Constructing Perceptual Categories

Jacob Feldman
Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology

Abstract

Human observers abstract away from some particular properties of observed objects, while representing other properties faithfully, thus identifying each observed object with a more general class of which it is interpreted as a "typical" member. Generalizing correctly—constructing a model of sample objects—involves an inductive leap that people can make confidently after just a single example ("one-shot categorization"), indicating heavy constraint on the set of possible category models. This paper shows how a parameterization of the structure of the observed object (interpretable as the space of dimensions of generative operations which brought the object into existence) entails a lattice that enumerates all the allowable categories and subcategories for the class of object, along with an inferential preference hierarchy among them. We constrain each model to be generic in its parameterization, so that each node in the lattice stands in for an entire class of objects that, all being parameterized the same way, can all be treated as equivalent to one another: the observed object's "natural" category.

1. Categorization and world structure

"Natural" categories are loci of property stability: classes of objects within which there is relatively little significant structural variation, but between which there is relatively much. In this sense, the very existence of categories depends on the highly regular nature of the world. Yet identifying parameter covariation in the world has enormous inherent inductive ambiguity, since with finite sample size or resolution there are always many more possible patterns of covariation than there are data to distinguish among them. Deciding just which parameters make plausible candidates for variation—building a mental model of the category—is thus essentially an act of inspired guesswork.

People seem to make such guesses well: the human capacity to categorize involves a shrewd ability to distinguish, on the basis of very little immediate evidence, between object parameters that will vary in novel unobserved instances of the category, and other parameters that will remain fixed. Our categorization competence, by employing in essence "a model of a good model," is thus able to cut down the field of category inductions into the range of the manageable. This allows a model to be built around the fixed parameters which is suitably flexible, sensitively incorporating the world regularity, and not getting entangled in possibly spurious correlations of features that are unlikely to correspond to real category processes in the world as it is actually structured.

Indeed, recent research in cognitive categories [16] suggests that human subjects do not consider all possible sources of feature covariation in an unbiased manner, but rather "zero in" on those feature pairs that can be fit easily into a coherent theory or model of the world, i.e., those that one would expect to covary in some plausible schema. Subjects were found to examine feature pairs that were part of their category model very carefully, being quite sensitive to any observed correlation, while ignoring even perfect correlations along "atheoretic" lines, i.e., those that they would have no sensible interpretation for even if they noticed them.

The necessity of reducing the number of candidate categorizations is brought into even sharper focus by the existence of "one-shot" categorizations: generalizations that human subjects can make from just one object. Consider, for example, the object in Fig. 1(a).

Fig. 1.

One might imagine many "solutions", but it seems that human observers typically agree on the following induced "regularity"; while the position of the dot along the segment might vary in other examples taken from the same "natural category," the dot would probably always fall somewhere on the segment. Clearly, this inductive distinction has something to do with the fact that the dot's translation away from the segment manifests a very special value, zero, in the observed object. It seems unlikely to have done so by accident; that is, unlikely unless that value was intrinsic to the definition of the category. Because it has this very special value in the observed object, we infer that it is supposed to have this value—that it is part of the defining characteristic of the class. Thus the model class is judged to be: segment with a dot on it. More examples are shown in Fig. 1(b).
observed object, critically, is well interpreted as a typical example of this class.

One-shot categorization creates an apparent paradox for the idea that categories derive their existence from the covariation of features in the environment, since, after all, we cannot observe even the first confirmation of a hypothesized covariation until a second example is evaluated. One aim of this paper is to show how this paradox—covariation without variation—can be resolved. In brief, property covariation will be seen as just a special case of the notion of "special" (non-generic) feature; such features will serve as key pointers to the correct category model. Fig. 1 is an example: a parameter that we, as observers, judged to be structurally important (namely, translation off the segment) took on a special value that it was unlikely to take by accident (namely, zero); so we incorporated this fact into our model of the object, and then generalized accordingly. Similarly, correlation of two (say, continuous valued) parameters creates a line (a 1-D object) in 2-space: failing exactly on this line (i.e., exhibiting the covariation that defines the category) is unlikely to happen by accident—that is, unlikely to happen unless the point really belongs to the category (cf. [4,14]). This notion will appear in more general form as "positive codimension" in Section 2.3. Unlike the case with observed covariation, though, if the observer has a prior conception of "special value", then just one example may be enough to qualify and trigger the model.

