

## Simplicity and complexity in human concept learning

Jacob Feldman

Dept. of Psychology, Center for Cognitive Science  
Rutgers University

It's a pleasure to be here today to speak with you, and I'd like to extend my gratitude to Drs. Lyle Bourne and Linda Bartoshuk for making it possible.

The topic I'd like to talk to you about today is, I think, one of the oldest and most basic in cognition: how we learn from examples. As most famously pointed out by Hume, when we make a finite number of observations of an enduring phenomenon, there is no strictly logical (i.e., deductive) basis for forming any firm generalizations about it. Instead we must "induce," that is, make educated guesses about what its general properties might be. The need for this is especially clear in the case of category learning, or as it is sometimes called, concept formation. We see a few examples of category—say, a straight-backed chair, a plush armchair, and three-legged stool—and must guess the true form of the category (*chairs*).

Categories differ widely, of course, in the ease with which people can learn them from examples. Some categories—*chairs*, say—are easily guessed from few examples. At the other extreme, extremely disjoint categories—say, the set including a hat, a piano, the sun, and the King of

Sweden—are so incoherent and seemingly irregular that it seems *no* finite subset would suffice to communicate the essence of the category; such categories are consequently very difficult to learn from examples. The contents of such a category can only be effectively communicated, it seems, by simply *listing* its contents verbatim: no regularities or common trends hold sway. This idea—that a psychologically incoherent (and thus unlearnable) category is one that cannot be compressed or summarized—will be important later.

This spectrum of subjective complexity is of profound importance in understanding how we learn, because it reflects the underlying mechanisms of induction: what makes some inductive hypotheses—potential concepts—more attractive to the human learner than others. When we have to wrap a concept around the few examples actually seen, we of course prefer the most *natural* concept available. But what exactly is a psychologically “natural” concept?

This question was extensively studied during the 1960s (a period that stretched from about 1953 to about 1973). During this period a great many experiments were conducted concerning the learning of artificial concepts, usually defined by some logical combination of two binary (Boolean) variables. These studies produced some extremely beautiful and stable results, epitomized by the comprehensive studies by Lyle Bourne (summarized in his 1966 book, *Human Conceptual Behavior*). The general thrust of these studies was that the logical form of the rule defining a category can have a decisive effect on the difficulty subjects have in learning it. The most notorious specific conclusion, around which literally thousands of studies revolved, was that *conjunctive* (“and”) concepts were easier to learn—subjectively simpler—than *disjunctive* (“or”) concepts.

Indulging in a bit of retrospective psychoanalysis, it seems that some of the huge amount of interest in this issue derived from the apparent divergence it suggested between *logical* complexity, on the one hand *psychological* complexity, on the other. Conjunction ( $a \wedge b$  in mathematical notation, in which the variables  $a$  and  $b$  refer to features or properties of the objects) and disjunction ( $a \vee b$ ) are of equal complexity by almost any conceivable mathematical definition. Yet the subjective, psychological difference between them was stable and reliable. This contrast seems to suggest some

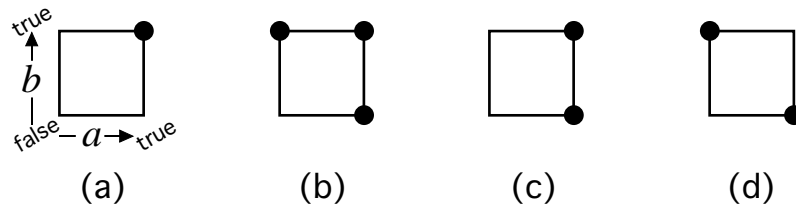


Figure 1. Logical form of some concepts studied in the 1960s as illustrated schematically in abstract Boolean 2-space: (a) conjunction (b) disjunction (c) affirmation (d) exclusive-or.

special, mysterious, and quintessentially *human* bias in favor of disjunctive concepts—an enticing prospect to students of human learning.

In order to fully appreciate the variations in logical form at play here, it is convenient to depict Boolean concepts visually as patterns in a grid of spatial dimensions (Fig. 1). Almost all studies in the 1960s involved bivariate concepts, which we can depict in a two-dimensional square, called Boolean 2-space (top row). Here the horizontal (*a*) and vertical (*b*) axes each refer to a particular Boolean feature, for example size (say, *small* or *large*) and fruit type (say, *apple* or *orange*). The two values of each variable are conventionally referred to as *true* and *false*, though of course these labels are arbitrary and meaningless with certain features (e.g. *apple* vs. *orange*—neither one has any special claim to the role of “true” along the dimension *shape*). The four vertices of the grid thus each refer to a single uniquely-defined object: small apple, large apple, small orange, large orange.

