Journal of Mathematical Psychology 53 (2009) 203–221

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Categorical invariance and structural complexity in human concept learning

Ronaldo Vigo *

Cognitive Science Department, Indiana University at Bloomington, Bloomington, IN, 47401, USA
Psychology Department, Ohio University, Athens, OH, 45701, USA

ARTICLE INFO

Article history:
Received 26 June 2007
Received in revised form
12 March 2009
Available online 10 June 2009

Keywords:
Concept learning
Categorization
Rule-based classification
Logical manifold
Categorical invariance
Logical invariance
Structural complexity
Boolean complexity
Invariance
Complexity
Concepts

ABSTRACT

An alternative account of human concept learning based on an invariance measure of the categorical stimulus is proposed. The categorical invariance model (CIM) characterizes the degree of structural complexity of a Boolean category as a function of its inherent degree of invariance and its cardinality or size. To do this we introduce a mathematical framework based on the notion of a Boolean differential operator on Boolean categories that generates the degrees of invariance (i.e., logical manifold) of the category in respect to its dimensions. Using this framework, we propose that the structural complexity of a Boolean category is indirectly proportional to its degree of categorical invariance and directly proportional to its cardinality or size. Consequently, complexity and invariance notions are formally unified to account for concept learning difficulty. Beyond developing the above unifying mathematical framework, the CIM is significant in that: (1) it precisely predicts the key learning difficulty ordering of the SHJ [Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42] Boolean category types consisting of three binary dimensions and four positive examples; (2) it is, in general, a good quantitative predictor of the degree of learning difficulty of a large class of categories (in particular, the 41 category types studied by Feldman [Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633]); (3) it is, in general, a good quantitative predictor of parity effects for this large class of categories; (4) it does all of the above without free parameters; and (5) it is cognitively plausible (e.g., cognitively tractable).

1. Introduction

1.1. Invariance

Invariance principles are ubiquitous in the physical sciences. A simple example of one such principle in physics involves the lines of magnetic force. These lines form closed loops that have the property that for any volume enclosing them the number of lines entering the volume is equal to the number of lines leaving the volume. This property is sometimes referred to as the property of incompressibility and it remains regardless of the way that we alter the space defined by the volume. That is, the property is invariant in respect to deformations of the space such as stretching, twisting, cutting, or bending.

Another example of invariance-based principles in physics involves both the theory of Special Relativity and General Relativity. Both theories aim to identify under what space–time coordinate transformations the laws of physics remain invariant. Lastly, topology, one of the most vital fields of mathematics, is roughly defined as the study of those properties of figures that are invariant under continuous transformations that have a continuous inverse.

In a similar spirit, cognitive scientists have resorted to invariance principles in order to explain a variety of cognitive phenomena such as memory, perception, and concept learning. For example, in pioneering work, Garner (1963, 1970) and Garner and Felfoldy (1970) developed the idea that the degree of invariance of a form determines how good a pattern is perceived to be. More specifically, Garner (1974) explores the relationship between what he calls "goodness of pattern" to the concept of information redundancy in the form of regularities in the stimulus structure.

Under this approach, structural regularities are characterized in terms of the size of mutually exclusive subsets of the set of visual stimuli. These subsets are generated by subjects asked to group together visual patterns under a similarity criterion. A good model of the generated subsets is to regard them as equivalence classes of patterns that can be generated from each other by a combination of rotations and reflections. Garner (1963) showed that the size of these equivalence classes is a good predictor of goodness of pattern judgments: for example, the smaller the size of a subset of visual patterns that are judged similar (i.e., the greater its redundancy and regularity), the higher the degree of pattern goodness that will be assigned to a visual pattern within the subset.

Like Garner, Leyton (1992) in his book "Symmetry, Causality, and Mind" uses invariance principles to explain cognitive phenom-

* Corresponding address: SCOPE Lab, 337 Porter Hall, Ohio University, Athens, OH 45701, USA.
E-mail addresses: rvigo@indiana.edu, vigo@ohio.edu.

0022-2496/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2009.04.009

ena. However, this time the phenomena in question are memory processes. Leyton posits that the mind assigns to any shape a causal history explaining how the shape was formed. By examining shape in terms of its inherent symmetries, he argues that symmetry is a key component to cognitive processing. Thus, he is able to reduce aspects of visual perception to symmetry or invariance principles.

In tune with this idea, Imai (1977) and more recently Hahn, Chater, and Richardson (2003) proposed that similarity judgments may be determined by transformational distance. More specifically, the degree of similarity between two objects is characterized by these researchers in terms of the number of transformations necessary to convert one into the other. Hahn et al. (2003) give experimental evidence that seems to show that transformational distance strongly influences similarity judgments. Interestingly, unlike the previously discussed theories of invariance, rather than ascertaining under what transformations an object's structure remains invariant, Hahn and associates ask how many (and what type of) relevant transformations are necessary to achieve a kind of deterministic variance.

In our theory, we follow this tradition of using invariance as an explanatory principle for cognition, but we do so from an entirely different perspective and aim. We wish to introduce an elegant and natural mathematical framework for the study of invariance in human concept learning in terms of the languages of logic and analysis. This approach is rather natural since, as we shall see, the concepts we will focus on (i.e., Boolean concepts) are concepts learned from categories that are plainly definable in terms of logical rules or logical expressions from Boolean algebra in the first place. We focus on Boolean categories since these are the simplest nontrivial cases we can study. In fact, Boolean categories also play an important role in the landmark theories of human concept learning that we shall use as benchmarks for our own.

Two of these landmark theories, the exemplar theory (Medin & Schaffer, 1978; Nosofsky, 1986) and prototype theory (Rosch & Mervis, 1975), may be characterized as mental process theories based on similarity assessment representations. Under prototype accounts of human concept learning concepts are prototypical representations or central tendencies of the exemplars associated with the concept. A novel stimulus is then classified on the basis of its similarity to the prototype. On the other hand, under exemplar accounts of concept learning (Medin & Schaffer, 1978; Nosofsky, 1986), a novel stimulus is classified by determining how similar it is to the stored exemplars of a category and those of a contrasting category.

For example, Nosofsky's Generalized Context Model (GCM) (1984, 1986) is a formal generalization of the exemplar theory of Medin and Schaffer (1978). The model owes a great deal to the theory of multidimensional scaling (Shepard, 1962, 1974) and to Luce's choice probability theory (1959). The GCM is based on three key ideas. The first of these is the characterization of subjective similarity as an inverse exponential function of the Euclidean distance between stimuli in some psychological space as described by Shepard's universal law of generalization (Shepard, 1987).

The second is Luce's classic Similarity Choice Model (SCM) for predicting identification responses (Luce, 1963). In Luce's model, the probability that a stimulus i elicits a response j in an identification experiment is given by

$$p(r_j|s_i) = \frac{b_j s(i,j)}{\sum_{k=1}^{n} b_k s(i,k)} \tag{1.1}$$

where b_j and b_k are parameters indicating response biases, $s(i,j)$ is a similarity measure between the stimuli i and j, and the index k in the denominator ranges over the set of stimuli that are eligible as responses in the experiment.

The third key idea in the GCM is the inclusion of a parameter w_i representing the selective attention allocated to dimension i. The distance between stimuli in some psychological space is then defined as $\boldsymbol{d}(x,y) = \sum_i w_i \cdot |x_i - y_i|$ for separable (selectively attendable) dimensions. In the GCM the probability that category $\boldsymbol{C}$ will be selected given stimulus y is

$$p(\boldsymbol{C}|y) = \frac{b_{\boldsymbol{C}} \cdot \sum_{x \in \boldsymbol{C}} F_{\boldsymbol{C}_x} \cdot s(x,y)}{\sum_{\alpha} \left[b_\alpha \cdot \sum_{k \in \alpha} F\alpha_k \cdot s(k,y) \right]}. \tag{1.2}$$

This is the ratio between: (1) aggregate similarities of the stimulus y to the exemplars in $\boldsymbol{C}$, and (2) the sum of the aggregate similarities of the stimulus y to the members of the contrasting categories α. The parameter b_α is the response bias for category α and $F\alpha_k$ is a parameter indicating the frequency of occurrences of the exemplar k in category α (Kruschke, 2006). Thus, the processes of similarity, attention, and choice are combined to explain and predict the likelihood of a correct categorization decision, making the GCM a probability-based process account of classification behavior.

This and other exemplar accounts have great intuitive and parsimonious appeal and have led to connectionist models that have been successful at quantitatively predicting degrees of concept learning difficulty. One of the most successful of these models is ALCOVE (Kruschke, 1992)—a connectionist model of exemplar-based categorization inspired by Nosofsky's Generalized Context Model (1986). The dynamic adaptive nature of these process theories stands in sharp contrast to the "static" structural accounts of human concept learning discussed above: these aim to determine the degree of concept learning difficulty based solely on the structural properties of the categorical stimulus (Garner, 1974) and do not make assumptions about what is possibly taking place in the human mind before categorical decisions are made and while concepts are being learned or formed.

Indeed, in our *analytico–structural* account of Boolean concept learning, we will introduce a measure of the structural characteristics of the stimulus that are good predictors of the degree of learning difficulty experienced when learning it. As discussed, our stimulus-driven modeling paradigm is not very different from that of psychophysics, which examines what Gibson (1966) refers to as the energetic properties of the stimulus (its physical properties) as opposed to its informational properties as reflected in its structure.

1.2. Cognitive complexity

Most structural accounts do not have at their core invariance notions. Ever since the seminal work of Miller (1956) which sought to identify the limits of human short-term memory capacity, there has been much interest on how the complexity of a task or of a stimulus may influence our ability to perceive it, memorize it, learn it, and make decisions and inferences about it. We refer to this property as the cognitive complexity of the stimulus. An example of this idea in applied domains is found in the work of Wang and Shao (2003) on program design. These researchers proposed a measure of software code complexity based on cognitive weights. Cognitive weights are numbers representing the degree of difficulty for comprehending particular types of basic control structures in programming code. These weights combine to give an overall sense of the clarity, effectiveness, and correctness of the software.

Accordingly, the idea of using complexity measures for modeling cognition has been reexamined in the work of Feldman (2000) and Vigo (2006). Feldman (2000) proposes a structural model based on a measure of complexity. The complexity of a Boolean category is defined by Feldman as the length of the

shortest propositional formula that is logically equivalent to the original full DNF formula describing the category type.

Under this approach lie the common suppositions that: (1) there are structural properties intrinsic to a stimulus that can be used to describe how complicated the stimulus is, and (2) that such properties are good predictors of the learning difficulty of the stimulus. Our own model will offer an alternative notion of complexity that we shall refer to as "structural complexity" and that will be based on invariance principles.

More specifically, in the theory we introduce in Sections 3–5 the structural complexity of a categorical stimulus will be a function of both its inherent degree of invariance and its cardinality (size). This definition in turn will be used to predict the degree of learning difficulty associated with each categorical stimulus. In general, our approach may be viewed as a synthesis of invariance and complexity notions. Thus, on average, categories with the greatest degree of categorical invariance – and hence, a relative lowest degree of structural complexity – should be easiest to learn. This is because invariance tells us the way in which objects in a category are interrelated at a higher level.

We shall discuss Feldman's *m*-complexity model (i.e., minimization-based complexity model) in detail in Section 3 in order to bring to the fore some of the problems that our alternative invariance-based model of complexity will be able to solve. We shall also discuss briefly a more recent structural account by Feldman (2006) based on the idea of the spectral decompositions of Boolean categories.

