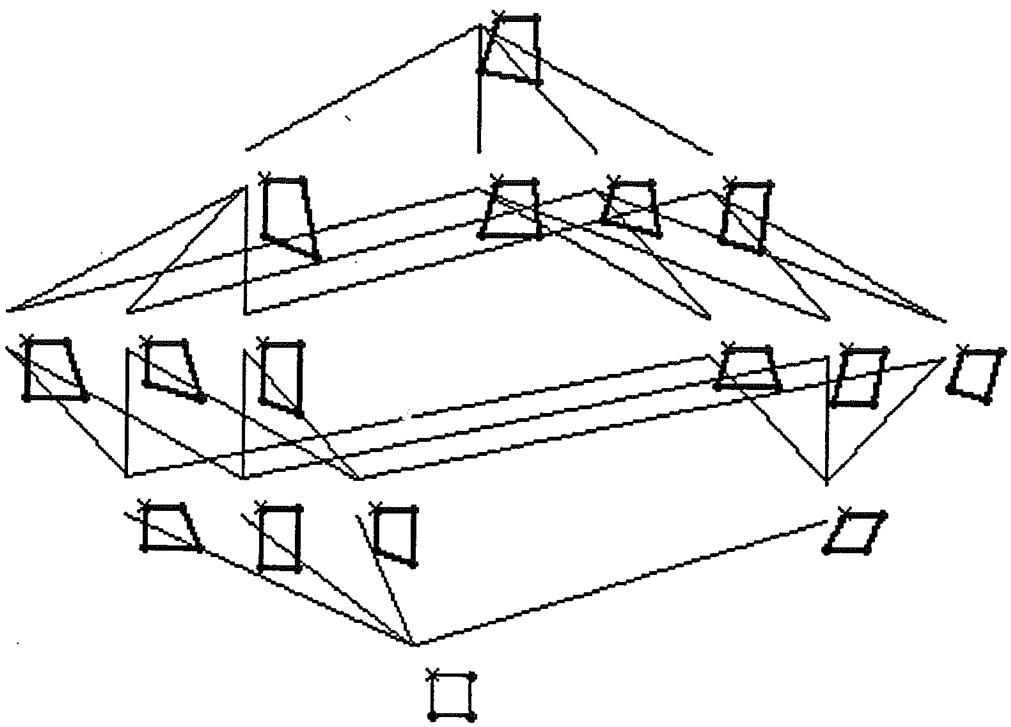


Fig. 4-9. A collection of "typical" 4-gons.

The top node appears simply to be a deformation of the square, exactly as desired. Many other familiar types—rectangle, parallelogram, rhombus (diamond)—make appearances, seeming very much "themselves" (i.e., matching our mental stereotypes), quite unlike their nearly unrecognizable "random" guises from the original lattice. The various unnamed types are clearly recognizable as more or less regular types of object, ones which we as human observer could identify with their brethren in their natural categories with ease. Again, the possibility of synthetic Bongard problems suggests itself.

4.3.2. Context: `always`

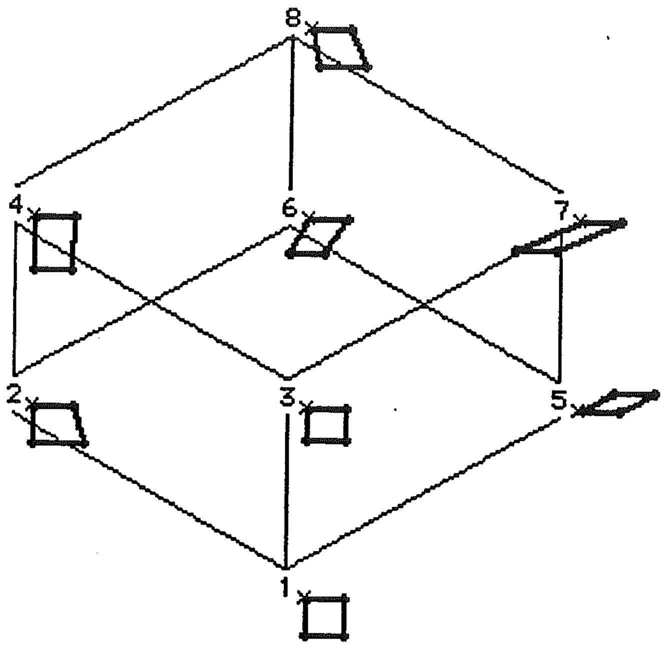
We now consider how one of the constraint functors, `always`, has the effect of constraining model class to some smaller, more constrained "context," i.e. a sub-universe even more constrained than the overall model class. This interpretation of the meaning of the functor follows naturally from the fact, discussed in Ch. 3, that `always(a)` has the effect of restricting a lattice to a particular sublattice, namely the lattice "hanging" off the node `a`. That is, the constrained sublattice contains only those objects that bear the property in question—the set of category models in which property is an ever-present certainty.

When we apply this notion to structural objects such as 4-gons, it has the effect of cutting the lattice down to include only those cases that obey the newly ubiquitous property. We simply illustrate this by means of an example. We start with the lattice of Fig. 4-8, one of the large 4-gons lattices. One of the codim-1 nodes (node number 16), i.e., one of the singleton properties, is `(a(p1,p2) - a(p3,p4), 0)`, meaning that the sides `p1p2` and `p3p4` are parallel (the difference in their angles is zero).

We add the constraint to the constraint list under which the lattice is generated using the command

```
?- constrain(always(16)).
        yes
```
(Eq. 4-7)

giving rise to the much smaller lattice in Fig. 4-10.




```
?- O = [p1,p2,p3,p4],C = [(p1,p2),(p1,p3),(p3,p4),(p2,p4)],
   fix_connections,categories(2,O,C,Tree,L).
   O = [p1,p2,p3,p4],
   C = [ (p1, p2), (p1, p3), (p3, p4), (p2, p4)],
   Tree = [[p1,p2],[[p1,p3],[p3,p4]]],
   L = [ (1,
          [ (a(p1,p2) - a(p1,p3), 1.570796),
            (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
            (d(p1,p2) / d(p3,p4), 1)]),
        (2,
          [ (a(p1,p2) - a(p1,p3), 1.570796),
            (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
            (d(p1,p2) / d(p3,p4), gen)]),
        (3,
          [ (a(p1,p2) - a(p1,p3), 1.570796),
            (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
            (d(p1,p2) / d(p3,p4), 1)]),
        (4,
          [ (a(p1,p2) - a(p1,p3), 1.570796),
            (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), gen),
            (d(p1,p2) / d(p3,p4), gen)]),
        (5,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), 1)]),
        (6,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), gen)]),
        (7,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
        (8,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), gen),
            (d(p1,p2) / d(p3,p4), gen)])]
```

Fig. 4-10. A smaller, more constrained version of the 4-gons lattice, in which the upper side is constrained to be always parallel to the lower side.

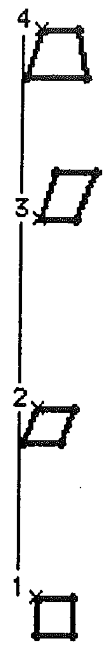
Notice that even though we added only one constraint, the lattice shrunk considerably (from 20 cases to 8). Partly this is due to internal constraints among the properties; partly it is due to the propagation of this constraint through the new sub-universe. This is in effect the "discrete natural modes" idea in action: our interpretations of the causal relationships in force require that we model inferred constraints as propagating *consistently* through the entire local universe. While the world may in principle add or subtract cases one at a time, we are free to move up and down our mental lattices by steps of unit codimension.

4.3.3. An `implies` chain as a causal history: a 4-gon conclusion?

In Ch. 3 (see Eq. 3-24 and following text) we considered how a chain of implications can be thought to correspond to a sequence of stages in some fixed historical sequence in the generation of an object. We now observe this phenomenon within the 4-gons class.

First we restrict the lattice using `always` to a context in which the top and bottom segments are always parallel. Now the generic object is a 4-gon with top and bottom segments parallel, but no other distinguishing features. Omitting some details, we next cast all of the codim-1 row models (each one corresponding to a single property imposed on the generic object) into a chain using the `implies` functor. That is, we apply the `implies` functor to pairs of properties so that each one appears once as the first argument and once as the second argument, except for initial and terminal properties which appear only once as first and second argument respectively (i.e. in the form `implies(a,b),implies(b,c), implies(c,d),` etc. continuing until the entire codim-1 row has been mentioned).

A lattice resulting from an implies chain appears in Fig. 4-11. Note how the lattice mimics Leyton's (1984) conception of the generative history of the parallelogram (Fig. 2-7). A generative sequence analogous to the one he proposes is turned on its side, with a fixed codimensional sequence in effect substituted for a fixed temporal sequence.




```
?- O = [p1,p2,p3,p4],C = [(p1,p2),(p1,p3),(p3,p4),(p2,p4)],
   fix_connections,categories(2,O,C,Tree,L).
   O = [p1,p2,p3,p4],
   C = [ (p1, p2), (p1, p3), (p3, p4), (p2, p4)],
   Tree = [[p1,p2],[[p1,p3],[p3,p4]]],
   L = [ (1,
          [ (a(p1,p2) - a(p1,p3), 1.570796),
            (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
            (d(p1,p2) / d(p3,p4), 1)]),
        (2,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), 1)]),
        (3,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
        (4,
          [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
            (d(p1,p2) / d(p1,p3), gen),
                                  (d(p1,p2) / d(p3,p4), gen)])]
```

Fig. 4-11. An `implies` chain in the 4-gons world: a model of a fixed generative sequence?

Note how movement up the lattice can be seen in a natural way as the development of a generic object (within the context) in fixed stages from a square: first the square slants, then it stretches, and then finally the orientation of the left and right sides divert from each other. Each category, then, amounts to capturing a specimen of this species at a point prior to the discrete, stepwise initiation of the next stage. Finally note that since here we are modeling categories built from relatively arbitrary generative operations, rather than a fixed stereotyped descriptive sequence of a static shape as in Leyton's model, we are free to use a generative sequence different from that (the so-called Iwasawa decomposition of the affine group) that Leyton argued was universal in shape description. In the example, for instance, a square becomes a rhombus, rather than a rectangle, en route to becoming a parallelogram. In effect, each fixed generative sequence definable within a given concept set picks a different route up the full lattice.


```
?- O = [p1,p2,p3,p4],C = [(p1,p2),(p1,p3),(p3,p4),(p2,p4)],
    fix_connections,categories(2,O,C,Tree,L).
    O = [p1,p2,p3,p4],
    C = [ (p1, p2), (p1, p3), (p3, p4), (p2, p4)],
    Tree = [[p1,p2],[[p1,p3],[p3,p4]]],
    L = [ (1,
            [ (a(p1,p2) - a(p1,p3), 1.570796),
              (a(p1,p2) - a(p3,p4), 0), (d(p1,p2) / d(p1,p3), 1),
              (d(p1,p2) / d(p3,p4), 1)]),
          (2,
            [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
              (d(p1,p2) / d(p1,p3), 1), (d(p1,p2) / d(p3,p4), 1)]),
          (3,
            [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
              (d(p1,p2) / d(p1,p3), gen), (d(p1,p2) / d(p3,p4), 1)]),
          (4,
            [ (a(p1,p2) - a(p1,p3), gen), (a(p1,p2) - a(p3,p4), 0),
              (d(p1,p2) / d(p1,p3), gen),
                                     (d(p1,p2) / d(p3,p4), gen)])]
```

Fig. 4-11. An `implies` chain in the 4-gons world: a model of a fixed generative sequence?

Note how movement up the lattice can be seen in a natural way as the development of a generic object (within the context) in fixed stages from a square: first the square slants, then it stretches, and then finally the orientation of the left and right sides divert from each other. Each category, then, amounts to capturing a specimen of this species at a point prior to the discrete, stepwise initiation of the next stage. Finally note that since here we are modeling categories built from relatively arbitrary generative operations, rather than a fixed stereotyped descriptive sequence of a static shape as in Leyton's model, we are free to use a generative sequence different from that (the so-called Iwasawa decomposition of the affine group) that Leyton argued was universal in shape description. In the example, for instance, a square becomes a rhombus, rather than a rectangle, en route to becoming a parallelogram. In effect, each fixed generative sequence definable within a given concept set picks a different route up the full lattice.

- Chapter 5 -

Pychological Typicality and Modality: Experiments

5.0. Chapter preview

This chapter develops a simple category inference theory that allows the category lattices developed in Ch. 4 to be used as hypothesis generators in the induction of categories from examples. This scheme suggests that human observers' beliefs about the a priori probabilities of objects in the world should exhibit marked peaks—distinguishable modes—at the location of each distinct category model in the lattice. A series of experiments is reported in which subjects were asked to induce a category from a very small number of examples (1 or 3), and generate novel examples of the induced category; in each case, the frequency distribution of the examples they produce takes the predicted form, with a clear frequency mode at each modal category.

Sec. 5.1 develops further the notion of "psychological typicality," really no more than a probabilistic version of the genericity concept. In Ch. 4 it was argued that psychologically typical examples of a particular generative model, rather than being chosen *randomly* from a configuration space without bias, must be chosen in a very biased way from internal regions of the generic space in order to minimize confusion with lower dimensional models, i.e. non-generic cases. Now this idea is worked out in a simple probabilistic context, imposing a probability distribution over objects within a given category model. Briefly, a distribution is constructed in which a probability mode is centered at each lattice model, which then rolls off to near zero at distinct adjacent models. If an observer uses such a distribution to provide priors for a straightforward Bayesian inference, the observer will typically infer the correct category identification of an observed object. Thus the lattice theory is taken to suggest a particular general form for observer's beliefs about the probabilities of various objects within a model, with objects near the center of each model on the lattice being the most probable.

Sec. 5.2 presents a series of experiments designed to provide empirical corroboration of the modal kind of probability space described in the previous section, and more generally of the role of the lattice of modal categories in human category inference. Subjects examined various simple planar geometric objects, and were then asked to produce novel examples in the "same" category. The frequency distributions they produced were then compared with the hypothetical modal probability distributions, and were found to be statistically modal at points corresponding to lattice nodes, according neatly with the theory. Furthermore, predictions are tested for the widths (standard deviations) of these distributions, derived from simple probabalistic argument about how the inference scheme should work, discussed in 5.1. Thus with theoretical predictions of both the means and of the standard deviations of all the distributions available, the theory in effect making a successful *point* prediction about the shapes of subjects' probability distributions.

Sec. 5.3 places the results of these experiments into a broader perspective of categorization research. First, it is argued that the inference scheme sheds light on the relationship between two conceptions of categories that have often been at odds. On the one hand, clusters of objects manifesting graded degrees of typicality in a category show up commonly in human experimental data, including those reported here; on the other hand there are, categories that seem to have essentially "classical" (necessary and sufficient) definitions. Here, the a strict and extremely uniform category conception, the constraint manifold, gives rise to graded distributions of typicality, due only to internal requirements of probabilistically correct inference.

Second, it is argued throughout that these experiments give insight into the subtle relationship between observers' tendencies toward two conceptual poles of categorization strategy. On the one hand, perceivers might summarize observations in a way that simply reflects the statistical distributions observed to obtain in the world (a "bottom up" approach); on the other they might attempt to impose coherence and organization onto the observations ("top down") in accord with some prior goals or criteria. While categorizations are of course necessarily keyed to observations to some extent, the experiments presented here reveal quite clearly the predominance of subjects' *prior* concepts in determining the location of the modes. The modes that appear in these subjects' data, that is, are nowhere to be found outside their heads; in fact subjects' apparent beliefs in some cases run *counter* to their observations, at times creating probability peaks in the head where there are probability valleys in the world.