In diagrams like these, a *concept* corresponds to a particular subset of the four vertices, which we depict by heavy dots at the “positive” corners. In this system, conjunctive concepts have one corner positive (Fig. 1a), and disjunctive concepts have three (Fig. 1b). Considering these figures, one immediately sees that there are number of other structural possibilities. One is to have two positives on one side of the square (Fig. 1c); such a concept is called *affirmation* (or *negation*) in the literature, because one of the two features is always true (or false, as the case may be). Another is to have two positives at opposite corners (called *exclusive-or* or *biconditional*; Fig. 1d). All of these varieties were systematically studied by Bourne and others in the 1960s; they differ in difficulty in a very reliable order (usually given as affirmation/negation < conjunction < disjunction < exclusive-

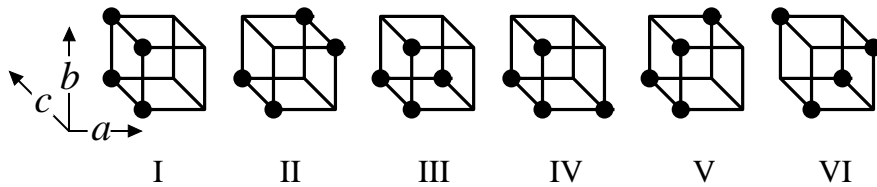


Figure 2. The six concept types studied by Shepard, Hovland, and Jenkins (1961), each of which has four positives defined over three features ( $a$ ,  $b$ , and  $c$ ). These types exhaust the space of possibilities for four positives and three features; every such concept has essentially the same structure as one of these six types. Shepard et al. found that they have a stable difficulty ordering,  $I < II < [III, IV, V] < VI$ .

or/biconditional). A number of researchers, notably Bourne (1974), offered satisfying quantitative accounts of this difficulty ordering.

However, there was one outstanding result that did not fit into this clean bivariate hierarchy: a 1961 study by Shepard, Hovland, and Jenkins, then virtually the only extant research to have considered concepts involving more than two variables. Shepard et al. (1961) considered 3-variable concepts having exactly four positives and four negatives. I'll refer to this as the case with  $D = 3$  and  $P = 4$ , with  $D$  meaning the number of features (dimensions) and  $P$  meaning the number of positive examples. Concepts with  $D = 3$  and  $P = 4$ , it turns out, come in exactly six distinct types (Fig. 2), designated types I – VI by Shepard et al. Mathematically, this typology is complete: no matter how you arrange four vertices in Boolean 3-space, you get something that has essentially the same structure as one of Shepard et al.'s six.

It is important in what follows to understand exactly what is meant here by “essentially the same structure.” As mentioned before, typically, the features here ( $a$ ,  $b$ , and  $c$ ) don't have an intrinsic “sign,” meaning we could just as well interchange the *true* and *false* labels. Similarly, since we are abstracting over the actual meanings of the variables themselves, we don't really care which of  $a$ ,  $b$  or  $c$  is pointing which way in the figures. These choices are really all arbitrary, and so changing them doesn't really change the nature of the concept we are talking about. What this means visually is that we are free to *rotate* and the diagrams rigidly through space (Fig. 3). As long as we don't change a concept's internal structure haven't changed its essential logical form.

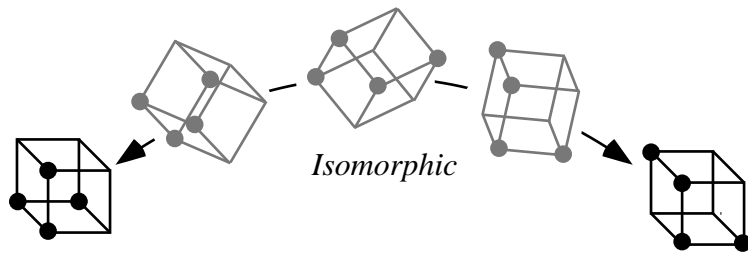


Figure 3. Two Boolean concepts are isomorphic—of the same logical type—if they are equivalent after interchanging the labels and signs of the features, which can be pictured visually as a rigid rotation through Boolean space.