1.3. Invariance, complexity, and process

Although the structural properties of a Boolean category may help us predict its degree of learning difficulty, we are left with the question as to how these structural properties influence learnability. In other words, what is the connection between the concept learning process and the structural properties of the categorical stimulus? To answer this question we propose a companion computational theory hypothesis to our structural account. In Sections 4 and 5, as we present our formal model, we shall fill in the missing details.

Under our hypothesis, it is invariance that determines the degree of perceived simplicity of the stimulus and it is excessive demands on working memory capacity during the invariance detection process that determine the degree of subjective learning difficulty. Intuitively, this makes sense. The more pattern perceived in a structure, the simpler it seems to be: the more tractable. Invariance plays the role of simplicity while category size plays the role of raw complexity in our model. However, these two properties act together in synergy to determine the overall structural complexity of a category.

In contrast, Garner (1974) believes that it is subset size (see the discussion above on Garner), rather than symmetry that is the pertinent factor in determining cognitive performance. He writes on page 17 (Garner, 1974): "If symmetry is so directly related to pattern goodness, why don't we just say that symmetry is the pertinent factor, rather than subset size? The answer lies partly in the fact that symmetry is a sufficient but not necessary condition for producing small subset sizes, even with these simple stimuli. Symmetry is simply one way in which the stimulus may be manipulated to produce variations in subset size, but its mode of action is via subset size rather than by symmetry per se."

This conclusion by Garner may have been the result of the absence of a function in his formal framework that could meaningfully measure the degree of total invariance of each stimulus independent of the equivalence class or subset it is grouped in. With such a measure at our disposal, we believe that symmetry information is necessary to make predictions about goodness of pattern judgments. Notwithstanding, it will be shown that in respect to concept learning, both the degree of symmetry and category size are intimately tied up and play an important and synergistic role in determining structural complexity and ultimately concept learning difficulty.

To derive an invariance measure of the stimulus we introduce a mathematical framework based in part on what is known in Boolean circuit theory as the Boolean differential operator. The framework introduces the notion of the *logical manifold* operator of a Boolean category in order to: (1) generate the degrees of invariance or symmetry of a category in respect to its dimensions, and (2) generate the space of degrees of symmetry of a Boolean category family (we shall describe these ideas in detail in Section 3). Consequently, it is proposed that the category types within a category family can be naturally ordered by both the maximal invariance value and the frequency of the maximal invariance value of their logical manifolds. This ordering points the way toward a method for meaningfully measuring degrees of "global" invariance.

The method involves measuring the Euclidean distance of the logical manifolds of the category types from a zero origin. This is significant since distance metrics have proven unsuccessful in making concept learning predictions in discrete feature domains. For example, Lee and Navarro (2002) found that ALCOVE (Kruschke, 1992), perhaps the most successful exemplar model, could not fit discrete-feature categories well without a metric sensitive to the featural structure of the space: as will be seen, in our approach, invariance is defined in terms of an ordinary metric that, by virtue of being applied to points in invariance space, is sensitive to feature structure.

Although our model is structural in nature, we can reasonably speculate whether the mathematical operators that compute the invariance of the categorical stimulus may also be construed as functional descriptions of an information processing system, and therefore as the basis for a theory of computation. To understand this point, it is helpful to consider Marr's three levels of description of an information processing system (Marr, 1982; Poggio, 1981): the computational or behavioral level of description (e.g., the functional description or the description of what the device does), the algorithmic and representational level of description (e.g., the description of the representation of acquired information and of the algorithms defined over such representation), and the physical implementation level of description (i.e., the description of the physical device that is capable of actualizing what is described by the first two levels).

More specifically, in Marr's work on vision, the computational level of description specifies the mathematical operations that facilitate the goal of a theory of vision: namely, to construct a three-dimensional representation of distal stimuli on the basis of inputs to the retina. For example, Marr introduces a differential operator (i.e., the Laplacian) capable of detecting intensity changes in light energy at an initial stage of processing as the basis of his computational level of description (Poggio, 1981).

Likewise, if we take our conceptual system's ultimate goal to be that of representing multiplicity of objects holistically, then the mathematical operations that determine the degree of invariance of a categorical stimulus may be understood as operations that facilitate such a goal. As such, our theory plays the role of a "computational theory" in Marr's sense of the term. This stands in stark contrast to the idea that our conceptual system is best described in terms of interacting mental processes or capacities (an idea consistent with Marr's representation and algorithm level of description). Henceforth, in order to make more intuitive the role that invariance principles may play in cognition, we shall emphasize the connection between our structural model and the aforementioned computational level of description.

On a different note, another significant and unique aspect of our logico–analytic structural model will be that it does not depend on choice probabilities nor on multiple parameters as is the case with the mental process models discussed in the previous section. This is significant since most cognitive models are of the statistical and probabilistic variety. Our successful excursion into non-probabilistic domains can only help expand the cognitive modeling horizon.

The outline for the remainder of this treatise is as follows. First, we will give a brief introduction to Boolean categories, the classification of these in terms of typed families, and Feldman's two models of Boolean complexity. We then shall discuss some of the challenges facing the minimization-based complexity model as a prelude to our theory. In Section 3, we will introduce our mathematical framework for economically characterizing and identifying degrees of invariances or symmetries in Boolean categories. This will be followed by our characterization of structural complexity in terms of degrees of invariance.

The core idea underlying our discussion will be that an *invariance-based* measure of structural complexity is a good predictor of Boolean concept learning difficulty; and more importantly, that invariance, and not complexity, is the fundamental principle underlying human concept learning. In other words, since invariance determines structural complexity, it also determines and explains why some concepts are more difficult to learn than others. This assumption will be referred to as the categorical invariance theory of concept learning or "CIT" and its formal description will be referred to as the categorical invariance model of concept learning or "CIM".

Lastly, in Section 5 we will examine how well the CIM, without free parameters, predicts the difficulty ordering of sets of categories of the type studied by Shepard, Hovland, and Jenkins (1961) consisting of three binary dimensions and four positive examples (i.e., 3(4) type categories), and more generally for the set of 41 category types studied by Feldman (2000). This is followed by a discussion of parity effects and the predictions made by our model when the 35 additional category types in down parity are included in our original set of 41 (76 category types in total). The paper ends in Section 6 where we summarize our results and discuss a list of open problems and research directions.

2. Boolean categories and complexity

2.1. Formal representation and notational preliminaries

Since we shall define Boolean categories using the language of Boolean algebra, we begin with a definition of a Boolean algebra and a Boolean expression (or formula). What follows should be regarded as a semiformal introduction to a few key concepts from Boolean algebra that are useful in the development of our categorical invariance model. For a more rigorous and detailed introduction to Boolean algebra, including its model theoretic details, the reader is referred to Mendelson (1970, 1979).

Definition 2.1. A Boolean Algebra is a sextuple $\langle \mathcal{B}, 0, 1, +, \cdot, ' \rangle$ consisting of an arbitrary (carrier) set $\mathcal{B}$ together with two (distinct) elements 0 and 1, a negation operator $'$, and two binary operations $+$ and $\cdot$ such that for all elements x, y, and z of $\mathcal{B}$, the following axioms hold: Commutativity: (1a) $x + y = y + x$, (1b) $x \cdot y = y \cdot x$; Associativity: (2a) $x + (y + z) = (x + y) + z$, (2b) $x \cdot (y \cdot z) = (x \cdot y) \cdot z$; Distributivity: (3a) $(x+y) \cdot z = (x \cdot z) + (y \cdot z)$, (3b) $(x \cdot y) + z = (x + z)(y + z)$; Indempotency: (4a) $x + x = x$, (4b) $x \cdot x = x$; Identity: (5a) $x+0 = x$, (5b) $x+1 = 1$, (5c) $x \cdot 0 = 0$, (5d) $x \cdot 1 = 1$; Negation: (6a) $(x')' = x$, (6b) $0' = 1$, (6c) $1' = 0$; and De Morgan: (7a) $(x + y)' = x' \cdot y'$, (7b) $(x \cdot y)' = x' + y'$.

We now define recursively the class of Boolean expressions. This is important since Boolean expressions (or Boolean formulae) will be used to define Boolean categories.

Definition 2.2. The class $\mathcal{F}$ of Boolean expressions (or Boolean formulae) is defined recursively as follows (where n is an arbitrary positive integer):

(1) The Boolean variables a, b, c, d, w, x, y, z, A, B, C, D, W, X, Y, Z, $x_1, \ldots, x_n$, $y_1, \ldots, y_n$, $z_1, \ldots, z_n$, and the constants 0 and 1 are elements of $\mathcal{F}$.

(2) If $\varphi \in \mathcal{F}$ and $\psi \in \mathcal{F}$ then $(\varphi + \psi) \in \mathcal{F}$, $(\varphi \cdot \psi) \in \mathcal{F}$, and $\varphi' \in \mathcal{F}$.

(3) Nothing else is an element of $\mathcal{F}$.

The Boolean variables above are variables with 1, 0, and elements of the carrier set $\mathcal{B}$ as their possible values. To simplify our exposition, we shall adopt certain notational conventions regarding the class $\mathcal{F}$ of Boolean expressions defined above. First, whenever there is no ambiguity, we shall drop parentheses from Boolean expressions. Also, following the notational conventions of the Boolean algebra of propositional logic, we will use the symbol "$\sim$" interchangeably with "$'$", the symbol "$\vee$" interchangeably with the symbol "$+$", and the symbol "$\wedge$" interchangeably with the symbol "$\cdot$". Finally, if φ and ψ are Boolean expressions, we shall abbreviate $\varphi \wedge \psi$ and $\varphi \cdot \psi$ with $\varphi\psi$. Boolean expressions and Boolean functions are intimately connected as may be seen from the following.

Definition 2.3. Let $\mathcal{B}$ be the carrier set of a Boolean algebra $\mathbb{B}$ and let $B = \mathcal{B} \cup \{0, 1\}$. An n-vary Boolean function F^n of $\mathbb{B}$ is a mapping from the cross product $\prod_{i=1}^{n} B_i$ to B (where i and n are positive integers and for any i, $B_i = \mathcal{B} \cup \{0, 1\}$). The value of the variable n is referred to as the degree of the function.

It is well known that every Boolean expression determines a Boolean function (for details see Church, 1956; Mendelson, 1979). Given a Boolean algebra $\mathbb{B} = \langle \mathcal{B}, 0, 1, +, \cdot, ' \rangle$ and a Boolean expression $\mu(v_1, \ldots, v_n)$ with variables $v_1, \ldots, v_n$, we can determine a corresponding Boolean function $\mu^n_{\mathcal{B}}(v_1, \ldots, v_n)$ such that for every n-tuple $(b_1, \ldots, b_n) \in \mathcal{B}$, $\mu^n_{\mathcal{B}}(b_1, \ldots, b_n)$ is the element of $\mathcal{B}$ obtained by assigning the values $b_1, \ldots, b_n$ to $v_1, \ldots, v_n$ respectively, and interpreting the symbols "$+$", "$\cdot$", and "$'$" to stand for the corresponding operations in $\mathbb{B}$.