5.1. Psychological typicality and statistical modality of categories

The previous section argued in that psychologically typical members of a given category model fall at intermediate points between special values, avoiding proximity to special values near which they would appear non-generic. This model can now be sharpened by proposing simply that typical objects fall in in a probability distribution centered at a "most typical point," maximally far from any special value, and tailing off towards zero at special values (Fig. 5-1). For simplicity it is assumed that the distribution is normal (Gaussian) with a hypothetical width to be discussed below. Further assume that distance along the horizontal in this space is linear in the psychological variable in question (e.g. angle). None of these assumptions is really critical, though, in that the focus of interest is really only on the qualitative shape of the distribution; the lattice theory lobbies for a particular form in which there are probability modes at each lattice mode. Moreover, the kind of data reported below may be used to confirm or falsify any of these assumptions, providing direct evidence about the more specific shape of the distribution in various domains. In most cases below, a a collection of simple Gaussian modes fit the data quite satisfactorily, suggesting that these assumptions are not inappropriate.

In effect, the notion of genericity is being extended here from a topological to a statistical conception, in which modal forms are associated with modes in a probability distribution. Each distinct "way of being typical" within a given category model (as discussed above) results in a distinct peak in the distribution.

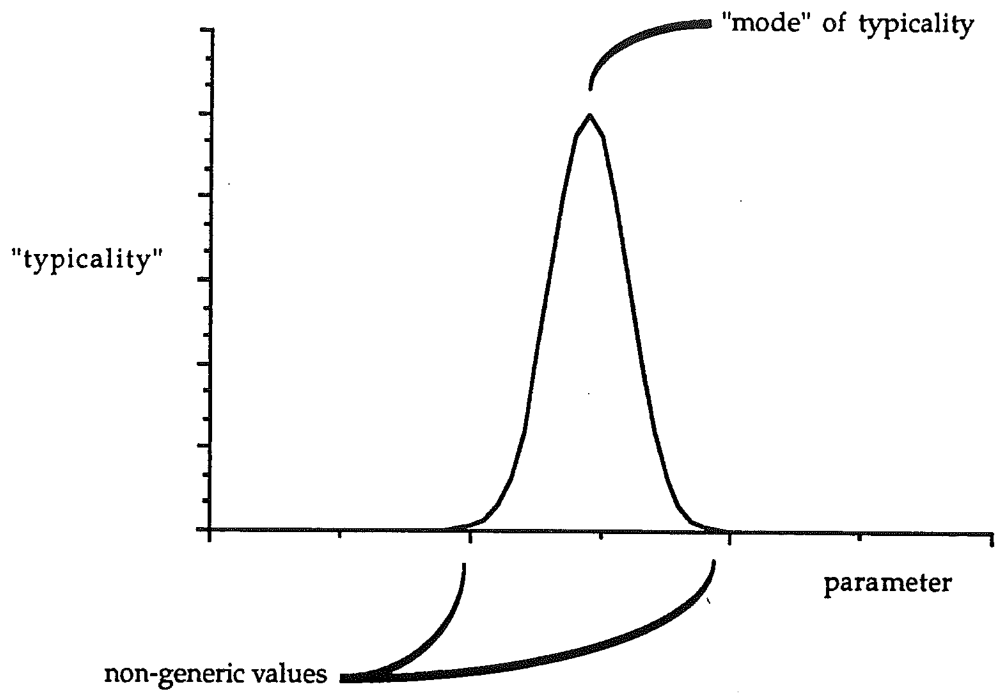


Fig. 5-1. A "mode of typicality" within a category model, in which typicality is maximal at a point maximally distant from non-generic values, and trails off gradually to near zero at non-generic values.

The probability distribution in the configuration space can be thought of as playing a role in category inference, in the following manner. Ch. 2 proposed that modal categories in a lattice serve the perceiver as a source of discrete hypotheses. In a very straightforward way, placing a distribution over these hypotheses allows these hypotheses to be compared in an orthodox Bayesian manner.[1] That is, say a particular modal point m_i on a parameter x has probability distribution $p_i(x)$, and each other mode m_j similarly has distributions $p_j(x)$ Further, say that each *model* m_i occurs with probability $p(m_i)$. The term $p_i(x)$ is treated as the prior probability of an object with parameter value x in the category model m_i: the hypothetical *frequency* distribution of these objects as they actually tend to occur in the world in the

[1]Notice that we hereby sidestep a notorious dilemma of Bayesian reasoning (and a major source of skepticism about Orthodox Bayesianism), namely the problem of ascertaining prior probabilities for an infinite range of unknown and uninteresting events. Here the category lattice explicitly delimits the hypotheses considered to just the modal ones, leaving the entire universe of hypotheses of unknown probability entirely out of the realm of consideration.

i-th category model of this model class. Then, the probability of each category hypothesis m_i (i.e., the conclusion that the observed object belongs to the i-th modal category) can be computed simply as the conditional probability of that category model given the x-value of the observed object, $p(m_i \mid x)$. That is, object x can be attributed to category model i with probability

$$p(m_i \mid x) = \frac{p_i(x)p(m_i)}{p(x)}, \qquad \text{(Eq. 5-1)}$$

where the denominator here sums the probability of the object x over all category models, i.e.

$$p(x) = \sum_j p_j(x)p(m_j), \qquad \text{(Eq. 5-2)}$$

which by stipulation includes *only* those listed on the mode lattice.

A critical and commonly raise problem in reasoning like this is that the various $p(m_i)$, the probability for each model i that occurs, are not known, nor is there really any way to know them for all conceivable models. The trick of using the mode lattice as a hypothesis generator, as proposed in Ch. 2, is that only models that appear on the lattice are assigned probabilities; it is stipulated that in the universe under consideration, only this highly constrained set of models are to be considered. Note again, that this is not presented as an objective fact, but rather an "organizing principle;" as with any inference scheme, violations of the constraint will produce false targets (i.e., structural relations generated by models not considered will lead to erroneous categorizations).

Hence, perhaps counter-intuitively, it is necessary to assign a significant prior probability (i.e, well above resolution zero) to each modal category model, including those at different dimensional levels. This point is discussed much more extensively in Jepson & Richards (1992). Otherwise, examples of lower dimensional categories, rather than being interpreted correctly, would be "better" interpreted as unusual examples of a some higher-dimensional category, which have more mass in the distribution at that point. To solve this problem for purposes here, it is assumed simply that all the areas of the various modes are all about equal—an extension of the famous "Principle of Indifference."[2] Then, each $p(m_i)$ is the same regardless of i, say $p(m_i) = \kappa$ for all i. Then Eq. 5-1 can be rewritten as

$$p(m_i \mid x) = \kappa \frac{p_i(x)}{p(x)}, \tag{Eq.5-3}$$

i.e. as a *likelihood*. That is, the most probable category for object x is simply that m_i that produces the most x's, i.e. the one for which $p_i(x)$ is greatest. Hence, with these assumptions it becomes trivially valid to interpret each object as an example of the category of which it appears most typical.

In short, an object is identified with its category by completely ignoring both the relative probabilities of different models (constrained to be nearly equal) and the probabilities of other models *not* on the lattice (stipulated to be "uninteresting"). In practice, heights were fit to these modes in the data below, rather than constraining them to be equal, since really all that is required is that they each be of the same order of magnitude and much larger than the noise floor. In the empirical data reported below, heights of different modes within a model class were in fact quite similar, despite differences in dimension of the underlying category models.

Clearly, the category choice as per Eq. 5-3 will always be correct and clear—the probability of the correct hypothesis near unity—so long as the modes are well separated from one another compared to their widths, so that the likelihood ratio at most points is high and in the correct direction. That is, again assuming normal distributions, the standard deviation of each mode should be such that the distribution rolls of nearly to zero by the next mode. We now propose that this is typically assumed to be so—a mild version of the "kind world principle." Bennett, Hoffman & Prakash (1989) present extensive arguments that meaningful perception is generally impossible without the world being constrained to be extremely "kind", a notion they too flesh out in terms of a bias in a probability space. Here, strictly speaking, nothing about the world per se is being assumed—rather, prior world probabilities are being attached to essentially mental constructs, in order to render the resulting inferences both simple and correct to the limit of the validity of our concept

[2]The notion, dating back to early conceptions of the theory of probability, that outcomes of unknown status should be assigned equal probabilities.

set. That is, the ideas of modality and genericity are molded into a probabilistic guise, with the effect of reorganizing our categorical inferences into a favorable form. (See remarks in Ch. 2 on the ontic status of the categories in the lattice.)

The final step in fleshing out the inference scheme, then is to assign hypothetical standard deviations to the distributions centered at each mode, guided by the notion motivated above of achieving high likelihood ratios and hence clear-cut category decisions for for most objects. The next section proposes a simple scheme for doing so. This scheme then serves to generate very specific predictions about the outcomes of experiments that in effect directly examined subjects' probability distributions for a variety of induced categories. Previewing, when subjects are asked to induce a suitable category from a small number of examples, and then produce novel examples of the same category, they do so exactly in accord with the argument above: their distributions have modes at maximally "typical" points of the induced category model, as well as at lower-dimensional non-generic points; and moreover, each such distribution is just wide enough to run approximately to zero by the next mode.

5.1.1. Motivation for experiments

Given the arguments in the previous section, it would be expected that the overall probability distribution in a configuration space should have modes centered in each "typical" region of the space—i.e. the spaces between the modal non-genericities—as well as modes centered at each of the non-genericities. The former simply represent the lower-codimension category hypotheses, and the latter represent the higher-codimension category hypotheses. That there are modes at both codimensional levels simply reflects that the categorizer must decide, in effect, how high a codimension the observed object deserves to be assigned—that is, how structured a category it should be interpreted as.

Higher-codimension modes are expected to have narrower distributions, reflecting the fact that each higher-codimension category represents in effect a point puncture in a larger space. That is, if a point does not fall within the distribution of a special value mode, than by default it pops back to the larger space.

Hence, it is proposed that the the distribution around a modal non-genericity be centered at the non-generic value, and be just wide enough that the probability runs down to zero at points as far as one just-noticeable-difference (JND) away from the non-generic point. Each of the various generic modes in the space, on the other hand, has a distribution centered at a point equidistant from the nearest non-generic points, and just wide enough so that the probability runs down to zero at each of the nearest non-generic points. This way, there is a broad distribution at each generic mode (but narrow enough to give large likelihood ratios), and a very narrow spike at each non-genericity (a "pulse", up to resolution).

Fig. 5-2 gives a canonical example, using the angle parameter θ with $\theta = 0°$ and $\theta = 90°$ as special, non-generic values.

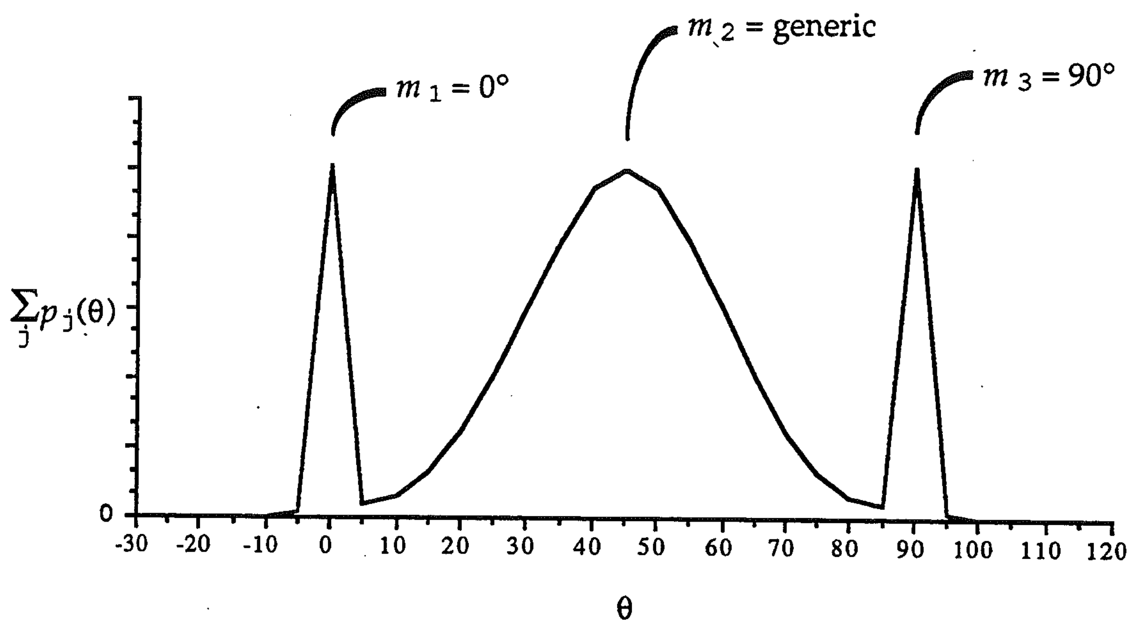


Fig. 5-2. A hypothetical modal distribution with one "generic" mode situated midway between two non-generic modes at θ=0° and θ=90°. The non-generic modes have distributions just wide enough to fall to resolution zero at one JND away from their centers (here, 0.01 times the standard mode height at ±5°). The generic mode falls to resolution zero at the at each of the non-generic modes (i.e., at ±45°). The "concepts" shown here are those of the author, and of course match those used in the computer-generated lattices of Ch. 4.

Concept Uniformity Hypothesis. Strictly speaking, the locations of the modes (special values) along the various parameters cannot be considered a priori predictions. Rather, they form a concept set adopted by the observer, and though one may argue at length about the rationale for picking one set over another in a given context (sensitive to the patterns of regularities in force in that domain), they are in the end merely stipulated internally by the observer.

Hence, critically, in order for the various subjects' data to be coherent, it must be assumed that they will all adopt more or less the same concept set as each other. With respect to choice of special values of parameters, that is, it is assumed that all the subjects are drawn from the same population; this critical linking assumption is called the *Concept Uniformity Hypothesis*. Furthermore, note that both the author and the reader belong to this same population, which given the universally-oriented arguments in Ch. 2, might be argued to include all adult human beings. Hence the concepts chosen by subjects should look reasonable to us as well, with the effect that the locations of the modes in the data reported below will not generally be surprising. The extreme uniformity of concept set across different people reflects the fact that these choices are not totally idiosyncratic, capricious choices, but rather represent clever bets on the part of the categorizing mechanism as to which values of parameters are likely to be causally important in environments at all like the ones to which we are accustomed.

Widths of the modes. Finally, in order to obtain fixed values for the standard deviations of each of the modes, two numbers must be selected: a resolution zero for the probability (since the normal distribution never goes all the way to zero), and a value for the JND, i.e. a resolution zero for the structural variable in question. Both of these choices affect the resulting distribution shape only very coarsely. The probability resolution zero here acts effectively as a significance level (in that it represents the acceptable risk that an alternative or default hypothesis could have produced the observations), and it is set conventionally at 0.01. As for the JND, in the experiments below the measurement precision with which objects produced by subjects were encoded in the analysis was employed as a coarse estimate (1/20 of a line-segment [one unit] for length, 5° for angle). The correspondence between JND and measurement precision is admittedly somewhat coarse, and in several cases below it appears to be slightly off. But again the distribution is not very sensitive to the choice and the resulting predictions fit the data satisfactorily.