This sort of equivalence is called an *isomorphism* (technically, this particular equivalence is *isomorphism under permutation of feature labels and exchange of sign*). Critically, equivalence under isomorphism carves up the space of possible concepts: we can sort concepts into bins depending on their essential structure, so that all concepts within a bin are essentially the same, but all the bins are essentially different from each other. Shepard et al.’s six types are one such typology—the one appropriate for three featural dimensions and four positives. Other numbers of dimensions or positives lead to other typologies (some of which I’ll get to in a moment). From the point of view of a psychologist interested in the effect of logical form on learning, these typologies give a crucial road map: they tell you exactly what forms to study—namely, the bins, or equivalence classes. Once you understand how each of these bins is treated psychologically, you understand everything that depends on logical form alone.

Note that this notion of equivalence, and the resulting typology, applies to the bivariate case as well, as was recognized in the 1960s. For example, affirmation and negation, though they are different in the details, are really the same *type* of concept, once we decide that we don’t care about which value is labeled positive and which “negative”—both are bivariate concepts in which membership depends only on the value of one variable. Similarly there is really only one kind of concept with one positive, namely conjunction—because all four vertices of the Boolean square are equivalent if we can rotate the square freely. Finally, there is really only one kind of concept with two positives on opposite corners, whether it is labeled either biconditional (when the corners are

oriented north-east and south-west) or exclusive-or (north-west and south-east). In terms of pure logical form, these three types are really all there is.

Now let's get back to the psychology. In their study of the three-features four-positives case, Shepard et al. found that the six essential types exhibited a very reliable difficulty ordering:  $I < II < [III, IV, V] < VI$  (with a tie among III, IV and V). This result has been replicated a number of times since, and has come to be viewed as a standard benchmark in the field. Notice that there is no simple way to explain it in terms of a preference for conjunctive concepts, or any of the other theoretical constructs proposed in the 1960s. More recently, a number of learning models (mostly exemplar models and connectionist networks) have successfully modeled it, but only in terms of the asymptotic behavior of a large and complex simulation with many parameters. There has never been any simple, theoretical account of the ordering: no way of deriving it from first principles of learning.

Almost lost amid the torrent of learning papers in the 1960s were two that proposed a different principle: *simplicity*. Neisser and Weene (1962) and, separately, Haygood (1963) had proposed that the known difficulty ordering of bivariate concepts could be explained in terms of the length of the logical formulas required to express them. For example, exclusive-or requires more symbols to express than conjunction ( $ab + a'b'$  vs.  $ab$ ), and this might be part of why it is more difficult to learn. (Here and from now on I'm adopting the standard mathematician's notation in which  $ab$  means  $a$  and  $b$ ,  $a + b$  means  $a$  or  $b$ , and  $a'$  means not- $a$ .) One big problem with this explanation, though, is that it fails to explain by far the most famous case: conjunction ( $ab$ ) vs. disjunction ( $a + b$ ), which require the same number of symbols (counting only variable names, not the  $+$  or  $'$  symbols, which are operators).

But the idea that simplicity was the overarching principle at work did not gain any adherents. Part of the reason, perhaps, is that to fully work out this idea required a mathematical notion of simplicity that had not yet, at that time, been developed. But at about the same time (around 1962) three mathematicians (Kolmogorov, Chaitin, and Solomonoff) independently developed the

required notion. Their idea, now usually referred to as Kolmogorov complexity, is that the complexity of a symbol string is the length of the *shortest* description that is required to faithfully express it. Simple strings can be very compactly described. Complex strings require longer descriptions. In the limit, if a string is so complex that it can't be compressed at all, one can simply *quote* it verbatim—list its members—as a way of expressing it. By this way of measuring it, then, complexity is intrinsically capped at about the length of the original object: if all else fails—i.e. with maximally complex strings—one can always enumerate their contents, which automatically takes about the same number of symbols as were in the original string itself. Kolmogorov complexity also has certain very desirable mathematical properties, chiefly that it is “universal:” in a certain well-defined sense the complexity of a string is independent from the details of the language in which you choose to express it.