For example, according to the standard truth-table definitions of the Boolean operators "or" (written as "$\vee$" or "$+$" in our notation) and "not" (written as "$'$" or "$\sim$" in our notation) the Boolean expression $a'+b'$ determines the Boolean function $f^2(a, b)$ which, in respect to the two-element Boolean algebra containing the carrier set $\mathcal{B} = \{0, 1\}$, maps the set of ordered pairs $\{\langle 0, 0\rangle, \langle 0, 1\rangle, \langle 1, 0\rangle, \langle 1, 1\rangle\}$, corresponding to the possible values of the variables a and b in the expression to the set $\{0, 1\}$ as follows: $f^2 : (0, 0) \rightarrow 1$, $f^2 : (0, 1) \rightarrow 1$, $f^2 : (1, 0) \rightarrow 1$, and $f^2 : (1, 1) \rightarrow 0$.

Henceforth, we shall use the following superscripted lower-case letters of the Latin alphabet to stand for particular Boolean functions: f^n, g^n, h^n; and the upper-case letters F^n, G^n, H^n to stand for arbitrary Boolean functions (where n is the degree of the function as defined in Definition 2.3). Also, whenever the arguments of the function are specified, we shall drop the superscript n. Finally, we shall refer to Boolean functions applied to their arguments also as "Boolean functions", even though they are technically not functions. For example, although the expression $f^2(a, b)$ refers to the result of applying a particular function f^2 to the variables a and b, we nonetheless shall, as is common practice, refer to the entire expression as a "Boolean function".

With this simple vocabulary we can define the notion of a Boolean category. Simply stated, Boolean categories are sets of

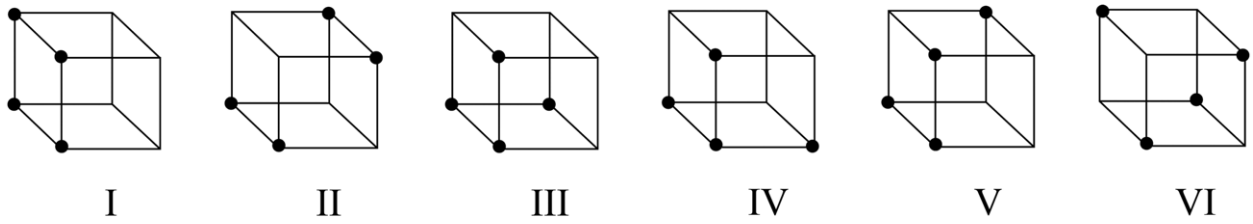


Fig. 2.1. Family of categories with three binary dimensions and four positive examples graphed in Boolean space. These are known as the Shepard, Hovland, and Jenkins (SHJ) category types.

objects defined by Boolean expressions. Thus, they are what some logicians refer to as the *extension* of a Boolean function (see Church, 1956 for a discussion of the *notion of extensionality*). These expressions define sets of objects when each of their distinct variables is interpreted as a distinct dimension. The values of the variables (i.e., 0 or 1) can then be interpreted as one of two possible features. For example, if we let the variable x stand for the binary dimension of color (binary because we restrict its range to two possible values) and y for the binary dimension of shape, then, by letting x = red, x' = black, y = round, and y' = square, we can define the Boolean set or category consisting of a black round object and a red square object with the Boolean expression $x'y+xy'$.

Boolean expressions or formulae that *completely* define Boolean categories are, like the expression in our previous example, in *disjunctive normal form* (DNF). Before defining what it means for a formula to be in disjunctive normal form, we define a *fundamental conjunction* as either (1) a literal (i.e., a negated or unnegated variable), or (2) a conjunction of two or more literals no two of which involve the same variable (where by a conjunction we mean a product of Boolean expressions and by a disjunction we mean a sum of Boolean expressions). For example, while xyz (abbreviation for $x\wedge y\wedge z$) is a fundamental conjunction, xyx' is not a fundamental conjunction since the literals x and x' involve the same variable x.

A Boolean formula φ is said to be in DNF if either (1) φ is a fundamental conjunction, or (2) φ is a disjunction of two or more fundamental conjunctions (as long as not all of the literals of any fundamental conjunction in φ occur in any other fundamental conjunction in φ). Furthermore, there is a special type of DNF, known as a *full* DNF, which we will be particularly interested in since the types of categories in Shepard et al. (1961) and in Feldman (2000, 2003a) assume this form.

A Boolean expression φ in DNF is said to be in *full* DNF in respect to its variables $x_1, \ldots, x_n$ if (1) any variable in φ is one of the variables $x_1, \ldots, x_n$ (i.e., φ is closed under the set of variables $\{x_1, \ldots, x_n\}$), and (2) each disjunct in φ contains all the variables $x_1, \ldots, x_n$. An important theorem in the Boolean algebra states that any Boolean function that is not a self-contradiction (i.e., equal to zero under all possible truth-value assignments to its variables) is logically equivalent to a DNF expression.

Categories defined by full DNF formulae have been studied extensively by investigators such as Shepard et al. (1961), Bourne (1966), and more recently by Nosofsky, Gluck, Palmeri, McKinley, and Glauthier (1994) and Feldman (2000). Of particular interest are the Boolean categories investigated by Shepard et al. (1961) consisting of three binary features or dimensions and four examples (four positives) and four non-examples (four negatives) for a total of eight stimuli. (Later, Feldman (2000) refers to these types of categories as the 3(4) family of "concepts" where the numeral 3 denotes the number of binary dimensions or features and the numeral 4 denotes the number of positive stimuli[1]). For example, suppose that the dimensions involved are those of shape, size, and color; if we let x stand for triangular, x' stand for round, y stand for small, y' stand for large, z stand for white, and z' stand for black, then one of the categories studied by Shepard et al. can be expressed by the Boolean formula $x'y'z' + x'y'z + x'yz' + x'yz$, where the symbol "+" stands for the logical operator *or*.

In other words, the formula perfectly defines the category associated with objects that are either round, large, and black, or round, large, and white, or round, small, and black, or round, small, and white. Note that each of the four conjunctions $x'y'z'$, $x'y'z$, $x'yz'$, and $x'yz$ in the formula above represents a positive example of the category, while the remaining four out of eight (2^3) possible conjunctions or logical products represent non-examples or negative examples of the category. Since in this paper we shall use Boolean formulae exclusively to define Boolean categories, then each Boolean formula in full DNF should be understood as denoting a Boolean category. That is, in our discussion, all Boolean formulae are mathematical representations of well-defined categories.

Although there are 70 $\left(C_4^8 = \frac{8!}{(8-4)!4!}\right)$ possible ways of selecting four positive examples out of a total of eight possible objects, subsets of these 70 are structurally isomorphic or structurally reducible to each other. This is apparent when considering that in a Boolean expression defining a Boolean category, the choice of labels (i.e., literals) for the various features as well as their negation is arbitrary, so that any Boolean expression generated from another as the result of a consistent reassignment of labels and their negation defines the same category structure as the original. For example, the category consisting of a black circle and a white square is structurally equivalent to the category consisting of a black square and a white circle, and both are defined by $xy + x'y'$ or by $xy' + x'y$.

Structural equivalence can also be illustrated geometrically using a Boolean cube (see Fig. 2.1). The relationship between the four points representing the four disjuncts in the Boolean expression is invariant in respect to rigid rotations of the cube. It turns out that there are exactly six such structural relationships that partition the set of all possible Boolean categories 3(4) (three dimensions and four positive examples) into six subsets or equivalent classes (for an in depth combinatorial discussion see Aiken (1951) or Higonnet and Grea (1958)). These six category types combined are referred to as the 3(4) family of Boolean categories. More generally, a *D(p)* family of Boolean categories is a set consisting of all the category types with D dimensions and p positive examples.

Members of the 3(4) category family are represented in Fig. 2.1 as points in Boolean Space and in Fig. 2.2 as Boolean functions in full DNF. Whenever convenient, we use Roman numerals as abbreviations for the longer Arabic numeral descriptions of category types involving the number of dimensions and positive examples: e.g., type I is a shorthand for 3(4)-1.

[1] Although the terms "concept" and "category" are often used interchangeably in the literature, we advice against such practice. Accordingly, in this paper, we reserve the term "concept" to refer to the mental object, mental event, or mental representation corresponding to, or associated with, some set of objects given as a categorical stimulus. We call such sets of objects"categories". Moreover, whenever such sets of objects are defined by a Boolean expression or rule, we shall refer to them as "Boolean categories", and more generally, as"Boolean sets".

Concept Type	Boolean Expression
3(4)-1	$x'y'z' + x'y'z + x'yz' + x'yz$
3(4)-2	$x'y'z' + x'y'z + xyz' + xyz$
3(4)-3	$x'y'z' + x'y'z + x'yz' + xy'z$
3(4)-4	$x'y'z' + x'y'z + x'yz' + xy'z'$
3(4)-5	$x'y'z' + x'y'z + x'yz' + xyz$
3(4)-6	$x'y'z' + x'yz + xy'z + xyz'$

Fig. 2.2. Boolean expressions corresponding to the Shepard, Hovland, and Jenkins category types.

These six types were studied by Shepard et al. (1961) who measured the degree of learning difficulty by the number of errors that subjects made until reaching a criterion for correct classifications. It was found that the six category types listed in Figs. 2.1 and 2.2 followed the following order of learning difficulty: I $<$ II $<$ III, IV, V $<$ VI. Thus type I problems involving the simplest concept or category structure yielded the least number of errors, followed by type II, followed by types III, IV, V which yielded approximately the same number or errors, and finally type VI. This ordering, known as the SHJ ordering, has served as a fundamental benchmark for models of human concept learning (for a discussion see Nosofsky et al. (1994)).

2.2. Minimization–complexity hypothesis

Feldman (2000), motivated by the Shepard et al. (1961) study, sought to find a connection between the degree of learning difficulty of a Boolean category and what he defines as its Boolean complexity. As defined by Feldman (2000, p.630): "The Boolean complexity of a propositional concept is the length of the shortest Boolean formula logically equivalent to the concept, usually expressed in terms of the number of literals (positive or negative variables)". Feldman's study examines an unprecedentedly large number of Boolean category families including the SHJ family: these included the 3(2), 3(3), 3(4), 4(2), 4(3), and 4(4) families for a total of 41 Boolean category types.

In addition, Feldman pays close attention to families where the number of positive and negative examples differ. For example, 3(2) categories each have two positives and six negatives and their mirror image has six positives and two negatives. Feldman refers to this distinction as a distinction in the parity of the Boolean category, where the category is in up parity when the number of positive examples is smaller than the negative examples and in down parity when the positives examples are swapped for the negatives.

Taking complexity (and parity) as his independent variables and proportion of correct responses as his dependent variable, he concludes from the data that, in general, subjective difficulty is well predicted by Boolean complexity and category parity. Clearly, since parity cannot be tested for the SHJ types, Boolean complexity is then the sole independent variable in regards to the SHJ types. Moreover, Feldman (2000, p. 630) claims: "When the SHJ types are considered from the perspective of mathematical logic, however, a simple explanation of the difficulty ordering emerges: the difficulty of the six types is precisely predicted by their Boolean complexity." He adds later in page 631 that: "These Boolean complexity values predict the order of empirical difficulty precisely. This exact correspondence has not previously been noted, though Shepard et al. speculated about it in their original paper, and the relation between Boolean complexity and human learning has never been comprehensively tested."

Under this approach lie the suppositions (1) that there are structural properties intrinsic to a stimulus that can be used to describe how complicated the stimulus is, and (2) that such properties are good predictors of the learning difficulty of the stimulus. Unfortunately, Feldman's minimization-complexity hypothesis has faced strong challenges. For one, Feldman himself has acknowledged that the simplification (i.e., minimization) heuristic of Feldman (2000) based on factorization alone may not be cognitively plausible (Feldman, 2003b).