Following the above argument, then, the overall structure of this paper will be first to axiomatize the notion of a parameterization, one along whose parameters it makes sense for the observer to look for special values (Section 2); propose a definition of "special value" along these parameters, really a zero in some parameterization (Section 3); and then, finally, show how choice of the parameterization and its zeroes entails a small, enumerable space of categories from which the observer can choose, which can be depicted schematically as a lattice (Section 4), in a way that was laid out more briefly in [7], and that is similar to [9]. We call the entailed types of regularities "modal," since they invoke modes of covariation that in some sense fall within the observer's conceptual structure. Hence the choice of parameterization directly entails constraint on the space of categories: the observer generalizes, and is in a sense trying to minimize the complexity of the generalization, but is capable of detecting regularities only so far as they do not range out of the small category space—that is, as long as they are on on the lattice. In short the observer organizes the categorization around modal regularities.

1.1. Categories as regularities

It is now a conventional view in cognitive science that natural categories somehow derive their existence from the correlation of features in the environment [see 15,20]. The regular, predictable quality of object properties typically manifests itself as the co-occurrence of discrete features. 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 (say) very readily, one tends not to encounter them as often. In this view, categories amount to equivalence classes of co-occurring 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."

Clearly, however, human observers cannot expect to detect, or in fact even to be able to conceptualize, every possible sort of regularity. There is an infinity of different "regularities" consistent with any sample, the vast majority of which are arbitrary, complex, and invoke concepts and patterns that the human observer would not recognize as in any sense "regular." Even restricting the complexity of hypothesized patterns, though, does not constrain away enough possible categorizations, since what is simple in one descriptive language may be complex in another. On the other hand, some higher dimension categories, while plainly constituting weaker inferences (being, for one thing, more prone to be confirmed by accident), seem to be more "built in" to our conceptual apparatus, and we pick them up in favor of formally simpler ones. The situation bears an intriguing analogy to the ubiquitous problem of "over-fitting" numerical data. A low-dimension surface (i.e., category) may be fit to any data, especially if its shape is allowed to be very convoluted. But the ability of such as surface to generalize properly for new points is often very poor, while poorer-fitting but flatter categorizations often tend to generalize better. The question to be answered before a theory of categorization can get off the ground is: what does it mean to "generalize better"?

Our conclusion, for the moment, is that while observers may prefer to minimize dimensionality in their induced categories, they are extremely constrained in what type of category they may invoke in order to do so. We propose that in human categorization this constraint takes the form of a closed class of categories and their relationships, the mode lattice, which is directly entailed by an induced parameterization, and which constitutes the entire universe of choices from which the observer is constrained to draw.

2. Parameterizations

A critical part of the problem of categorization is choosing a parameterization, an identification of the variables that most directly mediate between the apparent structure and deeper properties of an object. This choice in fact putatively determines all the categories available to the observer for a particular class of objects, in a way to
be laid out in Section 3. This section axiomatically defines the essential properties of parameterizations that make the category lattice schema possible. An intriguing comparison may be made between the perceptual/inferential treatment of parameterizations here and in [7], and the statistical treatment of the parameterizations of triangles in [11]; the statistical central tendencies of random triangles, it turns out, are sometimes peculiarly at odds with the perceptual central tendencies (i.e., special categories).

2.1. Genericity and category flatland

A parameterization is a set of variables that may vary within a category, as contrasted with others on which variance entails leaving the category, due to the violation of some defining category constraint. The parameterization may thus be thought of as a surface along which an object may travel freely, but which it may not leave without losing its categorical identity. (This idea will be made more explicit below.) Hence the objects in a uniform category inhabit a kind of “flatland” (after the mythical place where 2-dimensional creatures live [11]). To remain part of the category, definitionally, they cannot roam into alien dimensions (Fig. 2). On the other hand, objects that fall within the same flatland can be treated as equivalent, since they can roam freely about without changing their identity. This is the intuition behind the genericity constraint: parameterize each object so that it is generic in its parameterization—so that it may be modeled as a typical object of some kind. Then, each parameterization may be treated as a categorial unit; in Section 3 each such unit will be represented as a node in a mode lattice.

![Diagram of generic object and category flatland](image)

**Fig. 2.**

2.2. Generative models

The parameters in a parameterization may be meaningfully taken to entail object categories if they are interpreted as dimensions of some generative process or procedure. That is, we imagine some procedure in which a set of one-parameter operations—stretching, shearing, and so forth—is carried out on some primitive object successively, after the manner of [12,13]; and then take the values along the various parameters to be the magnitudes to which the various operations were carried out. Then, a generic point in the parameterized space would naturally correspond to a typical product of the stipulated generative process: that is, an object which was produced by some non-zero but otherwise undistinguished degree of bending, twisting, stretching and so forth.