In the realm of Boolean concepts, the natural analog of Kolmogorov complexity is what's called the *Boolean complexity*: the length of the shortest logical expression that is equivalent to the set of positive examples—called the *minimal formula*. The more you can reduce the logical expression of your concept, while still faithfully expressing it, the simpler the concept is. Normally we measure the length of the minimal formula by counting variable symbols only, not operators. For example, the concept *big apple or small apple* (expressed symbolically as  $ab + ab'$ , with  $a$  = apple and  $b$  = big) is logically equivalent to *apple* (i.e.,  $ab + ab'$  reduces algebraically to  $a$ ); it has Boolean complexity 1. Conversely the concept *big apple or small orange* ( $ab + a'b'$ ) can't be similarly reduced—no shorter formula is equivalent to it—so it has Boolean complexity 4. The same reduction trick can be applied to any Boolean concept, of any length, to give an estimate of its intrinsic complexity.

So how does Boolean complexity match up to the subjective difficulty of concepts—that is, to subjective complexity? The touchstone is Shepard et al.'s six types, whose difficulty ordering, you'll recall, was a bit tricky to explain. The Boolean complexities of the six types come out as 1, 4, 6, 6, and 10 respectively (see Fig. 4)—perfectly agreeing with the famous difficulty ordering

$I < II < [III, IV, V] < VI$ . The exact minimal formulae for each of the six types are given in the figure (lower panel).

This agreement gives a good prima facie boost to the theory that Boolean complexity dictates subjective complexity. A more comprehensive test, though, would require trying new cases. Bivariate cases were exhaustively studied in the 1960s, and Shepard et al. completely covered the  $D = 3, P = 4$  case. What about other values of  $D$  and  $P$ ? These had never been tested. The aim of my study (Feldman, 2000) was to do so.

The first step is to work out the typologies for other values of  $D$  and  $P$ . For each value of  $D$  and  $P$ , the typology completely changes, producing a different “family” of basic concept types, referred to as the  $D[P]$  family. The sizes of these families as you change  $D$  and  $P$  vary in a somewhat unpredictable way (for example,  $4[4]$  has 19 types—who would have guessed?)—though the correct combinatoric theory was worked out as early as 1951 (by Aiken and his staff at the Harvard computation laboratory; see Feldman, in press). When we need to refer to particular types within each family, we designate them with Roman numerals subscripted with the family name, like  $I_{3[4]}$ ,  $II_{3[4]}$ , etc., for Shepard et al.’s family.

For my study, I wanted to be as comprehensive as possible, but above  $4[4]$  there are simply too many types in each family to test conveniently; in the end I used families  $3[2]$ ,  $3[3]$ ,  $3[4]$ ,  $4[2]$ ,  $4[3]$ , and  $4[4]$ . This means that, up to isomorphism, the study considered *every* Boolean concept with three or four features and up to four positives. Fig. 4 shows the  $3[2]$ ,  $3[3]$ , and  $3[4]$  families, along with both the raw and minimal formula associated with each concept in each family. As you can see in the figure, just as with Shepard et al.’s six types (the  $3[4]$  family), each of the  $3[2]$ -family concepts is distinct from each of the other ones—they can’t be rotated through 3-space to become equivalent; and likewise for each of the  $3[3]$  concepts. (Of course, concepts from different families are never isomorphic to each other either, since they have different numbers of dimensions or positive vertices.) The  $3[2]$  and  $3[3]$  families each have three concepts, none of which as far as I know had every been tested psychologically before.















	Concept	Raw formula	Minimal formula	Complexity
3[2]	I 	$a'b'c' + a'b'c$	$a'b'$	2
	II 	$a'b'c' + a'bc$	$a'(b'c'+bc)$	5
	III 	$a'b'c' + abc$	$a'b'c'+abc$	6
<hr/>				
3[3]	I 	$a'b'c' + a'b'c + a'bc'$	$a'(bc)'$	3
	II 	$a'b'c' + a'b'c + abc'$	$a'b'+abc'$	5
	III 	$a'b'c' + a'bc + ab'c$	$a'(b'c'+bc) + ab'c$	8
<hr/>				
3[4]	I 	$a'b'c' + a'b'c + a'bc' + a'bc$	$a'$	1
	II 	$a'b'c' + a'b'c + abc' + abc$	$ab+a'b'$	4
	III 	$a'b'c' + a'b'c + a'bc' + ab'c$	$a'(bc)'+ab'c$	6
	IV 	$a'b'c' + a'b'c + a'bc' + ab'c'$	$a'(bc)'+ab'c'$	6
	V 	$a'b'c' + a'b'c + a'bc' + abc$	$a'(bc)'+abc$	6
	VI 	$a'b'c' + a'bc + ab'c + abc'$	$a(b'c+bc') + a'(b'c'+bc)$	10

Figure 4. Concepts from three of the families tested (3[2], 3[3], and 3[4]), showing the “raw” (uncompressed) formula, minimal formula, and Boolean complexity (length of the minimal formula in literals).