Secondly, the complexity values reported by Feldman (2000) are not in fact the minimal values for 9 of the 41 category types (Vigo, 2006). Of particular concern is the fact that the actual minimal expressions corresponding to the SHJ types do not reflect the SHJ ordering (see Vigo, 2006, for a discussion) and make the overall complexity predictions less significant. Henceforth, we refer to the complexity model that is based on the actual minimal expressions as the "strong m-complexity model", and likewise, we refer to the model implemented by Feldman (2000) in his Nature paper as the "weak m-complexity model".

Feldman (2006), aware of these difficulties, has recently introduced his spectral decomposition model. Like Garner (1970) the basic idea is that learning from examples involves the extractions of patterns and regularities. The formal model describes how a pattern (expressed in terms of a Boolean rule) may be decomposed algebraically into a "spectrum" of component patterns, each of which is a simpler or more "atomic" regularity. Regularities of higher degree represent more idiosyncratic patterns while regularities of lower degree represent simpler patterns in the original decomposed pattern. The full spectral breakdown of the component patterns of a Boolean category in terms of minimal component regularities is known as the power series of the pattern. These are expressed in terms of what he calls implication polynomials.

An implication polynomial Φ^K of degree K is an expression of the form $\sigma_1 \cdots \sigma_K \rightarrow \sigma_0$ (where each σ_i stands for a Boolean variable and each Boolean variable stands in turn for a particular dimension). These types of expressions are structurally equivalent or congruent to $\neg(\sigma_1 \cdots \sigma_{K+1})$ (see the discussion above on structural equivalence). It turns out that the union of categories of arbitrary degree can be expressed completely as the conjunction of the implication polynomials that each category satisfies. Given this fact, a natural measure of complexity would be to determine the number of rules of each degree (i.e., number of dimensions) that a given Boolean category satisfies. An obstacle to this is the fact that there are many possible structurally equivalent or redundant rules for each degree considered.

To address this problem, Feldman shows that in fact there is a minimal irredundant set of implication polynomials $\mathcal{P}(\mathbf{x})$, where $\mathbf{x}$ is a set of objects belonging to the category that is defined by the Boolean expression in question (for details see Feldman, 2006). Based on this proposition he defines a measure of complexity for Boolean categories in terms of their power spectrum. Letting Φ_x^K denote the set of implication polynomials in $\mathcal{P}(\mathbf{x})$ having degree K, the power spectrum of $\mathbf{x}$ at degree K is denoted by λ_x^K and is simply the size or cardinality of the set Φ_x^K (i.e., $\lambda_x^K = |\Phi_x^K|$). The algebraic complexity of a Boolean category is then defined by the weighted sum of its power spectrum:

$$\lambda^T = \sum_{K=0}^{D-1} w_K \lambda_x^K$$

where the weights w_K are linear increasing in K and sum to zero (where $w_K \propto K$, $\sum w_K = 0$, and $\sum |w_K| = 1$). Thus, the model requires a weight per level of decomposition.

With our model of invariance we shall approach Boolean category complexity from a considerably different perspective.

As discussed, our ultimate goal is to devise a structural Boolean algebraic and analytic account of human concept learning that predicts and explains human concept learning difficulty based on invariance principles. As it turns out, this can be achieved without free parameters within the model. With this aim in mind, we proceed to answer the following question: what is it about the intrinsic structure of some Boolean categories that makes them more difficult to learn than others?

We suggest that the answer lies on a radically different notion of complexity which henceforth we shall call "structural complexity or invariance-based complexity (*i*-complexity)". Unlike the minimization notion of complexity examined above (*m*-complexity), *i*-complexity is based on both the inherent degree of invariance of the Boolean category and its size (i.e., cardinality). This synthesis of invariance and complexity notions may be regarded as a first step toward the conceptual unification of the various structural accounts of concept learning discussed above. Since the key quantity in our definition of structural complexity is the degree of categorical invariance, we lay the foundations for it next.

3. Logical manifold theory

3.1. The derivative of a Boolean category

In the previous sections we discussed how the structural properties of invariance and complexity, independently, have played a key role in the development of structural models. Also, we characterized our modeling approach (to be introduced in the next few sections) as unique in that it unifies these two fundamental structural properties in a simple and direct manner. Indeed, in our model, complexity and invariance may be understood as inverses of each other! However, although complexity will play an important role in our model, it is invariance that is the crux of the model, since it is invariance (along with cardinality) that will determine the degree of structural complexity of a stimulus in the first place.

In view of this, the key to our model is to find a way of characterizing in an economical, intuitive, and natural way, the degree of invariance inherent to a Boolean category, and how it bears on its structural complexity. We accomplish this by introducing next a mathematical framework that is a rudimentary hybrid of Boolean algebra, discrete topology, and analysis. The framework is based in part on a concept that is analogous to some extent to the derivative in calculus.

The Boolean derivative was introduced by Reed (1954) in a discussion of error-correcting codes in electrical circuits. The basic concept has been mainly relegated to this very specialized domain of applied Boolean algebra. A more comprehensive study and generalization of the concept can be found in Thayse (1981). Unfortunately, the notation found in the aforementioned papers is somewhat clumsy and counterintuitive in respect to our goals; hence, we have devised a different notation using the hatted partial derivative operator suggestive of (two) discrete states.

In addition, the concepts that we are about to define, including the *logical manifold*of a category, the *logical norm*, *invariance signatures*, and the rest of our mathematical framework (in short, "logical manifold theory"), including the derivations of our invariance order, measures of invariance, and the invariance "laws" based on these, are all original contributions to the field. In fact, as far as we know, this is the first time that simple core concepts from analysis, discrete topology, and Boolean algebra have been integrated in this fashion with the goal of measuring degrees of categorical invariance. Next, we proceed with a discussion of the Boolean derivative.

The Boolean derivative is an effective formalism for revealing the invariance information intrinsic to Boolean categories. The derivative of a Boolean expression is defined as follows.

Table 3.1
Definition of the "exclusive-or" connective.

x	1	0	1	0
y	1	1	0	0
$(x \oplus y)$	0	1	1	0

Definition 3.1. Let F be a Boolean function of $n \geq 1$ variables $x_1 \cdots x_n$, then its Boolean partial derivative in respect to x_i (where $1 \leq i \leq n$) is defined as

$$\frac{\widehat{\partial} F(x_1 \cdots x_n)}{\widehat{\partial} x_i} = F(x_1 \cdots x_i \cdots x_n) \oplus F(x_1 \cdots \sim x_i \cdots x_n).$$

The symbol "$\oplus$" stands for the "exclusive-or" operator defined by the truth-table in Table 3.1.

Note that the Boolean partial derivative is analogous to the partial derivative in Calculus since in both we evaluate how the dependent variable $F(x_1 \cdots x_i \cdots x_n)$ changes in respect to a change in the independent variable x_i. As in differential calculus, we define the nth-order Boolean partial derivative of a Boolean function as the result of taking the Boolean partial derivative of the Boolean partial derivative of the Boolean partial derivative etc. of the function a total of n times as illustrated below (note that $\frac{\widehat{\partial}^1 F(x_1 \cdots x_i \cdots x_n)}{\widehat{\partial} {x_i}^1}$ below is the same thing as $\frac{\widehat{\partial} F(x_1 \ldots x_n)}{\widehat{\partial} x_i}$):

$$\frac{\widehat{\partial}^n F(x_1 \cdots x_i \cdots x_n)}{\widehat{\partial} x_i^n} = \frac{\widehat{\partial}^n}{\widehat{\partial} x_i^n} \left(\frac{\widehat{\partial}^{n-1}}{\widehat{\partial} x_i^{n-1}} \cdots \left(\frac{\widehat{\partial}^1 F(x_1 \cdots x_i \cdots x_n)}{\widehat{\partial} {x_i}^1} \right) \right).$$

Henceforth, we shall use the following shorthand for the partial Boolean derivative of a Boolean function:

$$\frac{\widehat{\partial} F(x_1 \cdots x_n)}{\widehat{\partial} x_i} = \widehat{\partial}_{x_i} F(x_1 \cdots x_n).$$

Although higher-order Boolean partial derivatives (i.e., of order greater than one) are all equal to zero, they are nonetheless useful in generating equivalent but distinct Boolean expressions. The partial Boolean derivative of a Boolean expression acts as a test for logical equivalence in that it determines whether or not the Boolean expression obtained by replacing x_i with ${x'}_i$ is equivalent to the original.

For example, if we let $F(x, y) = xy$, then the resulting Boolean derivative of $F(x, y)$ in respect to x is $xy \oplus x'y$. If we construct a truth table for this new expression we obtain the value true or 1 when y is true and the value false or 0 when y is false. Thus, for this expression, the value of its Boolean derivative can be predicted by the value of the variable y. The key idea to remember here is that the derivative of this function is *contingent* upon the value of y.

To facilitate our analysis, we evaluate the Boolean derivatives of Boolean expressions using truth tables. Truth table analysis allows us to identify under which truth-value assignments to its variables the Boolean derivative is true. If it is true under all possible truth-value assignments to its variables, then we say that the Boolean derivative is a tautology; on the other hand, if it is false under all its possible truth-value assignments then we say that it is a self-contradiction. When the Boolean derivative is neither a tautology nor a self-contradiction, it is called a contingency.

Clearly, Boolean expressions that are tautologies or self-contradictions are truth-value invariant in respect to the truth-value assignments to their variables, while contingencies only exhibit a certain degree of truth-value invariance. We shall discuss this point in greater detail in the following sections. Fig. 3.1 gives the Boolean derivatives for the SHJ category types in respect to each of their dimensions. We computed these using truth-table analysis as illustrated in Fig. 3.2 for the 3(4)-3 category type. In the

Concept Type	Partial Boolean Derivative in respect to x	Partial Boolean Derivative in respect to y	Partial Boolean Derivative in respect to z
3(4)-1	*tautology*	*self contradiction*	*self contradiction*
3(4)-2	*tautology*	*tautology*	*self contradiction*
3(4)-3	$xyz'+xy'z'+x'yz'+x'y'z'$	$xyz+xy'z+x'yz+x'y'z$	$xy'z+xy'z'+x'yz+x'yz'$
3(4)-4	$xyz'+xy'z+x'yz'+x'y'z$	$xyz'+xy'z'+x'yz+x'y'z$	$xy'z+xy'z'+x'yz+x'yz'$
3(4)-5	*tautology*	$xyz+xy'z+x'yz+x'y'z$	$xyz+xyz'+x'yz+x'yz'$
3(4)-6	*tautology*	*tautology*	*tautology*

Fig. 3.1. Boolean partial derivatives of the category types studied by Shepard et al. (1961).

same figure, contingencies are represented by expressions in full DNF.

But how do these truth-value invariances translate into the degree of qualitative invariance of a Boolean categorical stimulus? Recall that category types are defined by Boolean expressions whose variables represent binary dimensions. Thus, the application of the partial Boolean derivative to an expression that defines a particular category first transforms the objects of the category along one dimension, and then logically "subtracts" the newly obtained category from the original category.