In addition to such simple mathematical operations as stretching and shearing, tapering, bending and twisting may also performed by simple matrix multiplication [2]. Conveniently for the identification of these mathematical operations with natural processes whereby objects’ shapes are determined, a number of simple mathematical operations have been shown to closely model natural growth of various kinds. Most famously, Thompson [21] showed 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 routinely, object shapes can be varied by using simple translations, rotations, and scale changes acting on just one part of an object with respect to the main body (rather than as usual acting on the entire object). Thus we can turn one part with respect to the main axis, or lengthen one part—corresponding to, say, bending an arm, lengthening a neck, and so forth. The examples in this paper were all generated by this sort of operation.

2.3. Some axioms

We now present some axiomatic constraints on the form of the constraint manifold, that attempt to capture the notion of a causally coherent collection of objects obeying a common regularity in some measurement space, and on the notion of a parameterization of that manifold. Space requires that the axioms be presented very briefly here; a fuller treatment may be found in [8].

We assume that observed objects fall in some space $\Phi = \phi_1...\phi_d$, the $d$ real-valued dimensions $\phi_i$ of which are just measurements defined on the objects that we care to take, prior to any analysis. Our central model of a unary, causally coherent category is a (locally) smooth subset (hence the term “constraint manifold”) of non-zero codimension embedded in this space (Fig. 2). This manifold embodies a constraint, condition, or rule defined...
in this space, that a point must satisfy in order to count as a member of the category.

A1 Continuity and smoothness. Assuming that the category is continuous—that for every point \( x \) in the category there is an entire neighborhood of points arbitrarily close to \( x \) that also belong to the category—satisfies an intuition that objects’ membership in a real-world category does not typically depend on an entirely idiosyncratic way on their structure, but rather that their membership is grounded in their structure via some reasonably transparent causal mechanism; equivalently, that the category in question contains objects produced by some causally coherent generative process (see [6,8]; cf. [17]). Assuming further that the subset is smooth (has derivatives of arbitrary order) embodies the idea that objects in a “natural” category all embody the same, uniform constraint. Continuity and smoothness are highly unusual in the space of all possible processes (they have infinite codimension in the space of all point sets). Even this simple axiom, then, ascribes an enormous amount of structure to the world.

A2 Non-zero codimension. The codimension of the manifold is simply the difference between the dimension of the manifold and that of the space in which it is embedded; for instance a plane curve has codimension one. The codimension is generally equal to the number of independent conditions that member points must satisfy (see [18]; also see [9,10]). Non-zero codimension expresses the idea that the category excludes members in a non-trivial way, and that objects are unlikely to satisfy the defining constraint by accident. This requires that the category model picks out some causal structure with respect to the less structured model implicitly embodied by the parameter space at large. In this sense, the definition of “structure” is always relative to a prior, more general, model.

Note that non-zero codimension means that the manifold can always be expressed as a deformation of some proper subspace of the naive space \( \Phi \). Hence, rather than expressing the manifold in terms of the naive parameters, we can reparameterize the manifold by some new, intrinsic parameters, \( \xi_1, \ldots, \xi_k \), where \( k \) is the dimension of the manifold, and hence \( d-k \) its codimension, where the new parameters may be thought of naturally as unit translations along the surface of the constraint manifold; these operations will then be re-interpreted as the parameters of a new, lower-dimensional space (see Fig. 2,3). In order to fully express the new space, these operations must obey:

A3 Transversality. This condition simply requires that the new parameters span the full space of the constraint manifold, i.e., they cross each other in a “typical” fashion. In essence, this means that each of them is independent of the others, that is, that no two of them point in the same direction. This axiom allows us to conclude, as desired, that the dimension of the new space, in which the manifold is flattened out, is strictly less than the dimension of the naive space.

A4 Intrinsic parameterization. Finally, in order to complete the picture, we must assume that the new parameters are intrinsic to the surface of the manifold, i.e., that they lie along its surface. To clarify this idea, consider that any category that has only one free parameter, like the dot-line category discussed above, will always appear as a 1-D manifold in whatever naive space it is depicted. We seek to collapse over the various shapes it may take in different spaces, by considering (some function of) distance along this manifold to be the most essential, “natural” expression of the single free variable. The same general idea carries over to higher dimensions.