There is one additional complication before the study can be completely laid out. With Shepard et al.'s study there were four positives drawn from the eight objects, and thus also four negatives. But most of the families I was interested in have *different* numbers of positives and negatives; for example 3[2] concepts each have two positives and six negatives. Each such concept has a “mirror image” concept in which the labels are swapped, so that the positives become negatives and the negatives positives. I'll refer to the “orientation” of each concept in this sense as its *parity*, designating the version with the smaller half labeled positive as the *Up* version and the mirror image as the *Down* version. What is interesting about this is that, from a logical point of view, Up and Down versions of a concept are essentially the same concept—specifically, the Down version corresponds to the same formula as the Up version with an extra “not” sign placed in front of it. Operator symbols don't count towards the complexity, so Up and Down versions of the same concept *always* have the same complexity. They differ only in this parity, which is a new variable orthogonal to complexity.

What is the psychological significance of a concept's parity? Will Up and Down versions of concepts be treated identically by subjects? In the experiments, it is very straightforward to test this: we simply include every concept in both Up and Down versions, and treat the parity factor as an independent manipulation fully crossed with Boolean complexity. (Actually, they aren't quite crossed, because 3[4] concepts only come in one version, because they have equal numbers of positives and negatives.) For simplicity of notation, from here on I will use  $P$  to mean the number of examples in the “smaller half” of the concept, which by definition is positive in the Up version and negative in the Down version.

In testing all these concepts, I was primarily interested in the ease subjects had in learning their members. So the procedure was simply to show the subject the entire space of objects (8 for  $D = 3$  cases, 16 for  $D = 4$ ), separated into positive and negative groups, with the positive group labeled “Examples” and the negative “Not examples” (see Fig. 5). First, the subject would study these for a fixed period of time (a few seconds, the exact period depending the case). Then, the subject was tested on all the objects in random order. We can then look at their performance (percent

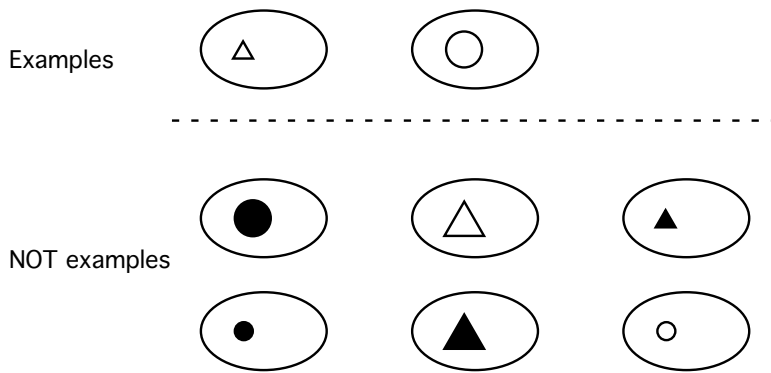


Figure 5. A sample “learning” screen as viewed by subjects. This screen shows a concept of type  $\Pi_{3[2]}$ .

correct) as a function of the structure of the concept—specifically, of its Boolean complexity and its parity. Each subject saw concepts from only one  $D[P]$  family, including all concepts from the family in both parities. Thus the manipulations of complexity and parity were both within-subjects, while family was between-subjects.

The results are summarized in Fig. 6, which collapses over all the families to highlight the effects of Boolean complexity and parity. Both factors plainly and systematically influence performance. As complexity increases, performance worsens steadily; more complex concepts are more difficult to learn than less complex cases, with a roughly constant advantage for Up cases over Down cases.

The main result—the complexity affect—suggests that human learners are doing something like *minimization* or *compression* when they represent concepts. As the subject studies the set of examples, he or she seeks to encode them in as compact a manner as possible. The more effectively the examples can be compressed—the lower the Boolean complexity—the more successful this strategy will be, and the more effectively the examples will be retained.