5.2. Experiments

This section gives a general overview of the series of experiments, describing the general method, the manipulations carried out in the seven experiments, and the general method by which the data were collected and analyzed. After the general description, method, results, and analysis are presented in detail for each experiment.

Description of general method. In each of the following experiments, subjects were asked to produce category judgments for simple geometric objects in a simple pencil-and-paper task. Subjects viewed a set of sample objects (either 1 or 3 depending on the experiment), and were then asked to produce 6 "different" examples of the same category. The class of objects used were either a line segment with a dot on it (henceforward line-dots for short, and called "blickets" in the experiments) or two segments sharing an endpoint (called V's for short, and "zorks" in the experiments). For example, in Experiment 1, the first page showed an object consisting of a line segment with a dot placed one-quarter of the way along from one end, accompanied by the instruction "Here is an example of a *blicket*." On the second page of each form, the subject is given the instruction, "Please draw 6 different examples of a blicket."

Next, the third page asked the subject to write a short verbal description of the category. Finally, on the fourth page the subject was probed with 6 novel objects and for each one asked whether or not it was a member of the category. The four pages were always presented in the same order, as pilot studies suggested that the latter two questions, which do not provide the principal data, had the effect of causing subjects to "over-regularize" their responses, as they attempted to render their novel examples consistent with the verbal description of the category and their responses to probe questions. The details of the verbal descriptions, as well as responses to probe questions, are not reported, as they tended to be inconsistent with both each other and with the subjects' novel examples.

The phrasing of both the ostension instructions ("Here is an example of a *blicket*" for one sample and "Here are some *blickets*" for 3 samples) as well as the probe instruction ("Please draw 6 different examples of a blicket") were intended to cue the subject to draw a category interpretation rather than a fixed-object interpretation of the verbal label, without explicitly instructing the subject to vary some parameters in novel examples. The mapping between coherent concepts and lexical entries, i.e. the presumption that subjects will tend to identify names for things with categories that they see as reasonable, has been corroborated elsewhere (see Landau & Jackendoff, in press). Subjects did generally drew the category interpretation, understanding that novel examples did not need to resemble the sample object(s) exactly, as for example virtually every subject in the line-dots experiments varied the orientation of the objects freely, evidently attaching no causal importance to the absolute orientation of the sample.

Similarly, as detailed below, there was substantial consensus among subjects as to *which* parameters to vary—position along the line for line-dots and angle for zorks. Subjects who drew an entirely different induction from the sample object produced responds that were, for the purposes of the frequency plots, uncodable (dot off the line for line-dots, unconnected line segments for V's, and in some cases, an amusing variety houses and dragons). Such responses were for the most part quite rare, though; subjects by and large demonstrated an impressively uniform view of what constituted a natural induction from each sample set.

A sample form experimental form appears in its entirety in Appendix B. Each session took less than five minutes. The large majority of subjects were drawn from the university community at large. Some but not all subjects participated in both a line-dots and a V's session; when one subject was tested in both conditions, the order of the two conditions was determined randomly.

Overall plan of experiments. Experiments 1-6 differ only in the type and number of the sample object(s) appearing on the first page. Experiments 1-3 used line-dots (*blickets*), and Experiments 4-6 used V's (*zorks*). Experiment 7 used line-dots much as in Experiment 1, the position of the dot along the line in the sample objects was manipulated between subjects, in an attempt to disentangle the positions of the location of the observed mode from the position of the sample object; this will be explained below.

In Experiments 1 and 4 (*1 line-dot* and *1 V*, respectively) subjects were presented with a single sample object in hypothetically generic configuration. Here the subjects' frequency distributions are expected to reflect the entire lattice of the model class in which the sample object is generic (i.e., to have modes at generic points as well as modal non-genericities in the configuration space).

In Experiments 2 and 5 (*1 midpoint line-dot* and *1 right V*, respectively), the sample object presented was a non-generic case. These two experiments test the prediction that subjects' sensitivity to modal non-genericities would lead them to hypothesize tight categories in which the non-generic value of the parameter is not allowed to vary in novel examples; that is, the hypothesis is that they would induce only the higher-codimension category, not the generic one.

In Experiments 3 and 6 (*3 line-dots* and *3 V's*, respectively), subjects were presented with three putatively generic sample objects rather than just one. This experiment aims to shed light on how subjects' produced frequency distributions interact with the frequency distributions they observe in the world.

In Experiment 7 (*Variable line-dots*), subjects were presented with just one example of a generic line-dot exactly as in Experiment 1, but the position of the dot on the line was varied across subjects, so that the subjects' frequency distribution could be compared with the environmental frequency distribution in a more sensitive way. That is, we attempt to disentangle the location of the maximally generic point in the configuration space from the location of the dot in the sample object. Furthermore, an analysis splitting these subjects into two non-overlapping groups by sample dot position (near-endpoint group vs. near-midpoint groups) provides a sensitive discrimination between the hypothesis that subjects' frequency distributions simply reflect the environmental distribution and the hypothesis that their distributions impose order and structure on the space—namely, the structure dictated by the mode lattice.

Summarizing, the experiments reported are:

Experiment 1: 1 line-dot
Experiment 2; 1 midpoint line-dot
Experiment 3: 3 line-dots
Experiment 4: 1 V
Experiment 5: 1 right V
Experiment 6: 3 V's
Experiment 7: 1 variable line-dot

Collection of data. The principal data collected were measurements of the critical parameters (position of dot along the line for line-dots, angle for V's.)

In the line-dots experiments, the critical measurement was the ratio of the distance from the dot to the nearer endpoint of the line to the length of the line, i.e. the position of the dot along the line as an absolute (unit-free) number. This was ascertained by hand by measuring with an elastic ruled with 11 equidistant marks. The position of the dot was then encoded as a a number from 0 to 10: 0 meaning at the endpoint, and 10 meaning at the midpoint. One unit in this scale is about the same size as the dot.

In the V's experiments, the critical measurement was the angle of the V. This was measured by hand to the nearest 5° using a protractor.

Analysis of data. After measurements for 6 individual objects per subject were taken, the frequency at each measured value was plotted. This typically produced a curve with some visible modes (peaks). This curve was then fitted to a number of summed Gaussian curves, the appropriate number determined by inspection of the raw data. A computer program (Systat) estimated the fit of each model, using a variant of Newton's method to find the best fitting model parameters, using least squared error as a loss function.

Generally, two models were fit to each data set. First, Systat fit a number of summed Gaussians with independent means, standard deviations, and heights. This will be referred to as the "fitted" model. The results of this calculation are reported below as the fitted values of these parameters, i.e. the best estimates of the actual shape of the distributions in the data set.

Next, the same data were fitted with the same number of summed Gaussians, again with independent heights, but this time with fixed predicted values for the means and for the standard deviations of the modes substituted for variables in the model. This will be referred to as the "fixed model," though note that it also has some fitted parameters (the heights) though fewer than are in the "fitted" model. Using the fit of this fixed model, we compute R^2, and compute F-ratio using the total amount of variance unaccounted for by this model as an error term. We use the number of parameters in the fixed model as the degrees of freedom in the numerator, and as the degrees of freedom in the denominator the number of data points (i.e. the number of parameter values for which we collected frequencies, e.g. 11 for line-dots) minus the degrees of freedom in the model minus 1 (see Cohen & Cohen, 1983). Notice that in all but one of the experiments (6), the final fitted model did not include a constant term, since we expected the floor of a probability distribution to be zero. However, we compute R^2 with reference to the total amount of variance, i.e. the total squared error of the constant model. Hence unlike standard polynomial nonlinear models, which include a constant term, the models fitted here in principle can fit the data even *worse* than the constant model—meaning that the F-tests performed here are very conservative.

Note that the fixed model, *not* the fitted model (which necessarily fits the data more closely) appears in the graphs. With the means and standard deviations fixed, the number of degrees of freedom of this model was typically simply the same as the number of modes in the data, i.e. one height fitted per mode. Note finally that with only the heights of the of the various modes fitted, and the total area of the graph in principle fixed by the number of data points collected, the fixed model that appears in the graphs amounts virtually to a *point* prediction of the data collected.

Table 1 below gives the predicted mean, the fitted means, the predicted standard deviation, and the fitted standard deviation for each mode in each experiment. Table 1 gives the fit of the "fixed" model using the predicted means and standard deviations, along with its F-ratio and significance level, for each mode in each experiment.

"Predictions" for standard deviations of modes. The standard deviations of the modes were predicted following a logic motivated by the theoretical arguments presented in Sec. 5.1.1; see above for a more complete discussion. Note that these "predictions" are not in any sense a priori, but for instance depend on the Concept Uniformity Hypothesis, discussed above. In one case, the "predictions" are actually post-hoc. The point rather is that for each experiment the predicted model serves as a *fixed* model of the data, i.e. a model in which constants are substituted for the means and standard deviations of each mode, with Systat estimating only the heights.

For "generic" modes, the standard deviation was selected so that the distribution would fall off to 0.01 of the height at the nearest non-generic value. For non-generic modes, the standard deviation was selected such that the distribution would fall off to 0.01 at plus or minus one JND, here defined operationally as one unit of measurement precision: 1 unit for line-dots (about the size of the dot), and 5° for angles. To achieve the desired roll-off to criterion 0.01, the distance in question (one JND or the distance to the nearest non-generic value) was divided by $\sqrt{-2\ln(.01)}$, or about 3.03. The estimated standard deviation in all cases detailed below corresponds to this quotient.

For example, for mode #2 in Experiment 1 (located at 5 units, or one quarter of the way along the line segment), we predict a roll-off to 0.01 at the two special values of 0 and 10 (endpoint and midpoint respectively), each 5 units away from the center of the mode. Hence the estimated standard deviation of this mode is 5/3.03 = 1.65. The resulting equation for this mode, expressing frequency f as a function of the position x of the dot along the line , was then

$$f(x) = h \exp\left[-\frac{1}{2}\left(\frac{x-5}{1.65}\right)^2\right] \qquad \text{(Eq. 5-4)}$$

where h (height) is the one remaining variable parameter to be fitted by the program.

With angles, the distribution at non-generic points tended to be slightly wider than predicted, as detailed below, suggesting quite plausibly that 5° is an underestimate of the actual size of one JND. The actual numerical predictions for the means and standard deviations are detailed for each experiment below.

Means and standard deviations were compared with their theoretical values using t-tests as a way of cross-validiating the regression technique, though it is a less sensitive test in that we normally predict a null result of no difference. These test are reported in detail, and the small number of cases where the fitted estimate of one of these parameters did differ from the prediction at $p < .05$ are marked in Table 1 , which lists all the predicted and fitted values for each mode in each experiment.

As discussed above, a small number of subjects in each experiment produced objects that did not belong to the intended category at all, indicating that they did not make the intended induction, and which hence were uncodable. The proportions of uncodable response are detailed by experiment below. These subjects were replaced by new subjects in the counterbalancing scheme.

Table 1 below provides the predicted and fitted means, and the predicted and fitted standard deviations, for each mode in each experiment. Table 2 provides the R^2, F, and p for the fixed (predicted) model for each experiment.

In each case, this fixed model is the *sum* of the Gaussians defined by the predicted means and standard deviations listed in Table 1. The seven experiments themselves, as well as the exact reasoning behind each prediction, are detailed one by one below.

Experiment	Mode	Predicted mean	Fitted mean (s.e.)	Predicted standard deviation	Fitted standard deviation (s.e.)
1. (1 line-dot)	1	0 (endpoint)	0.053 (7.469)	0.330	0.149 (61.439)
	2	5	4.868 (0.385)	1.650	1.671 (0.432)
2. (1 midpoint line-dot)	1	10 (midpoint)	9.849* (0.500	0.330	0.662 (0.464)
3. (3 line-dots)	1	5	4.390** (0.273)	1.650	1.422 (0.394)
	2	10 (endpoint)	9.261* (0.366)	0.330	1.341 (0.572)
4. (1 V)	1	55°	50.414°** (1.803)	11.551°	9.643° (2.190)
	2	90°	84.801° (3.556)	1.650°	8.405** (3.201)
	3	120°	120.640° (2.315)	9.900°	8.461° (2.279)
5. (1 right V)	1	90°	90.133° (0.929)	1.650°	9.497°** (1.649)
6. (3 V's)	1	60°	60.925° (32.806)	9.900°	32.11°** (8.703)
	2	90°	94.652° (5.529)	1.650°	27.942°** (2.792)
	3	120°	118.220° (18.305)	9.900°	24.687° (7.525)
7. (1 line-dot, variable)	1	0 (endpoint)	0.320 (0.860)	0.330	0.359 (0.795)
	2	5	5.158 (0.359)	1.650	1.526 (0.311)
	3	10 (midpoint)	13.797 (24.222)	0.330	2.926 (9.175)

Table 1. Comparison of predicted and observed means and standard deviations for each mode in each Experiment. Fitted values marked * and ** differ from the prediction by *t*-test at $p<.05$ and $p<.01$ respectively.

Experiment	N	R^2 (fixed model)	F	p
1. (1 line-dot)	18	0.7454	$F_{2,8} = 11.71$	$p < .005$
2. (1 midpoint line-dot)	16	0.7526	$F_{1,9} = 27.37$	$p < .001$
3. (3 line-dots)	18	0.6090	$F_{2,8} = 6.23$	$p < .025$
4. (1 V)	16	0.5340	$F_{3,33} = 12.65$	$p < .001$
5. (1 right V)	18	0.5090	$F_{1,35} = 17.63$	$p < .001$
6. (3 V's)	24	0.6045	$F_{4,14} = 5.35$	$p < .01$
7. (1 line-dot, variable)	28	0.8472	$F_{3,7} = 12.93$	$p < .005$

Table 2. Fits of final fixed ("predicted") models for each experiment.

5.2.1. Experiment 1: One line-dot

In this experiment, subjects induced a category from a single generic line-dot. We expect their resulting distribution to show modes reflecting the entire lattice of the model class in which this object is generic, as discussed above.

Method. The induction example was a line-segment with a dot placed one-quarter of the way from one end of the segment (i.e. a score of 5 in the coding system). For half the subjects, the line segment was at a shallow oblique angle (15° from horizontal); in the other half a steep one (80° from horizontal). Orthogonally, for half the subjects the dot was nearer to the upper end, and for half it was nearer to the lower end.

Subjects. 18 subjects were tested in this condition.

Results and analysis. The results for Experiment 1 are plotted in Fig. 5-3. The objects from 2 subjects (10%) were uncodable and were discarded.

There are clearly two modes in the frequency distribution: one at the endpoint (0), and one at about the location of the sample dot (5). The prediction is for the mode at 0 to have a standard deviation of 1/3.03 = 0.33, and the mode at 5 to have s.d. = 5/3.03 = 1.65. The regression estimates one mode to be at 0.053 (not different from 0, t_{10} = 0.0709, ns.) with s.d. = 0.149 (not different from 0.33, t_{10} = 0.0029, ns.); and the other mode to be at 4.868 (not different from 5, t_{10} = 0.3429, ns.) with s.d. = 1.671 (not different from 1.65, t_{10} = .0462, ns.). The predicted and estimated values for the means and standard deviations of these two modes, as for all the modes in all the experiments, are tabulated in Table 1.