For example, take the Boolean expression $x + y$ (i.e., "x or y") whose DNF is $xy'+x'y+xy$. In this latter expression, if we assign the value *black* to the variable x representing the color dimension and assign the value *circular* to the variable y representing the shape dimension, then we say that $xy' + x'y + xy$ defines the category consisting of three objects, namely, a black triangle, a white circle, and a black circle, as illustrated in Fig. 3.4. Accordingly, the expression $xy'+x'y+xy$ evaluates to true or 1 under only three out of four possible truth-value assignments: namely, the assignments (1, 0), (0, 1), and (1, 1). These three pairs of values correspond to (*black, triangular*), (*white, circular*), and (*black, circular*) under our interpretation, and the fact that they satisfy (i.e., make true) the expression $xy' + x'y + xy$ indicates that they are members of the category that the expression defines.

Now, the Boolean derivative of $xy' + x'y + xy$ in respect to x (the variable standing for the dimension of color) is the logical difference (as defined by the exclusive-or operator $\oplus$) between $xy' + x'y + xy$ (the expression that defines the original category described above), and $x'y'+xy+x'y$ (the expression that defines its corresponding *perturbed* category). In our example, this perturbed category consists of a *white triangle*, a *black circle*, and a *white circle*, as illustrated in Fig. 3.4, and like in the original category defined by $xy' + x'y + xy$, those truth-value assignments that satisfy the expression $x'y'+xy+x'y$ represent the objects or exemplars in the perturbed category.

The logical difference between the expressions that define the two categories simply tells us which objects in the original category are also in the perturbed category. This can be verified by examining Fig. 3.4. The basic idea is that, of the truth-value assignments that satisfy (i.e., make true) $xy' + x'y + xy$, those that *do not* satisfy its Boolean derivative are precisely those that are equivalent to $x'y' + xy + x'y$ and hence, those with corresponding exemplars or objects in the perturbed category defined by $x'y' + xy + x'y$. The reason is that the logical difference between two Boolean expressions evaluates to 0 only when the two expressions are equivalent (i.e., are satisfied or made true by the same truth-value assignments).

3.2. The logical norm

In Section 3.1, we showed that, with an appropriate measure of the Boolean derivative, it might be possible to express quantitatively the degree of partial invariance inherent to a Boolean category. In this section we introduce such a measure. The goal is to measure the degree of qualitative invariance that is revealed by an application of the Boolean partial derivative operator on a Boolean category (or more precisely, on the Boolean expression that defines it).

The intuition behind such a measure is that it must be a count of the number of items from the original category that are preserved in the derived (or perturbed) category. This is achieved by finding out for each possible truth-value assignment that satisfies (i.e., makes true) the Boolean expression that defines the category, whether or not it also *does not* satisfy the Boolean derivative of the Boolean expression that defines the category. The idea is that, if false, that particular conjunction representing an item in the category has been preserved in the perturbed category. If, on the other hand, the particular conjunction evaluates to true then it has not been preserved in the perturbed category.

For the remainder of this paper, a category family type will be denoted by $D(p)$, where D is the number of binary dimensions and p is the number of given positive examples. When referring to a specific category type, sometimes we shall use the notation $D(p) - t$, where t is a number used in Feldman's catalogue of categories (2003a) to denote the particular members within a family. Boolean expressions that we shall use to define Boolean categories are assumed to be in full disjunctive normal form. Such expressions provide a full description of the dimensional structure of a category. Finally, we use the capital bold letters of the Latin alphabet **F**, **G**, and **H** to stand for arbitrary Boolean expressions intended to define Boolean categories and we shall use the letters F, G, H for the Boolean functions determined by these Boolean expressions. Likewise, we shall abbreviate the Boolean expression corresponding to $F(x_1 \cdots x_i \cdots x_n)$ with $\mathbf{F}_{x_i}$ and $F(x_1 \cdots \sim x_i \cdots x_n)$ with $F_{x_i'}$.

Moreover, we define an instantiated fundamental conjunction (see Section 2 for a definition of a fundamental conjunction) as a vector of truth values that is the result of replacing each literal in the fundamental conjunction for a truth value depending on whether the variable is negated or not. Formally, we define the instantiation function γ as a mapping from a set of literals $L = \{t_1, \ldots, t_n\}$ to a set of truth values $V = \{0, 1\}$ such that $\forall t_i \in L, \gamma(t_i) = 0$ whenever t_i is negated and $\gamma(t_i) = 1$ whenever t_i is not negated (where i and n are positive integers). Then, the instantiated fundamental conjunction consisting of n literals $t_1 \ldots t_i \ldots t_n$ is the vector of truth values defined by $\mathbf{I}(t_1 \ldots t_i \ldots t_n) = \gamma(t_1) \ldots \gamma(t_i) \ldots \gamma(t_n)$. For example, $\mathbf{I}(xyz') = 110$ and $\mathbf{I}(xy'z'w) = 1001$. Throughout our discussion, and as previously alluded to, we assume that these vectors of truth values (or strings) correspond to stimuli in categories whenever each of their components is systematically and consistently interpreted as a value of the particular relevant dimension across all categories in question.

In addition, we define the set $\mathcal{A}_0(\mathbf{F})$ as the set of truth-value assignments represented also as vectors of zeros and ones (as above) for which the Boolean expression **F** (not necessarily in DNF) is false. Likewise, $\mathcal{A}_1(\mathbf{F})$ is the set of truth-value assignments represented also as vectors of zeros and ones (as above) for which

X	Y	Z	((~X ^ ~Y ^ ~Z) + (~X ^ ~Y ^ Z) + (~X ^ Y ^ ~Z) + (X ^ ~Y ^ Z)) ⊕ ((X ^ ~Y ^ ~Z) + (X ^ ~Y ^ Z) + (X ^ Y ^ ~Z) + (~X ^ ~Y ^ Z))
0	0	0	1 1 1 1 1 1 1 1 0 1 1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 0 1 0 1 1 1 0
0	0	1	1 1 1 0 0 1 1 1 1 1 1 1 0 0 0 1 0 1 0 0 0 1 0 1 0 0 0 0 1 1 1 1
0	1	0	1 0 0 0 1 0 1 0 0 0 1 1 1 1 1 1 0 0 0 1 0 0 0 1 0 0 0 1 0 1 0 0
0	1	1	1 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
1	0	0	0 0 1 0 1 0 0 0 1 0 0 0 0 0 1 0 1 1 1 0 1 1 1 0 1 0 0 1 0 1 0
1	0	1	0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 0 1 1 0 0 1 1 1 0 0 1 0 1 0
1	1	0	0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 1 1 0 0 0 0
1	1	1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0

Fig. 3.2. Boolean partial derivative in respect to x of the 3(4)-3 category type. The exclusive-or operation (in the middle of the Boolean expression) is represented by "⊕" while "or" is represented by "+" and "and" is represented by "^".

the Boolean expression $\mathbf{F}$ (not necessarily in DNF) is true. Also, we define $\mathcal{D}(\mathbf{F})$ as the set of fundamental conjunctions of a Boolean expression in full DNF and $\mathcal{D}_1(\mathbf{F}) = \{\mathbf{I}(c) \mid c \in \mathcal{D}(\mathbf{F})\}$ as the set of instantiated fundamental conjunctions of $\mathbf{F}$ in full DNF. This set can also be interpreted as the set of truth-value assignments that satisfy (i.e., assign a value of 1 to) $\mathbf{F}$. Clearly, $|\mathcal{D}(\mathbf{F})| = |\mathcal{D}_1(\mathbf{F})| = p$ and $\mathcal{D}_1(\mathbf{F}) = \mathcal{A}_1(\mathbf{F})$ (see Mendelson (1979, for a proof of the latter). We can now justify why it is that the Boolean partial derivative is indeed a qualitative indicator of invariance between the original Boolean category or category of items and the derived category.

To achieve this goal let us examine the intuitive way of measuring the invariance expressed by the Boolean partial derivative. First, consider that the Boolean partial derivative $F(x_1 \cdots x_i \cdots x_n) \oplus F(x_1 \cdots \sim x_i \cdots x_n)$ compares two Boolean expressions $F(x_1 \cdots x_i \cdots x_n)$ and $F(x_1 \cdots \sim x_i \cdots x_n)$ that define two categories respectively. We shall refer to the category defined by $F(x_1 \cdots x_i \cdots x_n)$ as the original category $\mathbf{F}_{\mathbb{O}}$ and to the category defined by $F(x_1 \cdots \sim x_i \cdots x_n)$ as the "perturbed" category $\mathbf{F}_{\mathbb{P}}$. If we represent the stimuli (i.e., objects) of the categories by the instantiated conjunctions of the DNF expressions that define them, then $\mathbf{F}_{\mathbb{O}} = \mathcal{D}_1(F(x_1 \cdots x_i \cdots x_n))$ and $\mathbf{F}_{\mathbb{P}} = \mathcal{D}_1(F(x_1 \cdots \sim x_i \cdots x_n))$.

Then, the simplest way of measuring the invariance expressed by a partial Boolean derivative is by taking the ratio between the number of exemplars that the original category and the perturbed category have in common and the number of exemplars in the original category. More formally, $\frac{|\mathbf{F}_{\mathbb{O}} \cap \mathbf{F}_{\mathbb{P}}|}{|\mathbf{F}_{\mathbb{O}}|}$. The problem with this expression is that it has been formulated in terms of categories and not in terms of the formal properties of the logical description of such categories. Of course, since what we wish to build is a model of invariance in Boolean algebraic terms, we then use the following definition instead.

Definition 3.2. Let $\mathbf{F}$ be a Boolean expression that defines a Boolean category. The logical norm or L-norm of $\mathbf{F}$ in respect to x_i is defined as

$$\left\| \frac{\widehat{\partial \mathbf{F}}}{\widehat{\partial x_i}} \right\| = \frac{\left| \mathcal{D}_1(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F}) \right|}{|\mathcal{D}_1(\mathbf{F})|}.$$

The basic idea behind this measure is that the number of truth-value assignments that satisfy $\mathbf{F}$ and that are in the set of truth assignments that do not satisfy the derivative of $\mathbf{F}$ are precisely the truth-value assignments representing the preserved objects in the perturbed category $\mathbf{F}_{\mathbb{P}}$. For a proof that this logical measure of qualitative invariance is equal to the intuitive measure $\frac{|\mathbf{F}_{\mathbb{O}} \cap \mathbf{F}_{\mathbb{P}}|}{|\mathbf{F}_{\mathbb{O}}|}$, see Proposition A.1 in the Appendix, which states that $\frac{|\mathcal{D}_1(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F})|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{|\mathbf{F}_{\mathbb{O}} \cap \mathbf{F}_{\mathbb{P}}|}{|\mathbf{F}_{\mathbb{O}}|}$.

Example 3.3. From this definition it is clear that the L-norm of a function whose Boolean derivative is a tautology is 0 since $\frac{|\mathcal{D}_1(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F})|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{|\mathcal{D}_1(\mathbf{F}) \cap \varnothing|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{0}{|\mathcal{D}_1(\mathbf{F})|} = \frac{0}{p} = 0$, the L-norm of a function whose Boolean derivative is a self-contradiction is 1 since $\frac{|\mathcal{D}_1(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F})|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{|\mathcal{D}_1(\mathbf{F}) \cap \{000, 001, 010, 011, 100, 101, 110, 111\}|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{|\mathcal{D}_1(\mathbf{F})|}{|\mathcal{D}_1(\mathbf{F})|} = 1$, and the L-norm in respect to x of the Boolean category $\mathbf{F} = x'y'z' + x'y'z + x'yz' + xy'z$ computes to $\frac{1}{2}$ since by truth table analysis (see Fig. 3.2) $\mathcal{A}_0(\widehat{\partial}_x\mathbf{F}) = \{001, 011, 101, 111\}$ and $\mathcal{D}_1(\mathbf{F}) = \{000, 001, 010, 101\}$ making $\frac{|\mathcal{D}_1(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_x\mathbf{F})|}{|\mathcal{D}_1(\mathbf{F})|} = \frac{2}{4} = \frac{1}{2}$.