An example should make this perfectly clear. Consider the class of “V”s, formed from two coterminal segments of equal length. The task of parameterizing this object is really just that of describing the relation between the two segments. Ignoring scale and position, one way to proceed is simply to fix a Cartesian frame to one of the segments, and parameterize the endpoint of the other segment as \( \langle \Delta x, \Delta y \rangle \) in this frame. Now, since the segments are constrained to be of equal length, the category follows a circular arc passing through this 2-D space—i.e., a manifold of codimension 1 (Fig. 3). Note, however, that the parameterization is not intrinsic to the manifold, since the manifold curves through the parameterized space. We can correct this easily by taking arclength \( \sigma \) along the manifold, or some function of it \( f(\sigma) \), as a single new parameter. Now, the parameterization is open and has the correct (intrinsic) number of parameters, namely one. This, then, is our hypothetical natural parameterization.

![Fig. 3](image)

Note that translation along the constraint manifold in both parameterizations changes the angle between the segments; the natural parameterization expresses this more “cleanly” (i.e., flatly) and with the correct zeroes (which will have implications later for the subcategories).

A5 Genericity. We would like to reject parameterizations of which the observed object would have to be considered an unacceptably unusual instance; this axiom gives us the leverage to do so. In essence, we want to parameterize such that observed object is typical of points in the new parameterization. We have assumed that each object can be seen as a weighted sum of the new,
transverse parameters: the genericity assumption simply requires that none of the weights is zero. If, by contrast, some object had a description in which one of the weights were zero, this would mean that the object fell in one of the subspaces of the parameterization; in this case, the object could have been parameterized with fewer parameters—that is, completely omitting the one on which it measured zero. That is,
\[ \alpha_1 \hat{t}_1 + 0 \hat{t}_2 = \alpha_1 \hat{t}_1 , \]
\[ 0 \hat{t}_1 = 0 , \]
and so forth.

The genericity constraint allows us to parameterize an object in such a way that it is generic; by the other side of the same coin, this means we can collapse over any differences between objects that are generic in the same parameterization. Subspaces of a parameterization—subsets in which the objects are generic in lower-dimensional generative sequences—thus amount distinct categories. The collection of such distinct categories are, by way of genericity, directly entailed by the parameterization. Section 3 will take advantage of this fact to yield as simple and compact notation for categories.

**Setting the zeroes.** The choice of zeroes on the parameters is particularly critical, as they will be the locations at which the genericity constraint requires a parametric collapse to a lower dimension. Again, space requires that the reader refer to [8] for a fuller discussion. Briefly, the idea is that some small set of values along each parameter are singled out as structurally special: zeroes, for example; multiples of \( \pi/2 \) on angular or periodic parameters; midpoints; and so forth. The argument is that such values are particularly prone to fall at maxima or minima of some utility function or other function diagnostic of structure. While, plainly, it is impossible to assign a probability to such a situation, the idea is that the observer's conceptual apparatus is equipped to consider it as a discrete possibility.

**Parametric collapse at modal non-genericities.** With a special value chosen, the parameter is redefined as the distance along the parameter away from the special value. Critically, then, this new parameter will vanish just when the parameter takes on the special value. At such a point—a modal non-genericity—a special subcategory occurs, one which characteristically does not involve the parameter in question in its generative model. Say for instance we have a special value \( s \) along the parameter \( \alpha_1 \). At \( \alpha_1 = s \) there will be a parametric collapse from a higher-dimension category \( \alpha_1 \hat{t}_1 + (\alpha_1 - s) \hat{t}_2 \) to the lower-dimension category \( \hat{t}_1 \). The genericity constraint dictates a shift from one "flatland" to an even flatter one.

The general organization of such shifts forms a lattice; this will be the subject of the next section.

3. The lattice of category models

The fact that genericity collapses over stable regions (subspaces) of the overall parameterization—that is, that we would like to treat as equivalent all objects that are generic in the same parameterization—leads to a useful notation. We represent each region (subspace) by a node in a directed graph, and then simply connect it to all the nodes that are subspaces of it of dimension one lower.

Thus we have \( a \) whenever \( b \) is a subspace of \( a \) and \( \dim b < \dim a \). This relation forms a lattice (a certain kind of partial order than can be pictorially depicted in a convenient way).

Fig 4 gives an example of the schematic representation. This notation deliberately disregards any differences within a region that contains only objects that ought to be parameterized the same way. That is, while a point in the \( x \)-\( y \) plane and a point in 3-space can both be expressed as a sum of three translations, only the latter should be so expressed: to parameterize the planar point in that way would violate the genericity axiom.