The fixed (predicted) model accounted for the data very well (R^2 = 0.745, $F_{2,8}$ = 11.71, $p < .005$); it is plotted against the data in Fig. 5-3. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

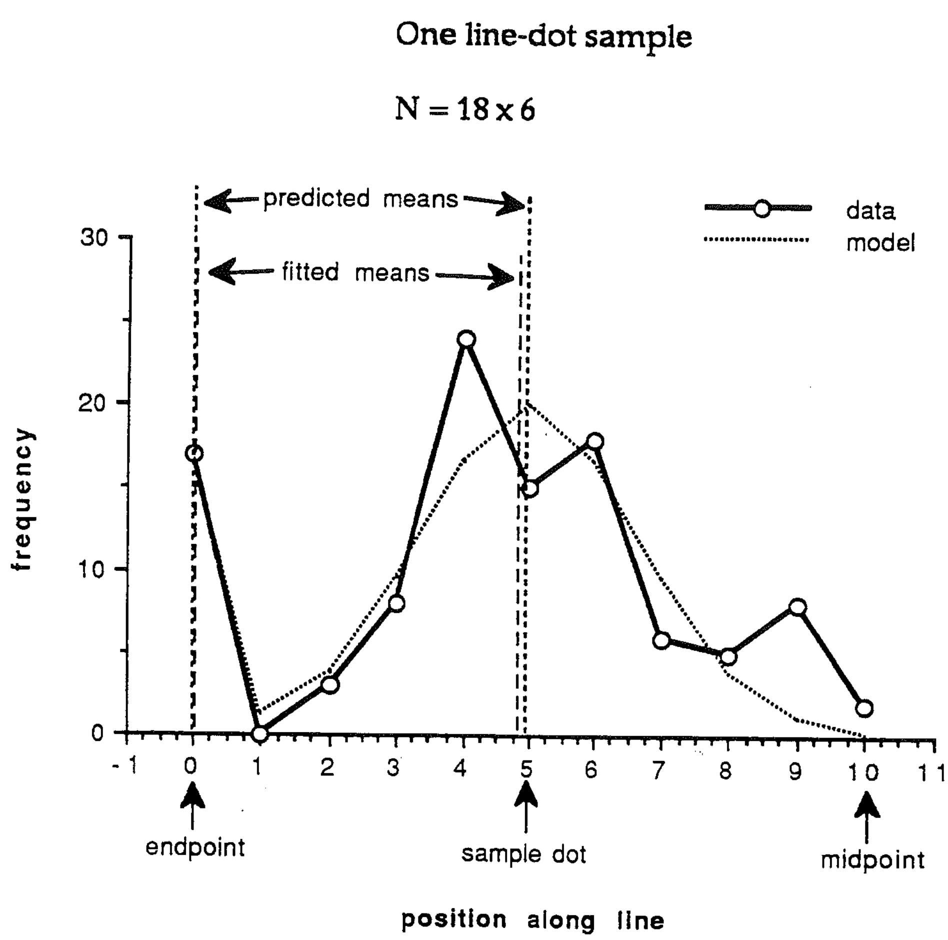


Fig. 5-3. Results for Experiment 1 (one line-dot). Model shown (dotted curve) is the "fixed" model using predicted means and standard deviations. Predicted and fitted means for each mode are also marked.

Discussion. The presence of the mode at the generic position at which the sample dot appeared here is quite intuitive, but the mode at the endpoint is perhaps less so, and constitutes a striking corroboration of the role of the lattice theory in the subjects' category judgments. In the two modes in the distribution we see, as clearly as possible the two modal categories of the simple category lattice for the dot-on-line model class: the codimension 0 case of a dot placed generically on a line, and the codimension-1 case of the dot at the endpoint, of which the codim-0 case may be though of as a perturbation. Even with a minimum of interpretation, the graph makes clear that subjects are generating examples from two separate, distinguishable categories or population processes. Indeed, the fact that the number of distinct distributions in distribution is actually greater than the number of examples in the induction base—the fact that subjects in effect induced *two* category populations from just *one* example—is strong prima facie evidence for the operation of some sort of mechanism enumerating discrete cases for the model class. The locations of the two modes is not so telling in this regard as the *number* of modes, and the fact that the distribution is modal at all.

The location of the "generic mode" at 5 coincides with both the location of the sample dot and with the location in the configuration space maximally far from both non-generic points. Hence it is unclear whether this mode simply reflects the environment or represents a construction subject to internal constraints (e.g. of distinguishability from higher-codimension categories). An attempt was made to decouple these two factors in Experiment 7. Notice, though, that the presence of the mode at the endpoint lobbies for the latter ("top-down") interpretation. After all, if the modes in this distribution simply reflect observed frequencies, then the mode at zero, at which subjects never observed an example, becomes quite difficult to account for.

5.2.2. Experiment 2: One midpoint line-dot

In this experiment, the induction sample was a single line-dot with a dot located exactly at its midpoint. As discussed above, the expectation here is that subjects will be unwilling to vary the position of the dot in novel examples of the category, unlike in Experiment 1, but rather zero in on the higher-codimension category model "dot at the endpoint."

Method. The induction example was a line-segment with a dot placed one-half of the way from one end of the segment (i.e. a score of 10 in the coding system). For half the subjects, the line segment was at a shallow oblique angle (15° from horizontal); in the other half a steep one (80° from horizontal).

Subjects. 18 subjects were tested in this condition.

Results and analysis. The results for Experiment 2 are plotted in Fig. 5-4. The objects from one subject (5.8%) were uncodable and were discarded.

There is clearly only one mode in this distribution, at about the midpoint, the location of the sample dot. The prediction is for this mode to be at 10 with a standard deviation of 1/3.03 = 0.33. The regression estimates the mode to be at 9.849 (not different from 10, t_{10} = 0.302, ns.) with s.d. = 0.662 (not different from 0.33, t_{10} = 0.7155, ns.). The predicted and estimated values for the mean and standard deviation of this mode, as for all the modes in all the experiments, are tabulated in Table 1.

The fixed (predicted) model accounted for the data very well (R^2 = 0.753, $F_{1,9}$ = 27.3748, p < .001) is plotted against the data in Fig. 5-4. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

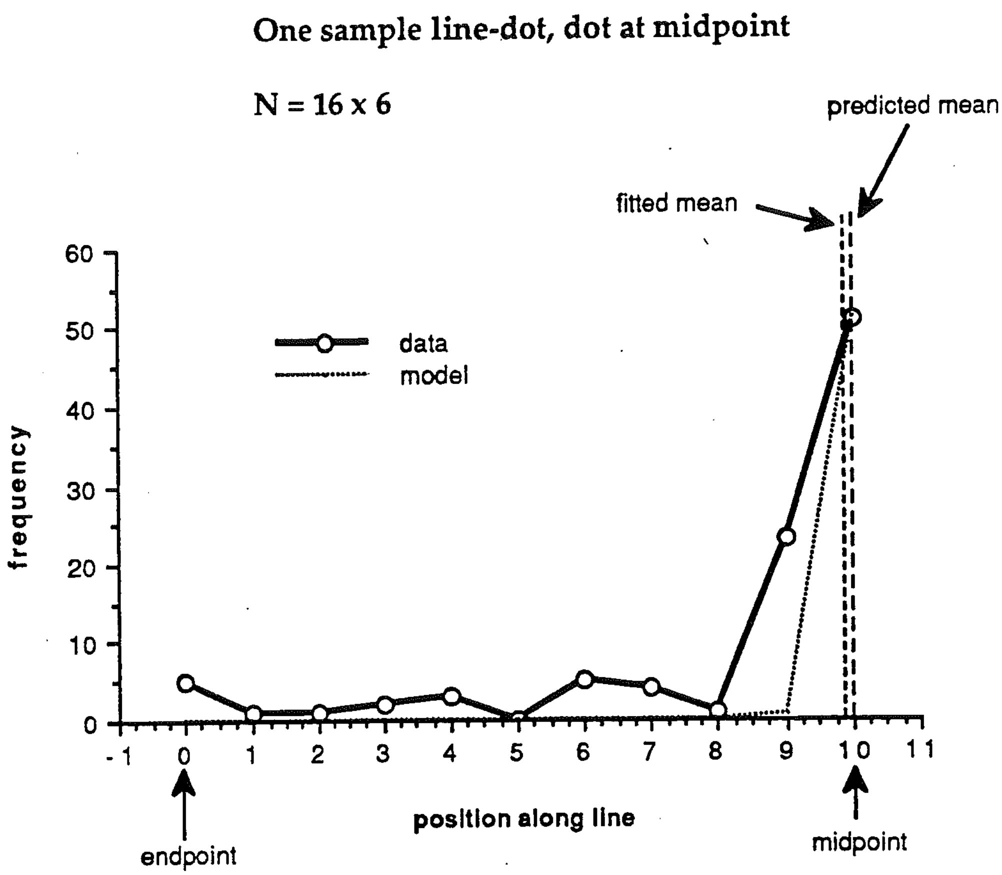


Fig. 5-4. Results for Experiment 2 (one midpoint line-dot). Model shown (dotted curve) is the "fixed" model using predicted mean and standard deviation. Predicted and fitted means of the mode are marked.

Discussion. Clearly, subjects in this experiment saw the dot at the midpoint as a fixed property of the category generating the sample object; they treat the midpoint position in a qualitatively different way from the quarter-point (generic) position of the last experiment in drawing category inferences. In other words, while the generic example invokes an entirely additional category, namely the endpoint case, as the subject constructs a 1-parameter lattice, the midpoint position does not. Putatively, the model class invoked by the midpoint example contains no free parameters, and hence the lattice contains only one node, the "origin".

5.2.3. Experiment 3: Three line-dots

In this experiment, the induction sample consisted of three generic line-dots. This experiment is intended to provide a sharper contrast between the distribution in the environment, which is carefully controlled, and the resulting distributions observed in subjects' output.

Method. The distribution of the positions of the line segments that each subject saw was carefully arranged to meet three constraints. First, it should exclude the special cases of 0 and 10, the idea being to show the subject only generic cases. Second, it should be perfectly flat when pooled across all subjects, so that the total "environment" itself contains no modes. Third, the *means* of each subjects' sample set should also be distributed as flatly as possible. To meet these three constraints, one-third of the subjects saw objects with scores of 1, 2, and 7; one third saw 3,4, and 8; and one-third saw 5,6, and 9. Hence the pooled distribution is flat from 1-9, and the means are evenly distributed at 3.33, 5, and 6.67. Hence any modality in the subjects' output cannot be attributed to modality in the distributions they observed.

The orientation of each object was determined randomly. Each subject saw objects of three different fixed sizes, permuted randomly. Half of the subjects saw two objects with dots closer to the upper half, and one with the dot closer to the lower half; the other half of the subjects had this reversed. This factor was crossed orthogonally with the distribution factor.

Subjects. 18 subjects were tested in this condition.

Results and analysis. The results for Experiment 3 are plotted in Fig. 5-5. No subjects' objects were uncodable, presumably reflecting the fact that 3 samples instead of only 1 greatly reduces the uncertainty about the nature of the intended category.

There are two modes in this distribution, one near the generic point, and one at about the midpoint. The prediction is for one mode to be at about 5, with standard deviation 1.65, and the other mode at 10 with a standard deviation of 1.65. Here the midpoint mode is expected to have the broad distribution as well, since it is not being invoked as a special case, but rather as the central tendency of the lined-dots measured from left to right (see below). The regression estimates one mode to be at 4.390, with s.d. = 0.1.422 (not different from 1.65, t_{10} = 0.5787, ns); and the other mode to be at 9.261, with s.d. = 1.341 (not different from 1.65, t_{10} = 1.7675, ns). Both modes are slightly different from prediction. The first mode is 0.610 away from 5, $t_{10} = 2.340$, $p < .01$. Similarly the mode near the midpoint is 0.739 away from 10, $t_{10} = 2.019$, $p < .05$. Both modes are generally in the right place; in particular the putatively generic one is still in a generic position, though not the position that is maximally distant from both non-generic points. The predicted and estimated values for the means and standard deviations of these two modes, as for all the modes in all the experiments, are tabulated in Table 1.

Generally the differences from prediction in this experiment seem to stem simply from uncertainty about the locations of points in question, due to the lack of a concrete exemplar from which subjects can generate objects. That is, the midpoint mode here does not come from any kind of a midpoint example, since subjects observed none. Similarly, the generic mode is coming solely "out of the heads" of subjects, since the examples were spread evenly throughout the generic portion of the space. In any case, note that despite these apparent slight differences from the predicted model, the predicted model still fits the data to a highly significant degree, a more meaningful test of the prediction.

The fixed (predicted) model accounted significantly for the data ($R^2 = 0.609$, $F_{2,8} = 6.2315$, $p < .025$); it is plotted against the data in Fig. 5-5. The "environmental" distribution, the pooled distribution of objects that subjects observed, is also plotted for comparison. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

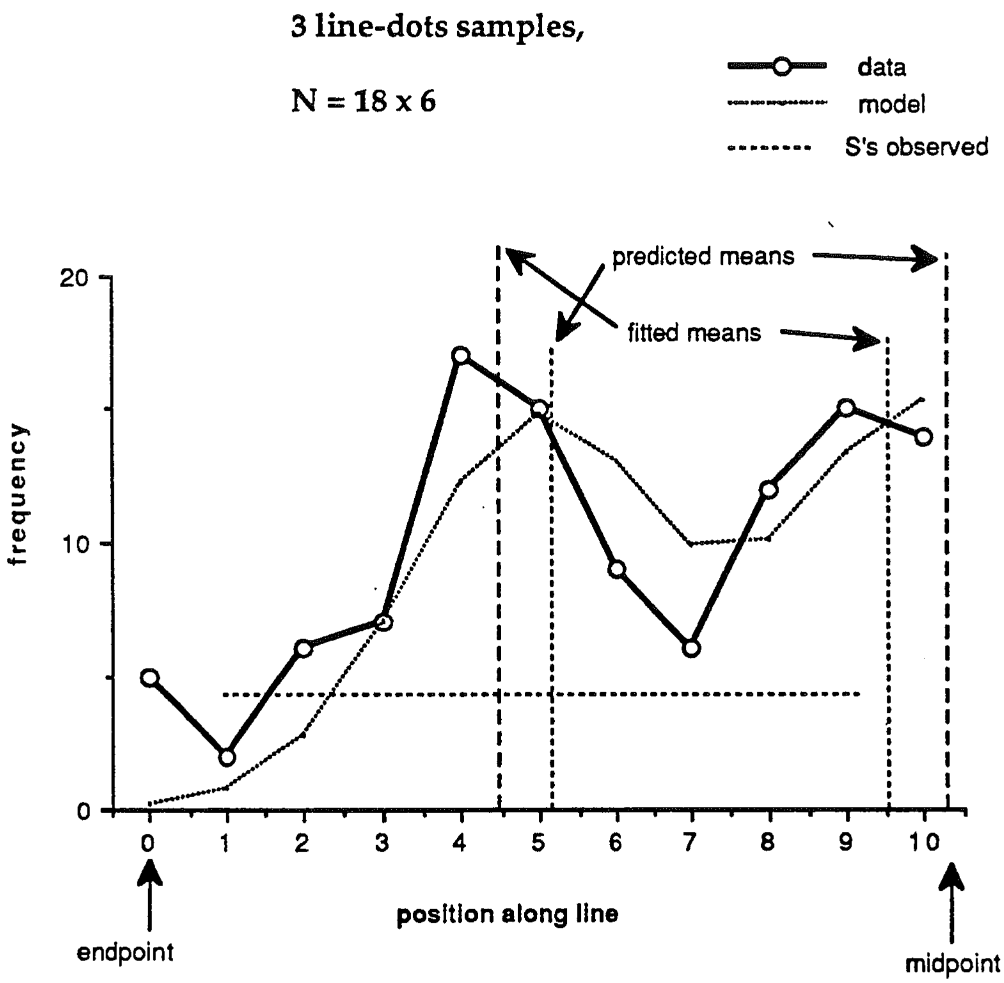


Fig. 5-5. Results for Experiment 3 (three line-dots). Model shown (dotted curve) is the "fixed" model using predicted means and standard deviations. Predicted and fitted means of both modes are marked. The "environmental distribution," the pooled distribution of objects that subjects' observed, is also plotted.