Again, it is important to recognize that, from a qualitative standpoint, the Boolean derivative may be construed as a relationship between two categories. As such, the L-norm is a direct measure of this relationship between the original Boolean category and the perturbed one with its changed dimensional value or feature (e.g., the color white to the color black). This is evident when considering the case of maximal logical identity between the two Boolean functions representing the two categories: that is, when every truth-value assignment to the Boolean derivative evaluates to 0 (i.e., a self-contradiction).

In this case, the two categories are identical since the Boolean expressions that define them are equivalent under every possible truth-value assignment to the Boolean derivative. More specifically, if the Boolean derivative of the Boolean expression that defines category $\mathbf{F}_{\mathbb{O}}$ evaluates to 0 or false (in respect to dimension d) for any of its truth-value assignments, then each of the exemplars or objects that is in $\mathbf{F}_{\mathbb{O}}$ is also in the perturbed category $\mathbf{F}_{\mathbb{P}}$. Likewise, if the Boolean derivative of the function that defines category $\mathbf{F}_{\mathbb{O}}$ evaluates to 1 or true for all its possible truth-value assignments (i.e., is a tautology), this indicates that every object in category $\mathbf{F}_{\mathbb{O}}$ will not be in $\mathbf{F}_{\mathbb{P}}$.

Although the Boolean derivative along with its L-norm gives us a means for measuring the degree of partial categorical invariance inherent to a Boolean category in respect to single dimensions, we still need a way of considering and combining invariances across the D dimensions of the Boolean category. To achieve this we introduce the notion of a *logical manifold*. Logical manifolds characterize the degree of invariance or symmetry of a Boolean category as a whole.

3.3. Logical manifolds

Thus far we have introduced the Boolean partial derivative as an operator on Boolean expressions capable of producing a Boolean expression that can potentially describe how much qualitative change a category undergoes as a whole when a featural change takes place along one dimension. Furthermore, we introduced the L-norm of a Boolean partial derivative as a simple quantitative measure of the degree of invariance between a category and its corresponding perturbed category.

Our goal is to measure the total intrinsic imperviousness to change (i.e., invariance) of a Boolean category that is subjected to a Boolean derivative transformation; we do this by computing the L-norm of each of the D Boolean partial derivatives corresponding to the category (where D is the number of dimensions in the Boolean expression needed to define the category). To do this, we introduce a new differential operator called the n-th logical

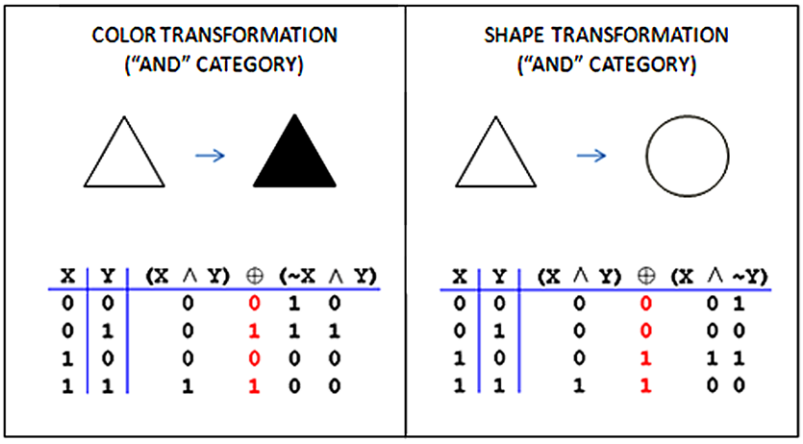


Fig. 3.3. Instance of the AND category transformed into its perturbed counterpart along the two binary dimensions of color (represented by the variable X) and shape (represented by the variable Y). Below each transformation is the Boolean derivative expression that determines which elements are shared by the original category and its perturbed counterpart. Note that no elements are shared in this example.

$$
\begin{aligned}
\Lambda(x+y) &= \left(\left\| \frac{\widehat{\partial}(x+y)}{\widehat{\partial}x} \right\|, \left\| \frac{\widehat{\partial}(x+y)}{\widehat{\partial}y} \right\| \right) \\
&= \left(\frac{\left| \mathcal{D}_1(x+y) \cap \mathcal{A}_0(\widehat{\partial}_x(x+y)) \right|}{|\mathcal{D}_1(x+y)|}, \frac{\left| \mathcal{D}_1(x+y) \cap \mathcal{A}_0(\widehat{\partial}_y(x+y)) \right|}{|\mathcal{D}_1(x+y)|} \right) \\
&= \left(\frac{\left| \mathcal{D}_1(x+y) \cap \mathcal{A}_0((x+y) \oplus (x'+y)) \right|}{|\mathcal{D}_1(x+y)|}, \frac{\left| \mathcal{D}_1(x+y) \cap \mathcal{A}_0((x+y) \oplus (x+y')) \right|}{|\mathcal{D}_1(x+y)|} \right) \\
&= \left(\frac{|\{(0,1),(1,0),(1,1)\} \cap \{(0,1),(1,1)\}|}{|\{(0,1),(1,0),(1,1)\}|}, \frac{|\{(0,1),(1,0),(1,1)\} \cap (\{(1,0),(1,1)\})|}{|\{(0,1),(1,0),(1,1)\}|} \right) = (2/3, 2/3)
\end{aligned}
$$

Box I.

manifold of a Boolean expression. This operator, when applied to a Boolean expression, characterizes the overall invariance of the category that the expression defines.

First, we define the logical gradient of a category as the D-tuple containing its partial Boolean derivatives in respect to each of its D dimensions. In contrast, we can say that a logical manifold is simply the object generated by taking the L-norm of each of the components of the *Boolean category gradient* (or *Boolean set gradient*). A convenient way of thinking about logical manifolds is in terms of vectors in invariance space. We shall revisit this idea in more detail when we introduce the CIM. Formally, we define a logical manifold as a Boolean differential operator on a Boolean expression **F** or, equivalently, on the function $F(x_1 \ldots x_D)$ that the Boolean expression **F** determines.

Definition 3.4. The nth logical manifold (where n is a positive integer) of a Boolean function $F(x_1 \cdots x_D)$ with D variables as arguments (where D is a positive integer) is a differential Boolean operator that transforms $F(x_1 \cdots x_D)$ into an ordered D-tuple of the L-norms of all its partial Boolean derivatives. We shall also refer to the resulting ordered set as the logical manifold of the Boolean category defined by the function $F(x_1 \cdots x_D)$. More formally,

$$
\begin{aligned}
&\Lambda^n(F(x_1 \cdots x_D)) \\
&= \left(\left\| \frac{\widehat{\partial}^n F(x_1 \cdots x_D)}{\widehat{\partial}{x_1}^n} \right\|, \left\| \frac{\widehat{\partial}^n F(x_1 \cdots x_D)}{\widehat{\partial}{x_2}^n} \right\|, \ldots, \left\| \frac{\widehat{\partial}^n F(x_1 \cdots x_D)}{\widehat{\partial}{x_D}^n} \right\| \right).
\end{aligned}
$$

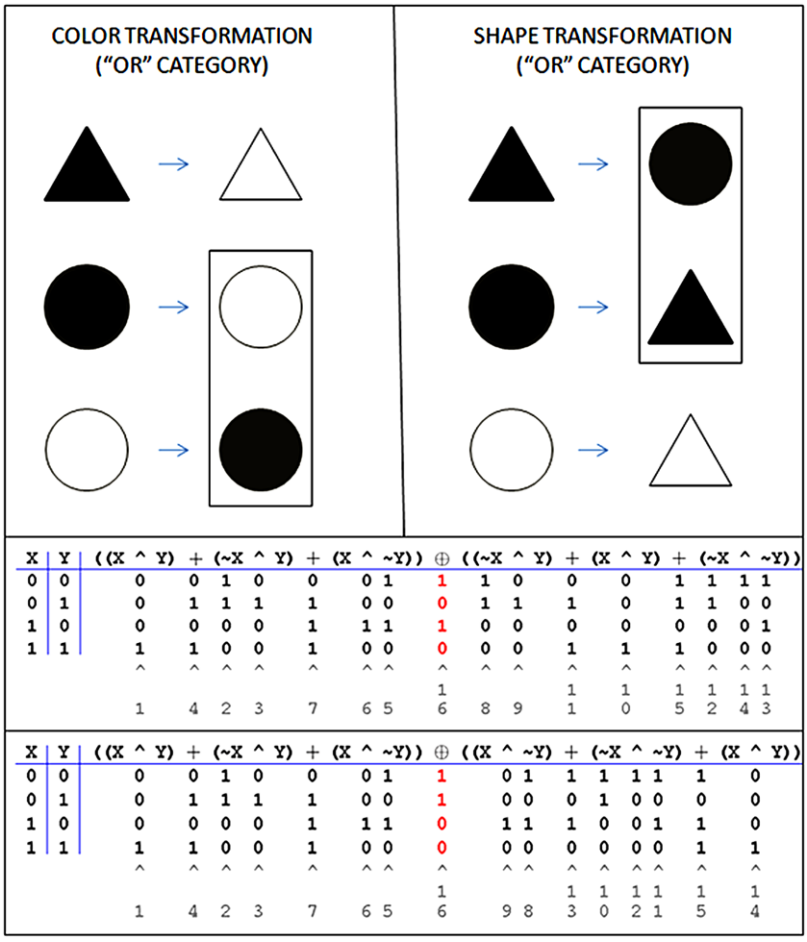
Remark. We shall adopt the convention not to specify the degree of the logical manifold when $n = 1$. That is, $\Lambda^1(F(x_1, \ldots, x_m)) = \Lambda(F(x_1 \cdots x_m))$. Furthermore, if the function F (on variables $x_1, \ldots, x_D$) is determined by the Boolean expression **F**, we shall also abbreviate $\left\| \frac{\widehat{\partial} F(x_1 \cdots x_D)}{\partial x_i} \right\|$ with $\widehat{\mathbf{F}}_{x_i} (1 \leq i \leq D)$, and $\Lambda(F(x_1 \cdots x_D))$ with $\widehat{\mathbf{F}}$ so that $\widehat{\mathbf{F}} = \left(\widehat{\mathbf{F}}_{x_1} \cdots \widehat{\mathbf{F}}_{x_D} \right)$.