![Fig. 4.](image)

(i) Example: \( V \)'s. Consider a simple 2-parameter category whose lattice takes the above form: a pair of segments joined at one endpoint, the \( V \)'s of the reparameterization example. The "natural" parameter expressing the difference between the orientations of the legs, as in Fig. 3, is the angle, or arc length 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 \( V \) may be conceived to have been determined. The parameters and their zeroes entail a lattice (Fig. 5).
At the top of the lattice is the codimension 0 category, the one in which a completely unconstrained \( V \) is generic. Non-generic \( V \)'s, of course, 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-length \( V \)'s. Each of these categories is codimension 1 in the overall space. At the bottom is the codimension 2 category, in which there are no degrees of shape freedom left to vary.

(ii) \textit{Example: line with dot.} The lattice for the relation between a line segment and a dot is shown in Fig. 6. Here, in contrast to Fig. 5, the lattice has only one codimension 1 case, “dot on line.” As per genericity, the category “dot on endpoint” is distinct from “dot on line,” even though the former is a subset of the latter.

This lattice, of course, putatively accounts for the categorical interpretation suggested in Fig. 1.

(iii) \textit{A 3-parameter example with multiple zeroes.} Now we consider a more complex example, with some added wrinkles: two segments crossing. We adopt a fairly straightforward natural parameterization of this 3-parameter class based on 1) the angle between segments, 2) the overhang of one segment across the other, and 3) the position of the intersection along one segment. The natural zero for parameter 1 is the right angle. For the second parameter it is the zero overhang (i.e. a \( T \)). For the third parameter we may easily adopt two distinct zeroes: one a midpoint, and the other an endpoint. The resulting lattice is shown in Fig. 7. The codimension 1 categories, which manifest the special properties in pure form, are a skew-symmetric cross, a right-angle intersection, and a \( T \).

Note that two cases in the codimension 1 row are listed twice because of symmetries: in each case, two objects that are group-equivalent actually have a non-empty intersection. For clarity, degenerate objects (for example, the object with zero angle, in which the two segments coincide) are omitted.
The lattice enumerates categories of objects, and puts a partial ordering on their regularity, in a compelling way. It constructs a constrained space of categories from which the observer may choose, but from which it is constrained not to depart. Any regularities not listed—and of course the vast infinite majority of conceivable ones are not—will be chronically missed. While much specification of human preferences among parameterizations remains, of course, completely open, the idea of the mode lattice provide a route whereby our conceptions of a meaningful regularity lend structure to our categorical interpretations.

4. Conclusion: Modal non-genericities

The notion of non-genericity often comes up in perceptual theory to refer special properties that we expect typically not to happen: for example, the special viewpoint in which two arbitrary non-coterminal line segments in 3-D appear, misleadingly, to be coterminal. Coterminous is thus inferentially safe because the relevant false target, in which the inference goes awry, is non-generic [14, 23]. The same reasoning may apply to more generalized image events [3]. A modal non-genericity is a non-generic configuration that we expect, in some sense, to occur—i.e., for which we have a coherent model. When such a non-generic regularity does occur, we may safely conclude that it was supposed to occur. That is, that its occurrence is causally tied to its category membership, and it is prone to occur again, reliably, in novel instances of the same category.

In a classic essay, the mathematician Weyl [22] noted our ubiquitous tendency to ascribe symmetries to the world, and to describe the world in such a way as to emphasize its symmetry. Similarly, in [18] Poston and Stewart have noted that many global world symmetries and regularities that are to in principle unverifiable (such as the multifarious symmetries of Euclidean space) may have very high (even infinite) codimension in the set of all possible worlds—and yet we take them for granted. Indeed, the tendency for mental categories to have more predictable internal structure than would arbitrary ones is the common thrust of many attempts to account for the human capacity to categorize, from Hume to Richards’ principle of “Natural Modes” [5, 10, 19]. The modal non-genericities that underlie category interpretations in this paper are, in essence, an instantiation of these ideas in a framework which may be specified and computed concretely.

Acknowledgements

I would like to thank Whitman Richards, as well as Allan Jepson, Michael Leyton, and two anonymous reviewers, for their comments on drafts of this paper. This work was supported in part by AFOSR 89-0504. Correspondence should be directed to the author at E10-120, Dept. of Brain & Cognitive Sciences, M.I.T., Cambridge, MA 02139; or at jacob@psyche.mit.edu.

References