Discussion. All in all, the data from experiment look remarkably like those from Experiment 1: two modes corresponding to two distinct categories, though here we have a midpoint mode where there we had an endpoint mode. Despite the fact that here the environment included a wide, flat distribution of generic objects, the subjects again produced exactly the same relatively narrow, sharp mode more or less at the center of the space of generic objects. The observed distribution, in other words, is not reflected in the induced distribution; rather, it had the effect only of moving the "extra" category invoked from the endpoint to about the midpoint.

To see why the second mode moved in this case, consider how the line-dots can be encoded by subjects. Though due to the symmetry, we are constrained to measure the position of each dot from the nearer end, so that in effect a score larger than one half (10) is not possible, subjects are free to view the objects in a different way. They can for instance adopt a reference coordinate system for each object with a defined left and right, and hence measure the position of the dot all the way from one end to the other. In this case, the midpoint of the segment (so long as it is not perfectly vertical, which it never was) makes a natural center for this coordinate system: a point maximally distant from both ends. Alternatively, if they adopt the system we use and measure only over a reflected half of a segment, then the natural center is at the quarter point of the segment. These two alternative ways of conceptualizing the objects, one treating the object as symmetric and one defining left and right ends—two generative models, in effect—give rise to the two distinct modal central tendencies.

The main point of the experiment, though, is simply that subjects produced a very modal distribution from observing an environment that was not modal, but rather in which the distribution was very flat. Most strikingly, they faithfully reproduced the mode at the midpoint, even though no midpoint line-dots were ever observed; this category comes free with the lattice, generated internally as part of the inference process.

5.2.4. Experiment 4: One V

In this experiment, subjects again induced a category from a single generic object, this time a V. We expect the resulting distribution to resemble that in Experiment 1, though this time there are two non-generic modes, one in the acute region and one in the obtuse region, in addition to the non-generic mode at 90°.

Method. The induction example was always a single V with a dot placed on its vertex (so that the we could probe the 180° case). Half the V's samples had an angle of 55°, and half of 120°, both angles chosen arbitrarily as generic-looking angles. The V's were all positioned at the same orientation, so that the (imaginary) bisector was at 45° from horizontal.

Subjects. 16 subjects were tested in this condition.

Results and analysis. The results for Experiment 4 are plotted in Fig. 5-6. The a relatively large number of subjects here induced a category other than than "V's": objects from 7 subjects (30.4%) were uncodable and were discarded. Most of these detached the two segments from the dot in one way or another, rather than varying the angle.

The distribution has three apparent modes, at nearly the locations expected: the locations of the two sample V's and 90°. The prediction for the acute mode is that the mean should be at 55°, the sample location, and the standard deviation should be (90-55)/3.03 = 11.55°; the non-generic mode at 90° should have a spike-like standard of 5/3.03 = 1.65°; and the obtuse mode should be at 120° and have a standard deviation of (120-90)/3.03 = 9.90°. Regression estimates the acute mode at 50.414°, more acute than predicted (t_{36} = 2.5156, $p < .01$)—perhaps drifting towards the more symmetrical choice of 45°. The standard deviation of the acute mode, on the other hand, did not differ from the prediction (t_{36} = 0.8712, ns). The right-angled mode was estimated at 84.801° (not different from 90, t_{36} = 1.4620, ns). Its s.d., though, was larger than expected at 8.405° (larger than 1.65°, t_{36} = 2.964, $p < .01$.) This clearly represents an underestimate of the size of the JND at this point in the space; subjects in this mode may well have been generating V's from a tight distribution around 90°, but they simply were not precise enough in their drawing to make this mode appear to be a spike. Finally, the obtuse mode appeared almost exactly at the right point at 120.640° (not different from 120°, t_{36} = 0.2764, ns) with standard deviation 8.461 (not different from 9.90, t_{36} = 0.6314, ns). The predicted and estimated values for the means and standard deviations of these three modes, as for all the modes in all the experiments, are tabulated in Table 1.

Despite several the close failures of two predictions, though, the fixed (predicted) model fit the data very well ($R^2 = 0.534$, $F_{3,33} = 12.650$, $p < .001$); it is plotted against the data in Fig. 5-6. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

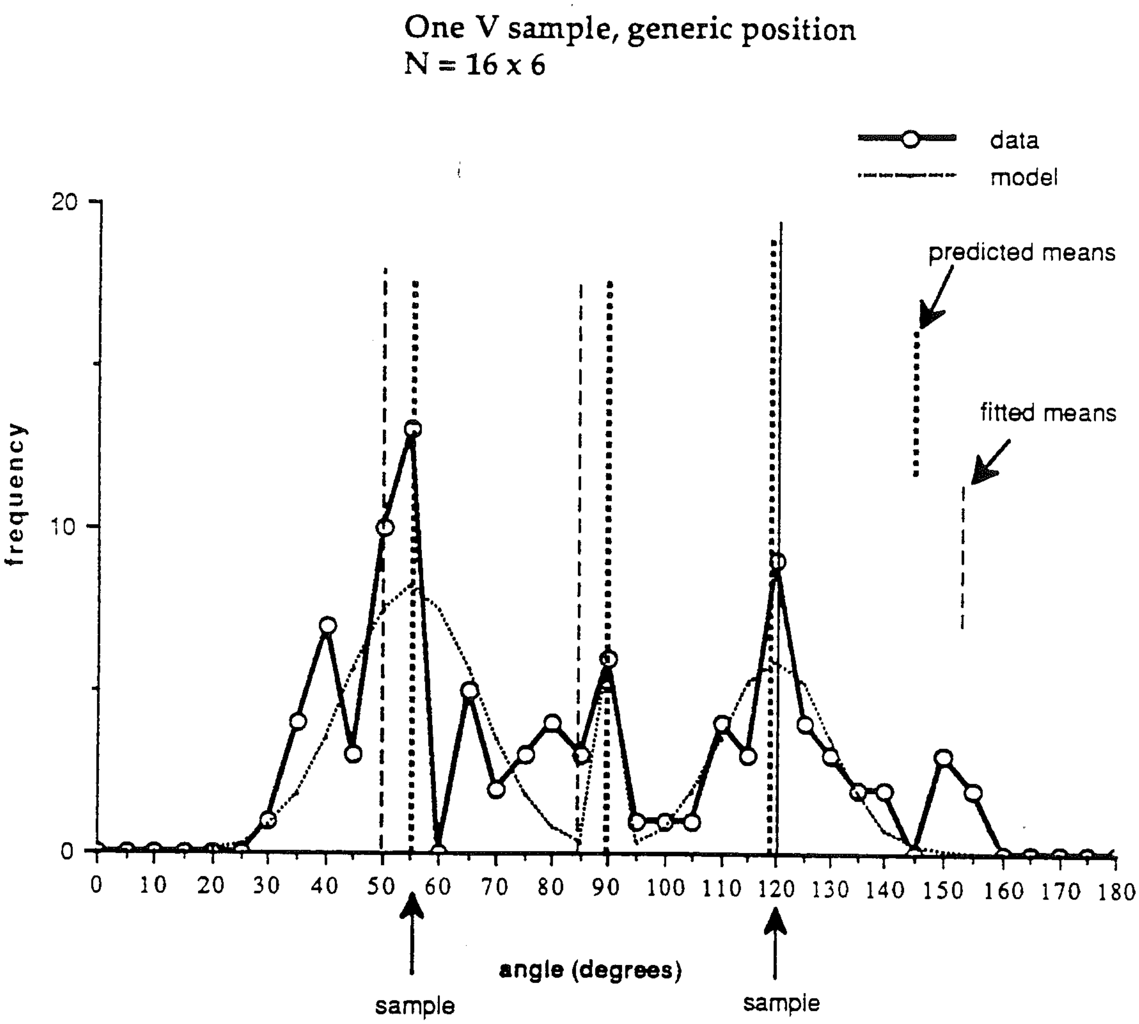


Fig. 5-6. Results for Experiment 4 (one V). Model shown (dotted curve) is the "fixed" model using predicted means and standard deviations. Predicted and fitted means for all modes are also marked.

Discussion. The results of Experiment 4 were, as expected, similar to those of Experiment 1. Again, subjects produced modes not only at the observed positions of the sample objects, but also at the non-generic case picked out by the lattice of the model class, here 90°. The data is certainly noisier than in the line-dots experiments, as may be expected from the relative difficulty of both perceiving and drawing angles as opposed to the position of dots. We might further speculate, though, that the noise in the data reflects that angle is not treated as as causally transparent a variable as position is—that is, a right angle might be a structurally important feature, but actual contact between two observed feature points (e.g. a dot and the end of a line) is even more obviously likely to be structurally important.

5.2.5. Experiment 5: One right V

This experiment is essentially a replication of Experiment 2 (one midpoint line-dot) with V's. Again we expect that presented with a special case, subjects will induce only the higher-codimension category, this time "right V's".

Method. Each sample object was a V just as in experiment 1, except with a right angle. Half of the subjects saw V's whose (imaginary) bisector pointed towards 60° from the horizontal, and half towards 320° from the horizontal.

Subjects. 18 subjects were tested in this condition.

Results and analysis. The results for Experiment 5 are plotted in Fig. 5-7. One subjects' objects (5.3%) were uncodable and were discarded.

As expected, there is only one mode in this distribution, at very nearly 90°. The predicted mean of this mode is of course 90°, and the standard deviation 5/3.03 = 1.65°. The regression estimates the mode at 90.133° (not different from 90, $t_{1,36}$ = 0.1432, ns). The standard deviation, on the other hand, is clearly larger than expected, being estimated at 9.497° (larger than 1.65°, $t_{1,36}$ = 4.758, p < .01). Again, we see this as simply reflecting an underestimate of the JND size for angles; subjects are attempting to produce a spike, and are producing a slightly less sharp one. The predicted and estimated values for the mean and standard deviation of this modes, as well as for all the modes in all the experiments, are tabulated in Table 1.

Despite the difference in the width of the distribution, the fixed (predicted) model accounted for the data extremely well (R^2 = 0.5090, $F_{1,35}$ = 14.878, p < .001), and is plotted against the data in Fig. 5-7. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

One sample V, right

N = 18 x 6

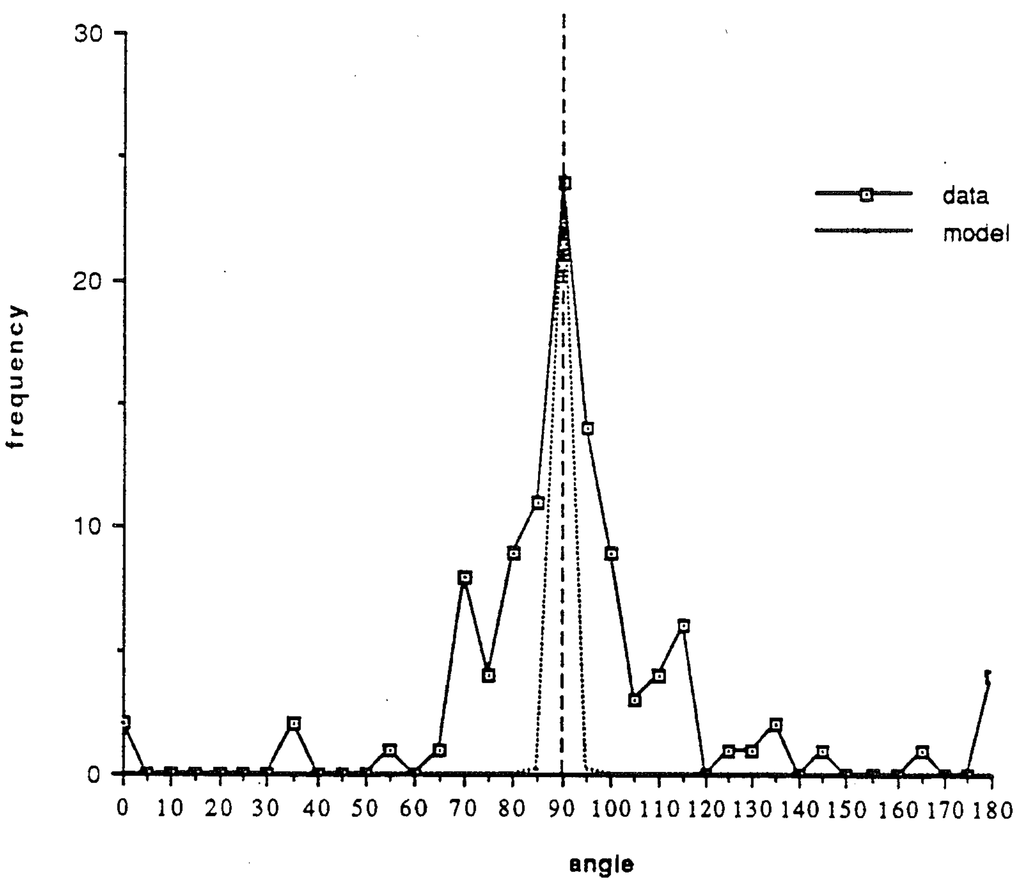


Fig. 5-7. Results for Experiment 5 (one right V). Model shown (dotted curve) is the "fixed" model using predicted mean and standard deviation. Predicted and fitted means of the mode are marked.

Discussion. As discussed above, this experiment clearly shows the marked difference between how special values and generic values and are used in category inferences. In Experiment 4, a single generic sample object engendered three distinct modes in subjects' probability distributions. When a special case is presented, though, subjects see it as such—a single, special type of object, to which other objects in the same class are likely to be identical in structure.

5.2.6. Experiment 6: 3 V's

This experiment simply corresponds to the three line-dots experiment. We expect more or less the same modes to appear in subjects' distributions as in the one-V case (Experiment 4)—i.e., one acute, one obtuse, and one at 90°—but this time the modes will not be triggered directly from the samples. Rather, the environmental distribution of observed angles, as in the 3 line-dots experiment, was carefully arranged to be flat, and to completely exclude non-generic cases.

Method. Again, an elaborate balancing scheme was constructed so that the pooled environmental frequency distribution seen by all subjects was perfectly flat, except for completely excluding non-generic cases, while the means of the individual subjects' sample sets were evenly distributed across the space. This was accomplished by presenting 6 different sample sets:

Sample set	Angles used	Mean angle of sample
1	25°,55°,130°	70°
2	30°,60°,135°	75°
3	35°,65°,140°	80°
4	40°,115°,145°	100°
5	45°,120°,150°	105°
6	50°,125°,155°	110°

Table 3. Individual induction samples used in Experiment 6.

Notice that each of the angles from 25°-65° and from 115°-155° is used exactly once over the whole set, so the overall distribution is exactly flat in these ranges. Notice as well that each subject saw either two acute and one obtuse angle or two obtuse and one acute, so no subject would be tempted to conclude that all members of the class must be of one kind or the other. Each subject saw objects of three different fixed sizes, permuted randomly. All the V's had one leg exactly horizontal. The concern here was to reduce the relatively number of "incorrect" inductions observed in the first generic V's experiment, by forcing subjects to focus on the angle between the legs.