Example 3.5. $(0, 0)$ is the logical manifold of xy, as can be seen by the following computation:

$$
\begin{aligned}
\Lambda(xy) &= \left(\left\| \frac{\widehat{\partial} xy}{\widehat{\partial} x} \right\|, \left\| \frac{\widehat{\partial} xy}{\widehat{\partial} y} \right\| \right) \\
&= \left(\frac{\left| \mathcal{D}_1(xy) \cap \mathcal{A}_0(\widehat{\partial}_x xy) \right|}{|\mathcal{D}_1(xy)|}, \frac{\left| \mathcal{D}_1(xy) \cap \mathcal{A}_0(\widehat{\partial}_y xy) \right|}{|\mathcal{D}_1(xy)|} \right) \\
&= \left(\frac{\left| \mathcal{D}_1(xy) \cap \mathcal{A}_0(xy \oplus x'y) \right|}{|\mathcal{D}_1(xy)|}, \frac{\left| \mathcal{D}_1(xy) \cap \mathcal{A}_0(xy \oplus xy') \right|}{|\mathcal{D}_1(xy)|} \right) \\
&= \left(\frac{|\{(1,1)\} \cap \{(0,0),(1,0)\}|}{|\{(1,1)\}|}, \frac{|\{(1,1)\} \cap (\{(0,0),(0,1)\})|}{|\{(1,1)\}|} \right) \\
&= (0, 0).
\end{aligned}
$$

Fig. 3.3 shows an instance of the AND category and the truth tables corresponding to the two Boolean derivatives in respect to each of the two binary dimensions. We can also see in the figure how the original AND category and its perturbed counterpart do not have any elements in common. This fact is described perfectly by the logical manifold of the expression that defines the category.

Similarly, we give an example for the OR category. Please note that the OR category is a 2(3) type category (consisting of two binary dimensions and three positive examples). In disjunctive



X	Y	((X ^ Y)	+	(~X ^ Y)	+	(X ^ ~Y))	⊕	((~X ^ Y)	+	(X ^ Y)	+	(~X ^ ~Y))
0	0	0	0	1	0	0	1	1	1	0	0	0	1	1	1	1
0	1	0	1	1	1	1	0	0	0	1	1	0	1	1	0	0
1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1
1	1	1	1	0	0	1	0	0	0	1	1	1	0	0	0	0
		^	^	^	^	^	^	^	^	^	^	^	^	^	^	^
							1			1	1	1	1	1	1	1
		1	4	2	3	7	6	5	6	8	9	1	0	5	2	4	3

X	Y	((X ^ Y)	+	(~X ^ Y)	+	(X ^ ~Y))	⊕	((X ^ ~Y)	+	(~X ^ ~Y)	+	(X ^ Y))
0	0	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0
0	1	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0
1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1
		^	^	^	^	^	^	^	^	^	^	^	^	^	^	^
							1			1	1	1	1	1	1	1
		1	4	2	3	7	6	5	6	9	8	3	0	2	1	5	4

Fig. 3.4. Instance of the OR category (consisting of two dimensions and three positive examples) transformed into its perturbed counterpart along the two binary dimensions of color (represented by the variable X) and shape (represented by the variable Y). Below the two transformations are the truth tables corresponding to each of the Boolean derivatives. By using the L-norm measure, these tables help us determine which elements are shared by the original category and its perturbed counterpart (in respect to the dimensions of color and shape respectively). Note that two out of three elements are shared in this example in respect to each of the two dimensions. The numbers below the tables indicate the order of evaluation of the Boolean operators.

normal form, a Boolean expression that defines this category is $xy + x'y + xy'$. However, in the following example, for the sake of clarity, we shall abbreviate this formula using the equivalent expression $x+y$. In Fig. 3.4, we give the truth tables for the Boolean derivatives of $xy + x'y + xy'$. The reader can then verify that the logical manifold of the latter is identical to the logical manifold of $x + y$ computed in Box I.

Example 3.6. Fig. 3.5 shows perceptual instances of the category gradients of the 3(4) category family types where the perceptual dimensions are that of color (black or white), shape (triangular or round), and size (large or small).

The table in Fig. 3.7 gives the logical manifolds for the 41 category types studied in Feldman (2000). Logical manifolds encode and reveal quantified invariances in terms of quantitative logical invariances. Recall that the L-norm of a Boolean function gives us a method for measuring degrees of categorical invariance in respect to specific dimensions. The logical manifolds give us an overall picture of the possible degrees of invariance of the category in respect to each categorical dimension. That is, they tell us how much of the original category as a whole remains the same after undergoing a change in dimensional value for each of its dimensions. But what does this have to do with structure? The degrees of invariance of a category (in respect to each of its dimensions) tell us how redundant each particular dimension is in respect to the entire category. But more generally, it measures the amount of relational information intrinsic to the category.

How humans detect invariance in a categorical stimulus is a more difficult question to answer. Fig. 3.6 illustrates a plausible cognitive mechanism for detecting categorical invariance that is consistent with the computational theory accompanying our structural model, the basic idea behind the figure above is that agents possibly detect invariances by comparing a set of objects to the set derived by applying a differential transformation (in respect to each dimension) on the original set. So, if the first dimension is color, then the four objects (in the case of the 3(4) type category) will be represented as the perturbed category consisting of the original four objects differing in color value in such way that every black object is now white, and vice versa. These four objects in the perturbed category are then compared to the objects in the categorical stimulus. The greater the match, the greater the invariance and the easier it is to learn and perhaps encode the particular category.

4. Categorical invariance

4.1. Invariance order

We have already seen how logical manifolds reveal patterns of logical invariance that are intrinsic to the category type at various levels of featural analysis. But we still have not derived

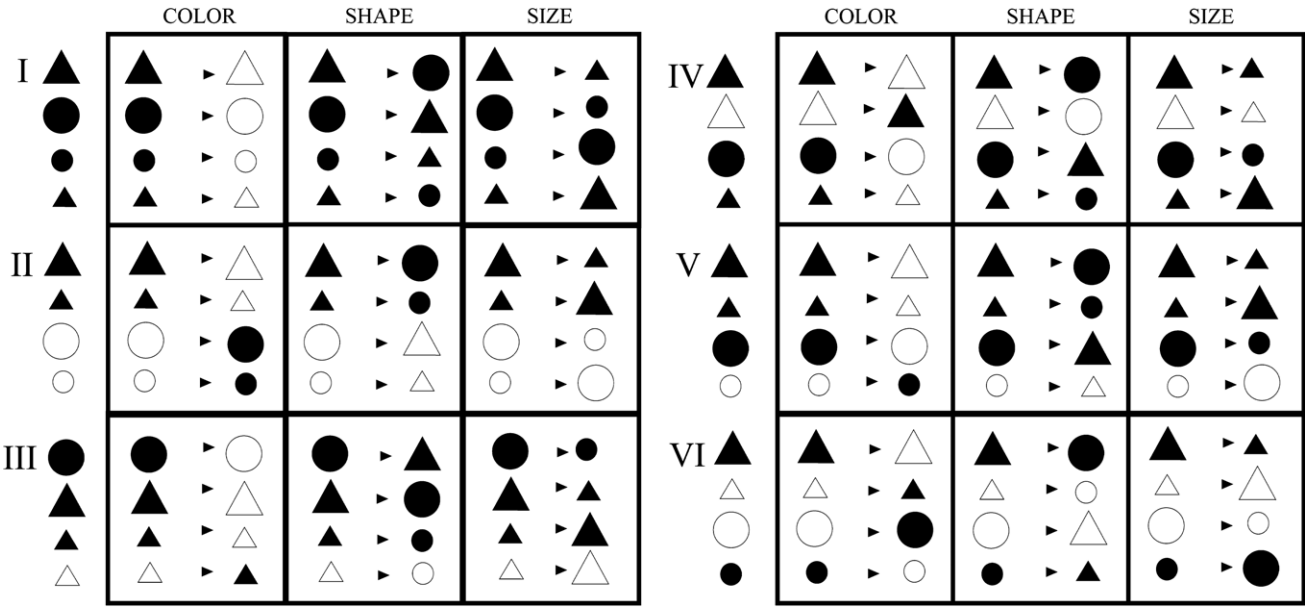


Fig. 3.5. Perceptual instances of category types in the SHJ family transformed three times (one for each binary dimension) by exchanging the values of each of its binary dimensions. Category types I, II, and III are listed in vertical order on the left. Types IV, V, and VI are listed in vertical order in the middle.

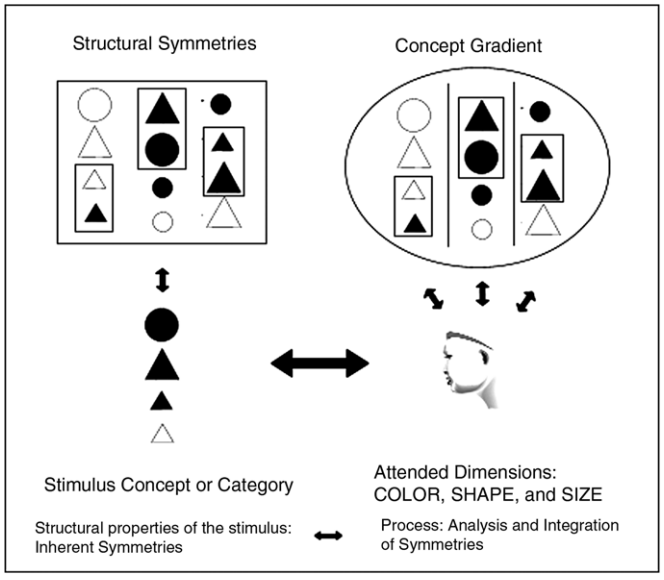


Fig. 3.6. Detecting categorical invariances implicitly: cognitive agents may detect invariances by comparing a set of objects to the set derived by applying a differential transformation (in respect to each dimension) on the original set.

a way of measuring the degree of invariance of a Boolean category as a whole. This is critical since, as explained, our definition of structural complexity will be based on a measure of categorical invariance. In this section we will examine some ordinal characteristics of logical manifolds that will be helpful in formulating such a measure. A natural way of ordering the logical manifolds of the 41 category types in Fig. 3.7 is lexicographically in terms of two dimensions: namely, the degree of invariance and frequency of their components.

Let $D(p)$ be the category family in question. Since logical manifolds have two distinctive aspects – the L-norms of the Boolean derivatives of the Boolean category and the frequency of the L-norms – that we believe are contributors to overall invariance, we define a two-dimensional order on the logical manifolds of the members of the category family $D(p)$. This is achieved by ranking the logical manifolds of the category family