This result is especially intriguing because it ties human concept learning to a very famous and ubiquitous principle—simplicity: the idea that observers ought to favor simple hypotheses. In the philosophy of science literature this is known as Occam’s razor, but the same idea turns up in a multitude of settings in other fields concerned with inference from examples—including

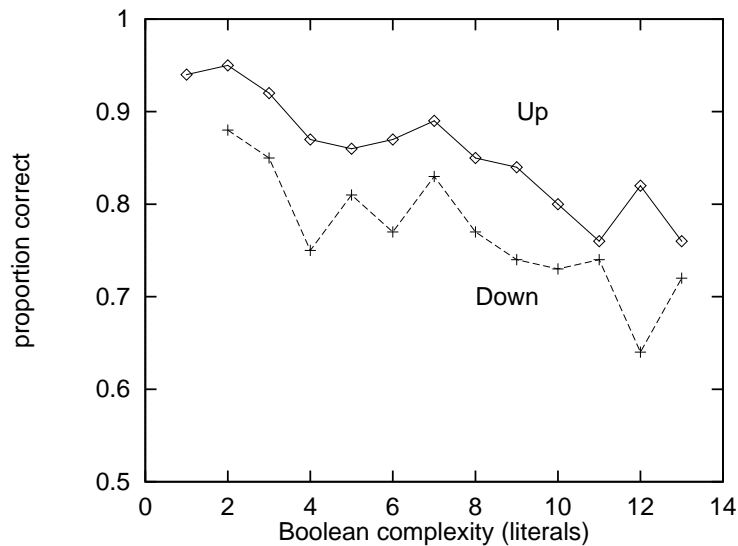


Figure 6. Results of the experiments. As Boolean complexity increases, performance steadily declines, with a roughly constant advantage for Up parity cases over Down parity cases.

machine learning and inferential statistics, where it is often called the Minimum Description Length principle, a term introduced by Rissanen (1978). This principle has been growing in influence during the past decade, where it lies at the heart of many of the most sophisticated and successful automated inference systems. In perception, the same idea is familiar in the guise of the Minimum principle, or in the Gestaltists' term *Prägnanz*. All these principles point in the same direction: observers profit by drawing the simplest interpretation available of what they observe. Surprisingly, this idea had never really penetrated the field of human concept learning (except for Neisser, Weene, and Haygood's doomed hypothesis) despite this seeming like a pretty apt place to apply it. But the data in Fig. 6 suggest that human category learners, too, obey a minimization principle.

The other main result, the parity effect, suggests that subjects have some kind of complexity-independent preference for looking at concepts through their positive examples. Indeed others had noticed the same tendency as early as 1953 (e.g., Hovland & Weiss, 1953); so this is not really news. The novelty here is to see this as a factor orthogonal to complexity, because seeing it this way changes the way you see older results—specifically the old conjunction/disjunction dichotomy.

Recall the typology of bivariate logical I discussed before. The way it ran, there turned out to

be only three essential types: affirmation, conjunction, and exclusive-or. So where does *disjunction* fit into this scheme? The answer is that disjunction is the same type as conjunction, except with opposite parity: disjunction is conjunction “upside-down.” A conjunctive concept such as  $ab$  has one positive example ( $ab$ ); its complement, in this case the concept  $(ab)'$ , has three positive examples ( $ab'$ ,  $a'b$ , and  $a'b'$ ), and can be rewritten as  $a' + b'$ —a disjunctive concept. Thus conjunction is the Up parity version of this basic type and disjunction is the Down parity version. The point is that the famous superiority of conjunctive over disjunctive concepts was just a reflection of the parity effect—Up parity cases are learned more easily than Down parity cases—and had nothing to do with complexity. But when a wider range of concept types is tested, as in this experiment, we see that there *is* a substantial complexity effect, even though it doesn't show up in the conjunction/disjunction comparison. Contrary to how it must have looked focusing on only that comparison, a lot of the variance in conceptual difficulty is driven by differences in complexity.

I think the idea of complexity minimization also sheds some light on a more recent controversy in the field of concept learning: the distinction between the encoding of rules vs. the storage of exemplars. Many recent concept learning theories have revolved around the explicit storage of specific examples: in such theories, new objects are evaluated by comparing them with stored exemplars. By contrast, many areas of learning, such as the acquisition of language, quite obviously involve the extraction of rules from what is heard (i.e. descriptive grammatical rules, such as how to form the past tense, what order to place adjectives in, which part of a sentence receives tense, etc.). Recently several theories have been proposed that mix these two strategies for learning: one component for extracting rules, and another component for storing examples that don't fit into the rule scheme (exceptions).