Subjects. 24 subjects were tested in this condition.

Results and analysis. The results for Experiment 3 are plotted in Fig. 5-8. Two subjects' objects were uncodable (7.7%).

As expected, there are three modes in this distribution: one at 90°, one at a generic acute angle, and one at a generic obtuse angle. In Experiment 4 where there was only one example, subjects apparently placed these two generic modes at the locations of the objects they observed. Here, the modes come from "within the subjects' heads", and so there is no really firm prediction as to where the modes will to be. In the analogous situation with line-dots, it was possible to specify a point maximally distant from all non-generic cases, namely the quarter point. Here the geometry does not allow this. Though in principle there are non-generic cases as 0° and 180°, they are degenerate—do not appear to V's at all, but rather single line segments. Hence while the generic mode must clearly be far from 90°, it is not clear just how far.

We solve this problem with a bit of post-hoc analysis. The regression fits the two modal means to 60.925° and 118.220° respectively—suspiciously close to the two points that are twice as close to 90° as to 0° and 180° respectively, namely 60° and its mirror image 120°. There are of course a number of reasons why these points make good modes: they trisect the 180° space; the 30° separation from 90° has all sorts of special mathematical properties; and perhaps most saliently three 60° angle make an equilateral triangle. In any case, we incorporate this reasoning and produce a post-hoc "prediction" of generic modes at 60° and 90°. Hence when we find that this "prediction" is confirmed it should not be taken too seriously. Rather, it is just a slightly reversed route to discovering the locations of the generic modes.

The data from this experiment were much noisier than other experiments, and several mild massages had to be performed in order to get the estimation program to converge. First, a constant term was added to the fitted model, corresponding to the very high noise floor (visible in the figure below); see the discussion above of the meaning of excluding this term in the other experiments. Second, we smoothed the data slightly by collapsing into 10° rather than 5° increments; i.e., we treated the data as if they were collected using a slightly duller instrument.

Following the same reasoning as with all the other modes, we predict standard deviations of 30°/3.03 = 9.90° for both the 60° and 120° modes, and 5°/3.03 = 1.65° for the spike mode at 90°. Again, the regression estimates the acute mode mean to be at 60.925 (not different from 60°, $t_{18} = 0.1432$, ns), and its standard deviation to be 32.122° (broader than expected, $t_{18} = 0.5520$, $p < .01$). The non-generic mean was estimated to be at 94.652° (not different from 90°, $t_{18} = 0.8414$, ns); like all of the other 90° modes in other experiments, though, its standard deviation at 27.942° was broader than the spike prediction ($t_{18} = 9.4169$, $p < .01$). Finally the obtuse mode was estimated at 118.220° (not different from 120°, $t_{18} = 0.972$, ns) and its standard deviation at 24.687 (not different from prediction, $t_{18} = 1.9650$, ns). The predicted and estimated values for the means and standard deviations of these three modes, as well as for all the modes in all the experiments, are tabulated in Table 1.

The slightly post-hoc but nevertheless distinctly modal fitted model fits the data satisfactorily ($R^2 = 0.6045$, $F_{4,14} = 5.3503$, $p < .01$). It is plotted against the data in Fig. 5-8. As in the three line-dots experiment, the "environmental" distribution, the pooled distribution of objects that subjects observed, is also plotted for comparison. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

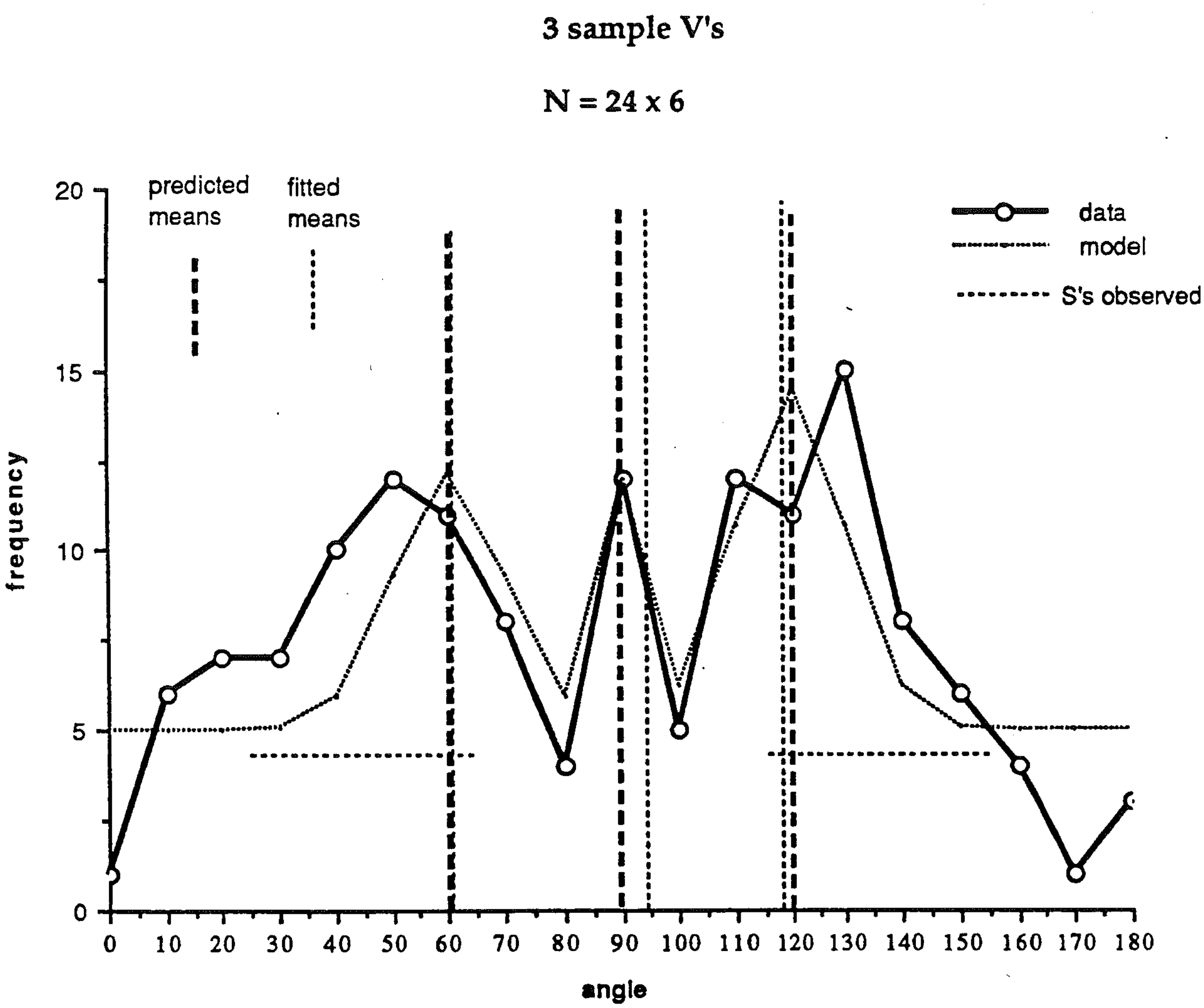


Fig. 5-8. Results for Experiment 6 (three Vs). Model shown (dotted curve) is the "fixed" model using predicted means and standard deviations. Predicted and fitted means of both modes are marked. The "environmental" distribution, the pooled distribution of objects that subjects' observed, is also plotted.

Discussion. As in Experiment 3 (3 line-dots), the critical point of this experiment is in contrasting the distributions produced by the subjects with the distribution they observed. As above, very clearly modal peaks occur in generic regions of the space at which the environmental distribution was perfectly flat (whether pooled or blocked into separate sample means). Even more striking is the non-generic mode, which covers an entire region of the space within which subjects never observed any examples at all. What is critical is that subjects do not simply fill in regions where no examples are observed, perhaps extrapolating regular distributions past where they actually operated. Rather, the 90° here must be seen as an invention out of whole cloth on the part of the subjects; it has no analog in the world at all.

5.2.7. Experiment 7: One variable line-dot

This experiment was designed to completely disentangle the variables contributing to the location of the generic mode in the results of Experiment. There, a mode appeared quite clearly at about one quarter of the way along the line segment, but by design it was impossible to tell whether subjects' were simply taking their cue from the sample in locating the mode there, or whether they were treating this point as some sort of a maximally generic case. Like the 3 line-dots and 3 V's experiments (in which a mode appeared at an empty region in the environmental distribution), this experiment was intended to help determine whether the modes we observe in subjects' data derive from summaries of observations or from internal inferential mechanisms.

In order to accomplish this, we again presented subjects with single line-dots as induction samples, but this time we varied the position of the dot between subjects, producing a flat pooled environmental distribution. First, as in the 3 line-dots and 3 V's experiments, we can compare the data with this flat environment. Second, and even more sensitively, we can separate the subjects into two groups according to what objects they observed: a near-endpoint group and a near-midpoint group. The "bottom-up" theory would have the two resulting distributions as widely separated from each other as the corresponding induction bases; the "top-down" theory would have the two groups produce more nearly the same distribution as each other, since both groups observe only "generic" objects.

Method. The induction example was again a line-segment with a dot placed some distance along it. One seventh of the subjects observed an object at each distance 2-8. The non-generic values 0 and 10 were of course omitted, and the values 1 and 9 were omitted as well, for fear that with only one sample they were too confusible with 0 and 10. The dot was always nearer to the lower end.

Subjects. 28 subjects were tested in this condition.

Results and analysis. The results for Experiment 7 are plotted in Fig. 5-9. The objects from 5 subjects (15.2%) were uncodable and were discarded.

As in all the other experiments, the frequency plot is clearly modal. Now, with examples shown from near the endpoint as well as the midpoint, all three possible modes are in evidence, and hence data the data were fitted to three Gaussians (rather than two as in Experiment 1). We again predict these generic mode to be at 5 and to have s.d. 5/3.03 = 1.65; and again we predict the non-generic modes to have means 0 and 10 and s.d. 1/3.03 = 0.330. The regression fits the endpoint mode at 0.359 (not different from 0, t_{10} = 0.3720, ns) with a standard deviation of 0.359 (not different from 0.330, t_{10} = 0.364, ns). The generic mode was estimated to be at 5.158 (not different from 5, t_{10} = 0.4401, ns), with a standard deviation of 1.526 (not different from 1.65, t_{10} = 0.3987, ns). Finally, the midpoint mode was fitted at 13.791 (not different from 10 due to a large amount of noise and resulting large standard error, t_{10} = .1567, ns); with a standard deviation of 2.926 (again not different from 0.330 due to large error, t_{10} = 0.2829, ns). Despite the null differences, the midpoint mode was clearly wider and further to the right than expected (actually well past the range of the data), reflecting a distribution that was flatter than true normal (see plot). The predicted and estimated values for the means and standard deviations of these two modes, as well as for all the modes in all the experiments, are tabulated in Table 1.

The fixed (predicted) model accounted for the data very well (R^2 = 0.8472, $F_{2,8}$ = 12.9326, $p < .005$); it is plotted against the data in Fig. 5-9. The fit of this model, as well as all the models for each of the experiments, is tabulated in Table 2.

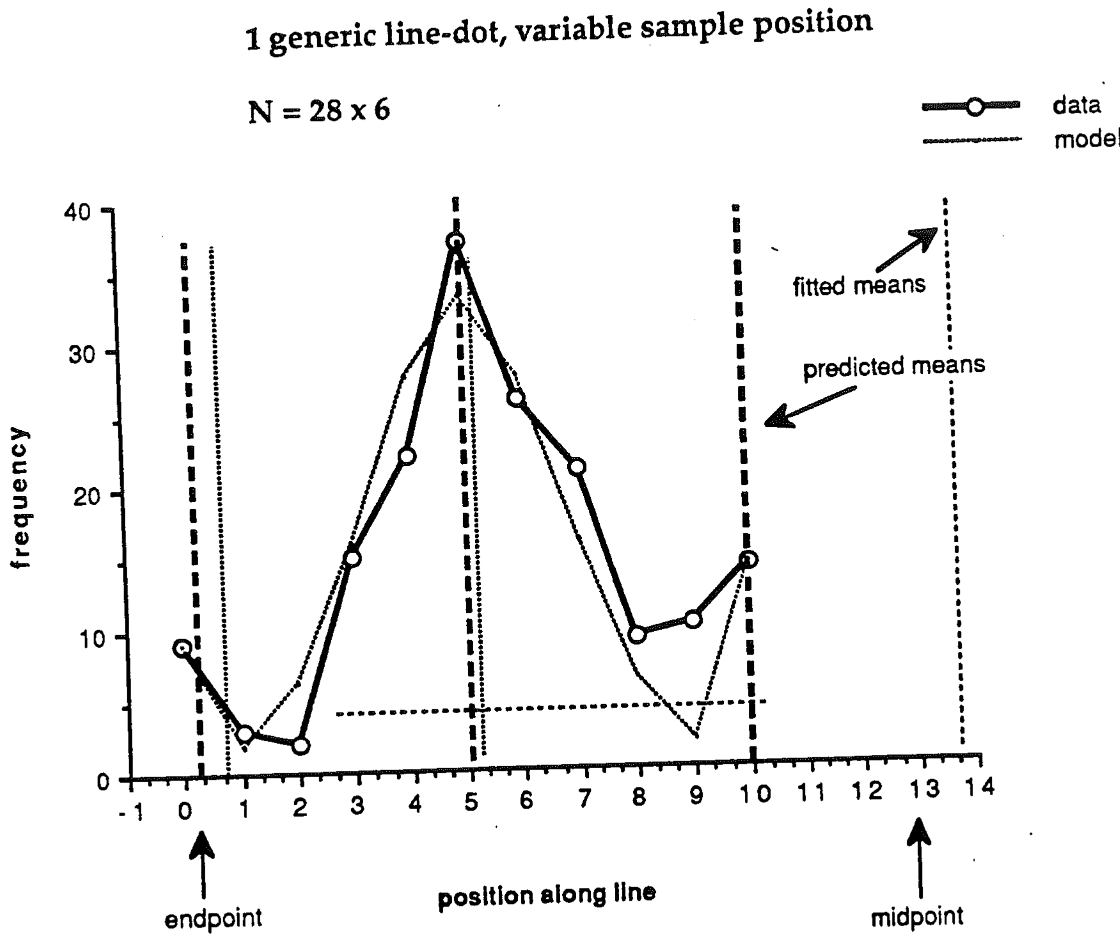


Fig. 5-9. Results for Experiment 7 (variable one line-dot). Model shown (dotted curve) is the "fixed" model using predicted means and standard deviations for the pooled subjects. Predicted and fitted means for each mode are also marked. The "environmental distribution," the pooled distribution of objects that subjects' observed, is also plotted.

- Chapter 6 -

Conclusion

6.0. Categorization as perceptual inference

This thesis makes an attempt to recast the categorization problem, often treated as an abstract cognitive problem, or even as an inherently philosophical one, as a problem basically of perceptual inference. In order to render it amenable to the same kind of computational approach as are other perceptual tasks, it is necessary to define the goal of the computation in as precise a way as possible. The categorization inference is simply the identification of an observed object with a larger class of which it is inferred to be a member; this inference affords us a window onto the complex pattern of property covariation that predominates in the natural world, giving rise to coherent natural categories of like objects.

To this end, a suitably general but non-trivial definition of "world regularity" was introduced. It was then shown that typical objects obeying such a regularity can be regarded as having been generated by the same generative model. For a given concept set—that is, for a given set of recognized properties—that set of objects arising from a particular regularity divide up naturally into equivalence classes of objects, each equivalence class corresponding to a distinct generative model. This division of the configuration space into distinct category models is a result of the so-called *genericity* constraint, which requires the observer to represent each object in a category model in which it is formally generic. The objects within an equivalence class, that is, are all typical of (generic in) the same generative model, and hence can be regarded as categorically uniform; while distinct equivalence classes correspond to objects in distinct, categorically distinguishable category models. For a given concept set, the set of distinct category models can be arranged naturally in a lattice, which depicts the dimensional relationships among the various models.

The resulting lattice collects together all of the category models that the observer considers potentially causally meaningful with respect to the chosen concept set. That is, the non-genericities incorporated in the lattice (as transitions from higher-dimensional models to lower-dimensional ones) are just those "coincidences" that the observer interprets as requiring a causal explanation: each non-genericity is an observed fact that the observer interprets as being intrinsic to the normal definition of the category. In this sense, the lattice serves as an explicit enumeration the "meaningful" configurations in the concept set, and hence serves as a kind of closed "category hypothesis generator," providing categories from which the observer can choose as dictated by particular observations.

Chs. 3 and 4 presented a system for constructing these lattices automatically by computer. The result for each family of objects is a set of distinct, special configurations that human observers are putatively willing to regard as distinct object categories. Ch. 5 reports a series of experiments in which evidence for these distinct category models was discovered in categories induced by human subjects from just one or three examples.

6.1. Summary of findings

A series of experiments were carried out in which subjects viewed a small set of simple, novel objects (each object was either a line segment with a dot positioned on it [line-dots], or two line segments with a common endpoint [V's]), with instructions suggesting that the object was to be taken to be an example of a category. The subjects were then asked to draw a number of novel examples of the category they had induced, and the frequency of their responses was plotted as a function of the varying structural parameter in the object class (position of the dot along the line for "line-dots", angle for "V's"). These plots all exhibited distinct peaks (modes), apparently corresponding to distinct object categories present in the subjects' category inferences. These modes correspond neatly to the distinct category models on the lattice appropriate for the sample object(s). The lattice theory aside, though, a number of strong conclusions about category inference can be drawn from the results.

•In two experiments in which the sample set consisted of only a single, generic object, subjects' induced distributions actually contained multiple modes (two modes for line-dots, one at the sample position and one at the endpoint; three modes for V's, one at each of two sample positions ["acute" and "obtuse"] and one at 90°). First of all, the fact that almost all the subjects drew the same category inference from just one example causes serious problems for any kind of a clustering model of human category induction,

References

Armstrong, S., Gleitman, L., & Gleitman, H. (1983) What some concepts might not be. *Cognition*, 13: 268-308.

Barr, A. (1984) Global and local deformations of solid primitives. *Computer Graphics*, 18, no.3, 21-30.

Barrow, H. & Tenenbaum, J. (1981) Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, 17, 75-116.

Bennett, B., Hoffman, D. & Prakash, C. (1989) *Observer mechanics: A formal theory of perception*. London: Academic Press.

Binford, T. (1981) Inferring surfaces from images. *Artifical Intelligence*, 17: 205-244.

Bobick, A. (1987) Natural object categorization. Technical Report 1001, M.I.T. A.I. Lab.

Bobick, A. & Richards, W. (1986) Classifying objects from visual information. A. I. Memo no. 879., M.I.T. A.I. Lab.

Bongard, M. (1970) *Pattern recognition*. New York: Spartan Books.

Buffart, H., Leeuwenberg, E., & Restle, F. (1981) Coding theory of visual pattern recognition. Journal of Experimental Psychology: Human Perception and Performance, 7(2): 241-270.

Chaitin, G. (1966) On the length of programs for computing finite binary sequences. *Journal of the Association for Computing Machinery*, 13 (4):547-569.

Chambers, J. & Kleiner, B. (1982) Graphical techniques for multivariate data and clustering. In P.R. Krishnaiah & L.N. Kanal (Eds.), *Handbook of statistics, Vol. 2*. North-Holland Publishing.

Clark, J. & Yuille, A. (1990) *Data fusion for sensory information processing systems*. Boston: Kluwer.

Cohen, B. & Murphy, G. (1984) Models of concepts. *Cognitive Science*, 8:27-58.

Cohen, J. & Cohen, P. (1983) *Applied multiple regression/correlation analysis for the behavioral sciences, Second edition*. Hillsdale, N.J.: Lawrence Erlbaum.

Darrell, T. Sclaroff, S. & Pentland, A. (1990) Segmentation by minimal description. *Proceedings of the Third International Conference on Computer Vision, Osaka, Japan*. Los Alamitos, CA: IEEE Computer Society Press.

Farlow, S. (1984) *Self-organizing methods in modeling: GMDH type algorithms*. New York: Marcel Dekker.

Feldman, J. (1991a) Perceptual decomposition as inference: continuous curvilinear processes. Cognitive Science Center Occasional Paper #42, Massachusetts Institute of Technology.

Feldman, J. (1991b) Perceptual simplicity and modes of structural generation. *Proceedings of the 13th Annual Conference of the Cognitive Science Society,* August, 1991.

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

Feldman, J., Jepson, A. & Richards, W. (1992). Is perception for real? *Proceedings of the Conference on Cognition and Representation,* SUNY Buffalo, April 1992, in press.

Fukunaga, K. (1982) Intrinsic dimensionality extraction. In P. Krishnaiah & L. Kanal (Eds.), *Handbook of statistics, Vol. 2.* North-Holland Publishing.

Gericke, H. (1963) *Lattice theory.* New York: Frederick Ungar Publishing.

Gold, E. (1967) Language learning in the limit. *Information and Control,* 10: 447-474.

Harman, G., Ranney, M., Salem, K., Doring, F., Epstein, J., & Jaworska, A. (1988) A theory of simplicity. *Proceedings of the 10th Annual Conference of the Cognitive Science Society, 1988.*

Hatfield, G. & Epstein, W. (1985) The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin,* 97, no.2, 155-186.

Hochberg, J. & McAlister, E. (1953) A quantitative approach to figural "goodness". *Journal of Experimental Psychology* , 46, 361-364.

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

Jain, A.K. & Chandrasekaran, B. (1982) Dimensionality and sample size considerations in pattern recognition practice. In P.R. Krishnaiah & L.N. Kanal (Eds.), *Handbook of statistics, Vol. 2.* North-Holland Publishing.

Jepson, A. & Richards, W. (1991) What is a percept? Occasional Paper #43, Center for Cognitive Science, Massachusetts Institute of Technology.

Jepson, A. & Richards, W. (1992) What makes a good feature? L. Harris & M. Jenkins (Eds.), *Spatial Vision in Humans and Robots,* Cambridge: Cambridge University Press, forthcoming. Also appears as A. I. Memo no. 1356., M.I.T. A.I. Lab.

Kanade, T. (1981) Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence,* 17: 409-460.

Keil, F. (1981) Constraints on knowledge and cognitive development. *Psychological Review,* 88(3) 197-227.

Kendall, D. & Kendall, W. (1980) Alignments in two-dimensional random sets of points. *Advances in Applied Probability,* 12:380-424.

Knill, D. & Kersten, D. (1991) Ideal perceptual observers for computation, psychophysics and neural networks. In R.J. Watt (Ed.), *Pattern Recognition by Man and Machine,* London: MacMillan.

Koenderink, J. (1990) *Solid shape.* Cambridge, MA: M.I.T. Press.

Koenderink, J. & van Doorn, A. (1976) The singularities of the visual mapping. *Biological Cybernetics,* 24: 51-59.

Koffka, K. (1935) *Principles of Gestalt psychology*. New York: Harcourt, Brace and Co.

Köhler, W. (1947) *Gestalt psychology*. New York: Liveright Publishing Co.

Kolmogorov, A. (1965) Three approaches to the quantitative definition of information. Problems of Information Transmission 1, no. 1, 1-7.

Landau, B. & Jackendoff, R., in press. "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences,* forthcoming.

Leeuwenberg, E., & Boselie, F. (1988) Against the likelihood principle in visual form perception. *Psychological Review,* 95, 485-491.

Leyton, M. (1984) Perceptual organization as nested control. Biological Cybernetics, 51, 141-153.

Leyton, M. (1988) A process-grammar for shape. *Artificial Intelligence,* 34, 213-247.

Leyton, M. (1989) Inferring causal history from shape. Cognitive Science, 13, 357-387.

Leyton, M. (1992) *Symmetry, causality, mind.* Cambridge, MA: M.I.T. Press.

Lowe, D. (1985) *Perceptual organization and visual recognition.* Boston: Kluwer Academic Publishers.

Lowe, D. (1987) Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence,* 31, 355-395.

Luce, R.D. (1990) Generalized concatenation structures that are translation homogeneous between singular points. Technical Report MBS 90-14, Irvine Research Unit in Mathematical Behavioral Sciences, University of California, Irvine.

Mark, L., Shapiro, B., and Shaw, R. (1986) Structural support for the perception of growth. *Journal of Experimental Psychology: Human Perception and Performance,* 12(2): 149-159.

Markus, M., Muller, S. Plesser, T. & Hess, B. (1987) On the recognition of order and disorder. *Biological Cybernetics,* 57, 187-195.

Marr, D. (1970) A theory for cerebral neocortex. *Proceedings of the Royal Society of London B,* 176, 161-234.

Marroquin, J. (1976) Human visual perception of structure. Unpublished Master's Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Martin-Löf, P, (1966) The definition of random sequences. Information and Control, 9, 602-619.

Matthews, G. & Hearne, J. (1991) Clustering without a metric. *IEEE Transactions on pattern analysis and machine intelligence,* 13 (2), 175-184.

Medin, D. & Ortony, A. (1988) Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Murphy, G. & Wisniewski, E. (1989) Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science, vol. 2: theory and applications*. Chichester: Ellist Norwood.

Murphy, G. & Medin, D. (1985) The role of theories in conceptual coherence. *Psychological Review*, 92:289-316.

Nosofsky, R. (1991) Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition*, 19:131-150.

Osherson, D. & Smith, E. (1981) On the adequacy of a prototype theory as a theory of concepts. *Cognition*, 9:35-58.

Osherson, D., Stob, M., & Weinstein, S. (1986) *Systems that learn*. Cambridge: M.I.T. Press.

Pearl, J. & Verma, T.S. (1991) A theory of inferred causation. In J. A. Allen, R. Fikes, & E. Sandewall (Eds.) *Principles of knowledge representation and reasoning: Proceedings of the Second International Conference.*. San Mateo, CA: Morgan Kaufmann.

Pentland, Al. (1986) Perceptual organization and the representation of natural form. *Artificial Intelligence*, 28:293-331.

Perkins, D. (1976) How good a bet is good form? *Perception*, 5, 393-406.

Pinker, S. & Prince, A. (1991) The nature of human concepts: insight from an unusual source. Unpublished manuscript.

Poston, T. & Stewart, I. (1976) *Taylor expansions and catastrophes*. London: Pitman.

Poston, T. & Stewart, I. (1978) *Catastrophe theory and its applications*. London: Pitman.

Quine, W. (1963) On simple theories of a complex world. Reprinted in M.H. Foster & M. L. Martin (eds), *Probability, confirmation, and simplicity: Readings in the philosophy of inductive logic*, New York: Odyssey Press, 1965.

Quine, W. (1985) Natural kinds. In H. Kornblith, (Ed.), *Naturalizing epistemology*. Cambridge: M.I.T. Press.

Quinlan, J.R. & Rivest, R.L. (1989) Inferring decision trees using the minimum description length principle. *Information & Control*, 80: 227-248.

Reuman, S. & Hoffman, D. (1986) Regularities of nature: the interpretation of visual motion. In A.P. Pentland (Ed.), *From pixels to predicates: recent advances in computational and robotic vision*. Norwood, N.J.: Ablex.

Richards, W. (1988) *Natural computation*. Cambridge: M.I.T. Press.

Richards, W. & Bobick, A. (1988) Playing twenty questions with nature. In Z. Pylyshyn (Ed.) *Computational processes in human vision: an interdisciplinary perspective*. Norwood, NJ: Ablex, 1988.

Richards, W., Feldman, J., and Jepson, A. (1992) From features to perceptual categories. *Proceedings of the British Machine Vision Conference*, Leeds, Great Britain, September 1992.

Rissanen, J. (1978) Modeling by shortest data description. *Automatica*, 14: 465-471.

Rock, I. (1983) *The logic of perception*. Cambridge: M.I.T. Press.

Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.) *Cognitive development and the acquisition of language*. New York: Academic Press.

Rosch, E. & Mervis, C. (1975) Family resemblances: Studies on the internal structure of categories. *Cognitive Psychology*, 7: 573-605

Rosch, E. Mervis. C. Gray, W., Johnson, D. & Boyes-braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology*, 8: 382-439.

Sacks, E. (1990) Automatic phase space analysis of dynamical systems. *Computing Systems in Engineering*, vol. 1(2-4), 607-614.

Sacks, E. & Doyle (1991) Prolegomena to any future qualitative physics. Technical Report CS-TR-314-91, Dept. of Computer Science, Princeton University.

Saund, E. (1987) Dimensionality-reduction using connectionist networks. A. I. Memo no. 941., M.I.T. A.I. Lab.

Schwartz, S. (1979) Natural kind terms. *Cognition*, 7(301-315).

Shepard, R. (1987) Toward a universal law of generalization for psychological science. *Science*, 237: 1317-1323.

Simon, H. (1972) Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79(5), 369-382.

Small, C. (1982) Random uniform triangles and the alignment problem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 91:315-322.

Smith, E. & Medin, D. (1981) *Categories and concepts*. Cambridge, MA: Harvard University Press.

Sober, E. (1975) *Simplicity*. London: Oxford University Press.

Thompson, D. (1917/1952) *On growth and form*. Cambridge: Cambridge University Press.

Ullman, S. (1979) *The interpretation of visual motion*. Cambridge, MA: M.I.T. Press.

Ullman, S. & Sha'ashua, A. (1988) Structural saliency: the detection of globally salient structures using a locally connected network. AI Memo No. 1061, M.I.T A.I. Lab.

Watanabe, P. (1985) *Pattern recognition: human and mechanical*. New York:John Wiley.

Weyl, H. (1952) *Symmetry*. Princeton, N.J.:Princeton University Press.

Witkin, A. & Tenenbaum, J. (1983) On the role of structure in vision. In H. Rosenfeld (Ed.), *Human and machine vision*. New York: Academic Press.

Witkin, A. & Tenenbaum, J. (1986) On perceptual organization. In A. Pentland (Ed.), *From pixels to predicates: recent advances in computational and robotic vision*. Norwood, N.J.: Ablex.

-Appendix A-

Classification of constraint manifolds by smooth shape:

Sketch of a theory

Qualitative model fitting and the continuity-causality principle. We have conceptualized each category of mutually co-originating objects as a smooth surface of some kind. Here we sketch the outlines of a theory by which they might be classified. The idea is that geometrically distinct constraint manifolds (i.e., surfaces with different canonic forms in the Morse classification) correspond to qualitatively different category lattices. The smooth classification itself is derived from Poston & Stewart (1976, 1978) and Koenderink (1990).