But the distinction between the “rule” and the “exceptions” gets a little hazy when one thinks about minimal formulae. Some concepts, like Shepard et al.'s type I (see Fig. 2), reduce to one very simple rule that covers all objects. Others, like type 3 of family 3[2] (see Fig. ??), are completely incompressible, which means their minimal formulae consist essentially of a verbatim list of

their members. But in between these extremes are some concepts whose minimal formulae have a component (literally, a disjunct) that covers most of the objects, plus one or more additional objects (again, more disjuncts) that *aren't* covered by the “main rule.” An example is type VII<sub>4[4]</sub>, whose uncompressed rendition is

$$a'b'c'd' + a'b'c'd + a'b'cd' + abcd,$$

which compresses to

$$a'b'(cd)' + abcd.$$

The first disjunct ( $a'b'(cd)'$ ) is the “rule,” and the second disjunct ( $abcd$ ), a single object, is the “exception.” But the exception is also in a sense *part* of the rule—part of the minimal formula.

In a sense, this makes the dichotomy between rules and exceptions a bit fuzzy. But alternately one can view this situation as clarifying the distinction between rules and exceptions, by showing exactly what the exceptions are exceptions *to*. When some objects need to be explicitly listed as part of the most compact rendition of the category, then you know in a deep sense that they are really exceptional. Compression points the way to a more rigorous basis for the distinction between rule and exception.

I'd like to end with a very superficial reflection about the underlying reasons for complexity minimization in concept learning. The principle of simplicity is so familiar that one hardly stops to wonder why it makes any sense. But why should we try to reduce a set of observed examples to a minimal form? What advantage does this afford us? Is it just the saving in storage space? To me, this saving seems a bit trivial in the context of a brain with  $10^{11}$  neurons.

I'd suggest—echoing an enormous amount of technical advances in statistics and machine learning—that minimization of complexity subserves the more basic goal of *extraction of regularities*. Our deepest cognitive impulse is to understand the world. And in a profound mathematical sense, compressing our description of it helps to accomplish this. How? Because all compression

schemes depend on finding and benefiting from regular tendencies in the data—places where the data is a bit redundant or repetitive or orderly. This is how a photograph with large uniform areas or repetitive textures can result in a very small file on your hard drive. (The image compression scheme known as JPEG—like all compression schemes—is based around a clever and systematic use of this idea.) In file-format compression schemes, an understanding of the form of regularities in the data is leveraged into a method for reducing file sizes. The flip side of this idea is that compressing data can be leveraged into an understanding of what regularities the data obeys—it is how regularities implicit in the data become *explicit*.

In the realm of concepts defined over Boolean features, this idea is particularly transparent. Take the concept  $ab + ab'$  discussed before (e.g., *big apple or small apple*). Compressing this algebraically to  $a$  (*apple*) makes explicit that all objects in this concept are apples—a regularity of this small world. By compressing the formula, the observer has discovered a grain of truth about this world. Ideally, one would like a theory of category formation that was organized around this principle, in order to fully understand how categorization relates to inference. In other more recent work, I have attempted to develop this idea into a more thorough-going “concept algebra,” modeling in more detail how human observers extract regular tendencies from the examples they observe.

The American Mathematical Society, in its newsletter, summarized the Boolean complexity result as *incompressible is incomprehensible*—an elegant phrase I wish I’d thought of first. But I’d rather turn it around: in a sense, *compressible is comprehensible*, or, perhaps, *compression is comprehension*. Minimization and inference are deeply intertwined, and I think one of the major challenges of psychology now is to understand how and why.

### References

- Aiken, H. H., & the Staff of the Computation Laboratory at Harvard University. (1951). *Synthesis of electronic computing and control circuits*. Cambridge: Harvard University Press.
- Bourne, L. E. (1966). *Human conceptual behavior*. Boston: Allyn and Bacon.
- Bourne, L. E. (1974). An inference model for conceptual rule learning. In R. Solso (Ed.), *Theories in cognitive psychology* (pp. 231–256). Washington: Erlbaum.

- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, 407, 630–633.
- Feldman, J. (in press). A catalog of Boolean concepts. *Journal of Mathematical Psychology*.
- Haygood, R. C. (1963). *Rule and attribute learning as aspects of conceptual behavior*. Unpublished doctoral dissertation, University of Utah.
- Hovland, C. I., & Weiss, W. (1953). Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology*, 45(3), 175–182.
- Neisser, U., & Weene, P. (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology*, 64(6), 640–645.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14, 465-471.
- Shepard, R., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75(13), 1–42.