The classification is according to their smooth geometry, but since each geometrically distinct surface type entails some distinct parameterization, the classification also feeds back to the types of discrete lattices discussed in Ch. 3. In this way, the intent is to identify the discrete patterns that discrete feature covariation must take in order to qualify as potentially signifying a smooth constraint model of some kind—the link to continuity and hence causality. Hence just as we classified the different ways that property covariation might occur among discrete properties in Ch. 3, we now propose to classify the different shapes of smooth relationship among continuous parameters. Ultimately, these two classifications should line up, giving an underlying formal structure to the inference of causal coherence in the world from observed covariation of both the continuous and discrete types. The alignment between the continuous and discrete classifications takes the form of a mapping between the superordinate lattices that interrelated canonic forms of covariation in both cases.

To see the mapping from a patch of smooth manifold to a parameterization and hence to a lattice, consider how the parameterization axioms of Ch. 2 require that the new transverse operations, among other things, be intrinsic to the surface, and be less numerous than the embedding space (i.e. the surface has codimension at least 1). Consider the codimension 1 case first. These conditions imply that there will be (at least) one direction that is normal to the surface, that is, perpendicular to all of the new parameters, but still lies within the space. We now take this "parameter" to stand in for some diagnostic function, another parameter that can have either a positive, negative, or zero value for points on the surface. We may alternatively think of this "diagnostic function" as simply some hidden property, an additional unobserved property of interest to the observer. The classification, then, is just the classification of the ways in which this hidden or diagnostic feature's value can vary as we move through the quadrants on the surface of the manifold. The unique plane to which this new parameter is normal is tangent to the constraint manifold at the point of intersection. This plane contains local linear approximations to all of the chosen parameters.

To make clear how a point on a constraint manifold becomes the origin of a parameter space and hence generates a lattice, we must adopt some well-defined directions along the manifold, as parameters for the new space. A natural choice from a mathematical point of view is the two orthogonal principal axes of curvature that Gaussian geometry gives us along a 2-D surface embedded in 3-space. For non-elliptical surfaces, it turns out that an even better choice is simply the directions of zero curvature. For cylindrical surfaces these two choices are the same (the minimum curvature direction is the zero curvature direction). For hyperbolic surfaces the zeroes segregate the surface into regions that are curving in the same direction, which is useful when we attempt to characterize points within each quadrant as categorically uniform. For elliptical points, we choose the principal axes (Fig. A-1).

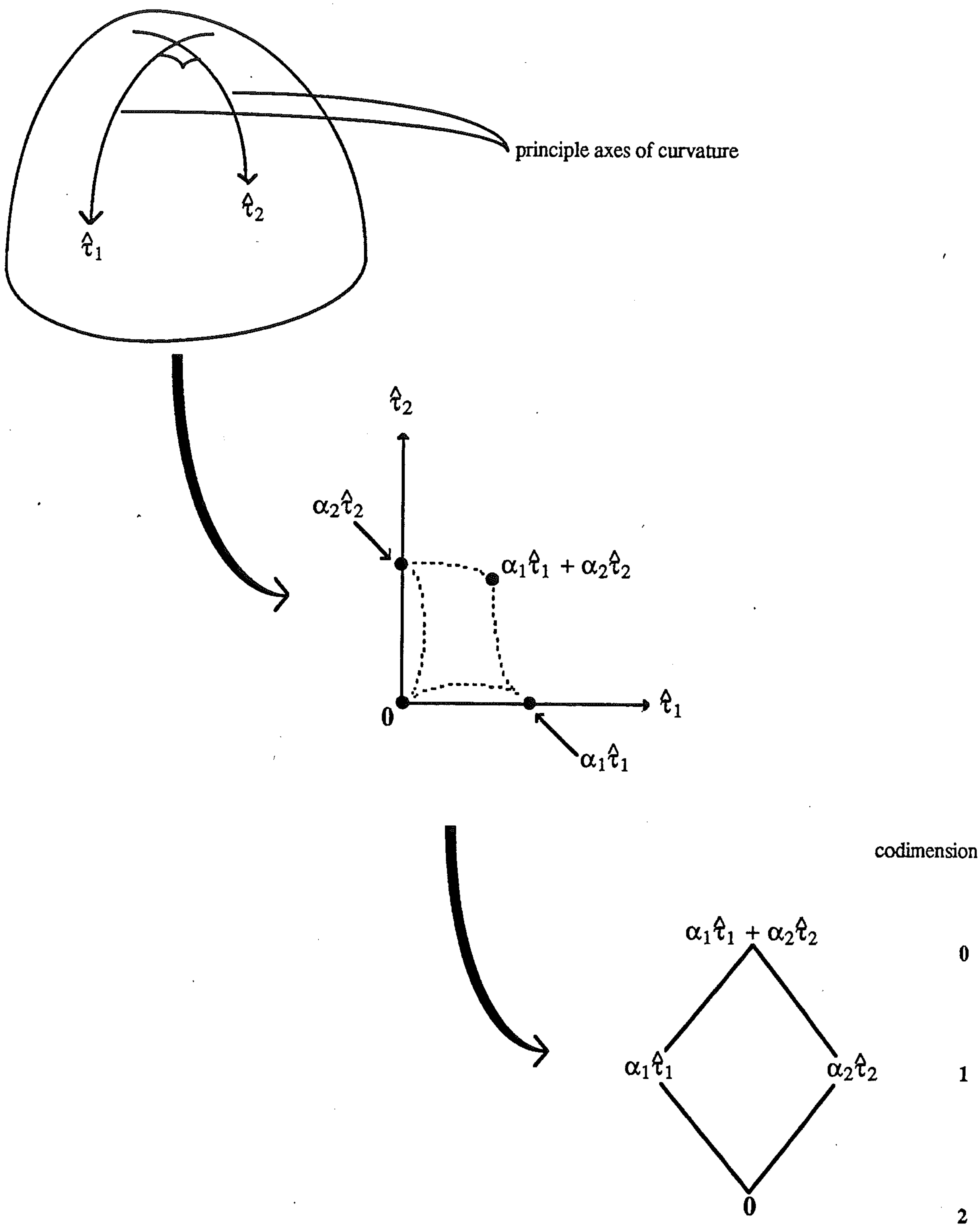


Fig. A-1. A typical parameterization of a point on a smooth surface, leading to the lattice of categories, i.e. regions on the surface of the manifold.

Morse classification. The key observation is that due simply to the way in which we have defined the operations—so that the local intrinsic coordinate system is tangent to the surface locally containing the parameters of the new space , this new diagnostic function will have a (generalized) *extremum* at the origin. This fact simply follows from how we have axiomatically chosen the orientation of the coordinate system.

To classify the codimension 1 surfaces, then, we simply need to classify the extrema. On this point, a theorem from differential geometry, the Morse Lemma (see Poston & Stewart, 1978) provides the appropriate tool: after equivalent relabelings, there are just $n+1$ qualitatively different types. In essence each type has a different number of dimensions pointing "up" (positive in our diagnostic function) and different number pointing down (negative). If some are up and some are down, the surface is a "saddle" of some kind; if all point the same way it is a simple extremum of some kind. (For logical completeness, the minimum is renamed an 0-saddle, a saddle with k dimensions pointing up a k-saddle, and the maximum the n-saddle. The familiar 1-saddle, a 2-D surface unique in 3-space, is of course the reason for the name.)

discrete representation, one property takes on both signs while the other is always forced to have one sign. Finally, in the superordinate codim-2 case, where curvature itself is zero, we have an inflection, so change in sign of one variable lawfully leads to change in sign in other variable . The same reasoning applies to two (and higher) dimension surfaces, except that now there are two dimensions of curvature, conveniently defined as the two components of Gaussian curvature.

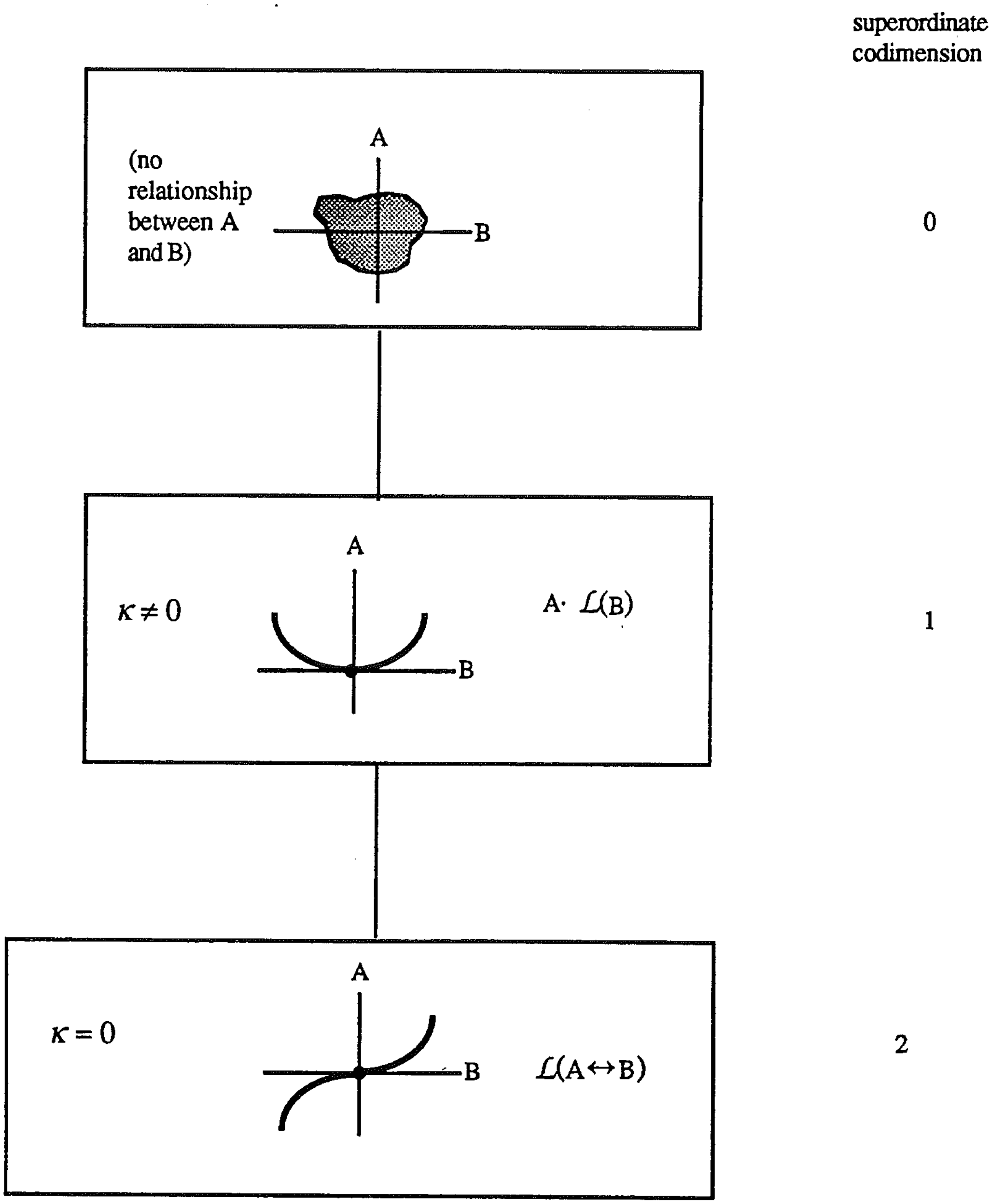


Fig. A-3. Superordinate lattice in two parameters, showing the different relationship two continuous parameters or discrete properties can manifest.

It is possible to construct a finer classification, in which some apparently qualitatively distinguishable types are distinguished. As curvature is in a sense the derivative of the tangent plane, the obvious choice for the next meta-parameter is the derivative of the curvature. The entailed superordinate lattice follows about the same lines as the above, except that it is naturally two layers deeper, and need not be laboriously described here. Essentially, the new meta-parameter allows sensitivity to extrema in the curvature. It seems natural that a point on a constraint manifold that falls at a maximum of curvature, for example—a condition that is not perturbable, but rather defines a special point—might denote a qualitatively different category model from one in which the curvature had an intermediate, generic value. Such points will naturally fall at zeroes (modal non-genericities) in the curvature-change meta-parameter. As for the discrete category model, such a point may be thought of as the origin of a defining operation that has a "sharp edge", i.e. in which the origin cannot be perturbed. Moreover, this meta-parameter allows us to distinguish between a planar point, at which all meta-parameters vanish, from an extruded inflection, which while also having both curvature components zero is clearly qualitatively different.

-Appendix B-

Sample experimental form

Forms in all experiments appeared in this manner. When three examples were presented on the first page instead of one, the instruction read *"Here are some blickets"*. The words *"page 1 of experimental form"* (etc.), seen here, did not appear in the experiments. For convenience, the forms are presented slightly smaller than they actually appeared.

Date______ Initials______

Thank you for participating in this brief experiment. (It should take about five minutes.)

There are no right or wrong answers; this is not an intelligence test! We are interested in what you think is a reasonable answer to each question.

→It is very important that you **finish** the questions on each page before turning to the **next** page!

Here is an example of a *blicket:*

(page 1 of experimental form)

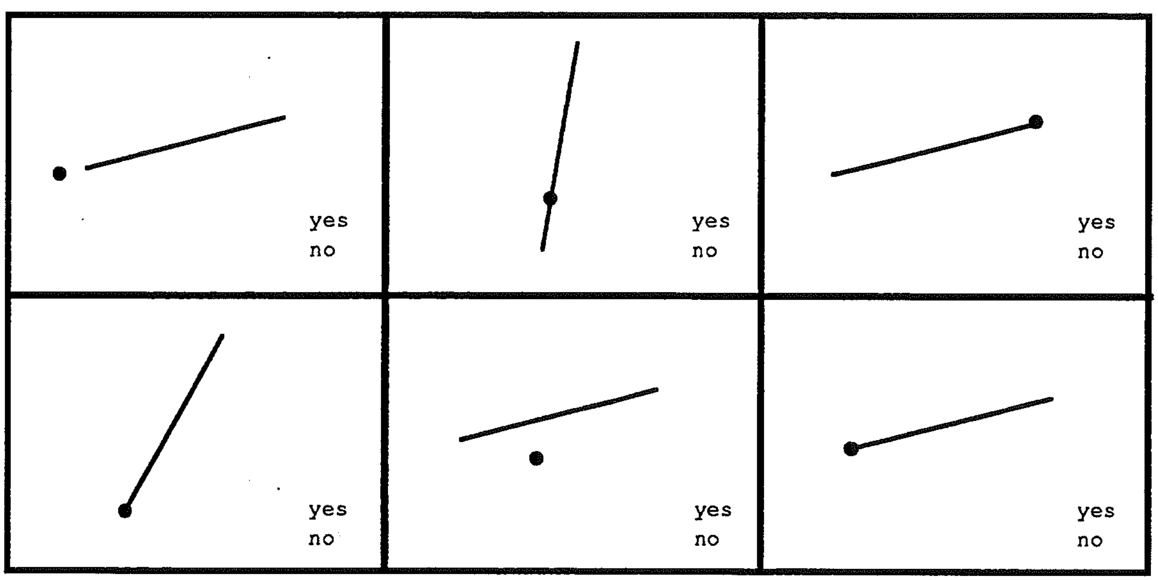
Please draw 6 different examples of a blicket.

(page 2 of experimental form)

In a few words, how would you describe a blicket?

(page 3 of experimental form)

For each of these, indicate (circle yes or no) whether you think it is also a blicket.



(page 4 of experimental form)