Type	*Boolean Formula*	*LM*
3(2)-1	*((x´y´z´)+(x´y´z))*	(0,0,1)
3(2)-2	*((x´y´z´)+(x´yz))*	(0,0,0)
3(2)-3	*((x´y´z´)+(xyz))*	(0,0,0)
3(3)-1	*((x´y´z´)+(x´y´z)+(x´yz´))*	(0,2/3,2/3)
3(3)-2	*((x´y´z´)+(x´y´z)+(xyz´))*	(0,0 ,2/3)
3(3)-3	*((x´y´z´)+(x´yz)+(xy´z))*	(0,0,0)
3(4)-1	*((x´y´z´)+(x´y´z)+(x´yz´)+(x´yz))*	(0,1,1)
3(4)-2	*((x´y´z´)+(x´y´z)+(xyz´)+(xyz))*	(0,0,1)
3(4)-3	*((x´y´z´) + (x´y´z) + (x´yz´) + (x y´z))*	(1/2,1/2,1/2)
3(4)-4	*((x´y´z´) + (x´y´z) + (x´yz´) + (x y´z´))*	(1/2,1/2,1/2)
3(4)-5	*((x´y´z´) + (x´y´z) + (x´yz´) + (xyz))*	(0,1/2,1/2)
3(4)-6	*((x´y´z´) + (x´y z) + (x y´z) + (x yz´))*	(0,0,0)
4(2)-1	*((x´y´z´w´)+(x´y´z´w))*	(0,0,0,1)
4(2)-2	*((x´y´z´w´)+(x´y´zw))*	(0,0,0,0)
4(3)-3	*((x´y´z´w´)+(x´yzw))*	(0,0,0,0)
4(4)-4	*((x´y´z´w´)+(xyzw))*	(0,0,0,0)
4(3)-1	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´))*	(0,0,2/3,2/3)
4(3)-2	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´))*	(0,0,0,2/3)
4(3)-3	*((x´y´z´w´)+(x´y´z´w)+(xyzw´))*	(0,0,0,2/3)
4(3)-4	*((x´y´z´w´)+(x´y´zw)+(x´yz´w))*	(0,0,0,0)
4(3)-5	*((x´y´z´w´)+(x´y´zw)+(xyz´w´))*	(0,0,0,0)
4(3)-6	*((x´y´z´w´)+(x´y´zw)+(xyz´w))*	(0,0,0,0)
4(4)-1	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´y´zw))*	(0,0,1,1)
4(4)-2	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yz´w´))*	(0,1/2,1/2,1/2)
4(4)-3	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yz´w))*	(0,1/2,1/2,1/2)
4(4)-4	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yzw))*	(0,0,1/2,1/2)
4(4)-5	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyz´w´))*	(0,0,1/2,1/2)
4(4)-6	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyz´w))*	(0,0,1/2,1/2)
4(4)-7	*((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyzw))*	(0,0,1/2,1/2)
4(4)-8	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(x´yzw))*	(0,0,0,1)
4(4)-9	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xy´zw´))*	(0,0,0,1/2)
4(4)-10	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xy´zw))*	(0,0,0,1/2)
4(4)-11	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xyzw´))*	(1/2,0,0,0)
4(4)-12	*((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xyzw))*	(0,0,0,1/2)
4(4)-13	*((x´y´z´w´)+(x´y´z´w)+(xyzw´)+(xyzw))*	(0,0,0,1)
4(4)-14	*((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(x´yzw´))*	(0,0,0,0)
4(4)-15	*((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xy´z´w))*	(0,0,0,0)
4(4)-16	*((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xy´zw´))*	(0,0,0,0)
4(4)-17	*((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xyzw´))*	(0,0,0,0)
4(4)-18	*((x´y´z´w´)+(x´y´zw)+(xyz´w´)+(xyzw))*	(0,0,0,0)
4(4)-19	*((x´y´z´w´)+(x´y´zw)+(xyz´w)+(xyzw´))*	(0,0,0,0)

Fig. 3.7. Logical manifolds (LM) for 41 category types.

in a lexicographical type order so that those logical manifolds containing the greatest number of maximal invariances are first in

Concept Type	Logical Manifold	Invariance Signature
3(4)-1	(0,1,1)	(1,2)
3(4)-2	(0,0,1)	(1,1)
3(4)-3	(1/2,1/2,1/2)	(1/2,3)
3(4)-4	(1/2,1/2,1/2)	(1/2,3)
3(4)-5	(0,1/2,1/2)	(1/2,2)
3(4)-6	(0,0,0)	(0,3)

Fig. 4.1. Logical manifolds and invariance signatures of the SHJ category types arranged according to their lexicographical ordering.

the order. We shall do this indirectly by constructing the set of the invariance signatures of the logical manifolds of the category types.

Definition 4.1. Let $\widehat{\mathbf{F}} = (\widehat{\mathbf{F}}_{x_1} \cdots \widehat{\mathbf{F}}_{x_D})$ be the logical manifold of the Boolean expression **F** that defines a Boolean category in the category family $D(p)$. We define $\max(\widehat{\mathbf{F}})$ as the greatest rational number in $\widehat{\mathbf{F}}$.

Example 4.2. $\max\left(\left(\frac{1}{2}, \frac{1}{2}, 1\right)\right) = 1.$

Definition 4.3. Let S be an ordered n-tuple $(x_1, \ldots, x_n)$. Then, we define $\mathfrak{F}_S(x_i)$, *(where* i and n are positive integers and $1 \leq i \leq n$*)* as the frequency or number of occurrences of x_i in S.

We now define the lexicographical order relation on the set of invariance signatures of the logical manifolds of a category family of type $D(p)$.

Definition 4.4. Let $\mathcal{L} = \left\{\widehat{\mathbf{F}}_1 \ldots \widehat{\mathbf{F}}_m\right\}$ be the set of logical manifolds of the categories in a Boolean category family of type $D(p)$. Corresponding to each $\widehat{\mathbf{F}}_k \epsilon \mathcal{L} (1 \leq k \leq m)$, there exists an ordered pair of real numbers $\left(\max(\widehat{\mathbf{F}}_k), \mathfrak{F}_{\widehat{\mathbf{F}}_k}\left(\max(\widehat{\mathbf{F}}_k)\right)\right)$. We call this pair the invariance signature of $\widehat{\mathbf{F}}_k$. Let $\mathfrak{M} = \left\{\max(\widehat{\mathbf{F}}_k) | \widehat{\mathbf{F}}_k \in \mathcal{L}\right\}$ and let $\mathfrak{F} = \left\{\mathfrak{F}_{\widehat{\mathbf{F}}_k}\left(\max(\widehat{\mathbf{F}}_k)\right) | \widehat{\mathbf{F}}_k \in \mathcal{L}\right\}$. The cross product $\mathfrak{M} \times \mathfrak{F} = \{(x, y) | x \epsilon \mathfrak{M} \wedge y \epsilon \mathfrak{F}\}$ is called the set of invariance signatures of $\mathcal{L}$. We can then define the lexicographical order relation $\succ$ on the set of invariance signatures as follows: $(x, y) \succ \left(x', y'\right) \leftrightarrow [(x > x') \vee [(x = x') \wedge (y > y')]]$. Fig. 4.1 above illustrates this ordering for the SHJ category types.

The lexicographical ordering on invariance signatures characterizes the importance of maximal partial invariance over lower degrees of partial invariance in the logical manifold of **F**. In other words, although they are both important contributing factors, we assume under this interpretation that the presence of one single maximal partial invariance in respect to a particular dimension has a greater ease of learnability impact on a cognizer's perceptual system than a greater number of lower degree invariances (in respect to the remaining dimensions) combined. Hence, our conception of partial invariance is a relative one.

In short, under this view, "goodness of pattern" (Garner, 1974) supersedes frequency of pattern. Accordingly, we assume that the degree of perceived invariance is directly proportional to the maximal invariance value in the logical manifold of **F** and the frequency of the maximal invariance value in the logical manifold of **F**. We know from measurement theoretic results that the lexicographical order as defined above is a strict weak order: that is, a relation that is asymmetric and negatively transitive. More specifically, a binary relation R is asymmetric on some set A iff $\forall x, y \in A, xRy \Rightarrow \neg(yRx)$ and it is negatively transitive iff $\forall x, y, z \in A, \neg(xRy) \wedge \neg(yRz) \Rightarrow \neg(xRz)$. Strict preference relations in utility theory are often interpreted as strict weak orders.

It is well known that strict weak orders are structures for which there exist order-preserving functions (i.e., homomorphisms) into the real numbers. Furthermore, the class of admissible transformations on these order-preserving functions is the class of monotone increasing functions making our scale an ordinal scale (for these measurement theoretic details see Roberts (1979). Given this result, we shall propose such a function in the following section for computing the overall invariance of a Boolean category.

It is important to recognize that invariance information by itself is not enough to characterize structural complexity. The reason is that two categories may have the same logical manifold and a very different number of elements. For example, the Boolean categories 4(2)-2 and 4(4)-19 have the same logical manifolds but the first has two elements while the later has twice as many. Clearly, the latter should be deemed more structurally complex. Thus, as pointed out on several occasions, both the degree of categorical invariance and category size are necessary quantities when determining structural complexity.

Furthermore, the ordinal measure of invariance introduced in this section is a local measure. That is, it only indicates relative degrees of invariance within category families. Our goal in the next section is to propose a quantitative measure of the overall (or global) degree of categorical invariance inherent to a Boolean category across category families (as defined by the general conceptual framework explained in the previous sections). We will also introduce our definition of structural complexity based on the said measure and on $|\mathcal{D}_1(\mathbf{F})|$, the cardinality of the Boolean category.

4.2. Degree of invariance

A natural and intuitive way of measuring the relative degree of categorical invariance of a Boolean category is by calculating the distance (in invariance space) between its logical manifold and the zero logical manifold $\widehat{\mathbf{0}}_\mathbf{D} = (0_1 \cdots 0_D)$ (where D, as before, is the number of arguments of the Boolean function that defines the Boolean category or the number of dimensions associated with the Boolean category in question). The zero logical manifold represents the minimal invariance point in invariance space. For brevity, we shall also write $\widehat{\mathbf{0}}_\mathbf{D} = (0_1 \ldots 0_D)$ as $\widehat{\mathbf{0}} = (0 \cdots 0)$.

Definition 4.5. The degree of categorical invariance $\Phi(\mathbf{F})$ of a Boolean category defined by the Boolean expression **F** and belonging to the family $D(p)$ is the Euclidean distance of its logical manifold from the $(0_1 \ldots 0_D)$ logical manifold.

$$\Phi(\mathbf{F}) = \left[\sum_{i=1}^{D} \left\| \frac{\widehat{\partial} F(x_1 \cdots x_D)}{\widehat{\partial} x_i} \right\|^2\right]^{1/2}.$$

Furthermore, by interpreting the logical manifolds of Boolean categories belonging to a *D(p)* family as vectors in D-dimensional invariance space, then, the length of such vectors (i.e., their norm) represent degrees of invariance. In other words, the degree of invariance of a Boolean category is the norm of the logical manifold of the Boolean function that defines it. This invariance measure preserves the lexicographical difficulty order $\succ$ within the various families. For example, the values for the 3(4) family (i.e., the SHJ family) are 1.414, 1, .866, .866, .707, and 0 respectively. It also shows the direct role that the Boolean derivatives and their L-norms play in determining degrees of invariance.

The conceptual unity achieved by this measure should not be underestimated. For instance, it is interesting to consider that Minkowski Boolean distance measures such as the Euclidean distance above had failed to account for the learning difficulty ordering of Boolean categories. However, when instead we

Concept Type	Logical Manifold	Invariance Signature	CIM	Observed Data (Nosofsky et al., 1994)
3(4)-1	(0,1,1)	(1,2)	1.657	.010
3(4)-2	(0,0,1)	(1,1)	2.000	.032
3(4)-3	(1/2,1/2,1/2)	(1/2,3)	2.143	.061
3(4)-4	(1/2,1/2,1/2)	(1/2,3)	2.143	.065
3(4)-5	(0,1/2,1/2)	(1/2,2)	2.343	.075
3(4)-6	(0,0,0)	(0,3)	4.000	.143

Fig. 5.1. CIM predictions. The observed data column is from Nosofsky et al. (1994). In Nosofsky et al., learning difficulty is given by the average probability of categorization error over 25 blocks of 16 trials for each category type.

consider the distance, not between Boolean objects and/or between Boolean categories, but between the logical manifolds of the Boolean categories and the logical manifold representing minimal invariance (i.e., the $\widehat{\mathbf{0}}$ logical manifold), a much more successful picture emerges.

5. Structural complexity

5.1. Complexity as a function of invariance

With the invariance measure introduced in Section 4.2 we can now define our notion of structural complexity. First, we notice that the higher the degree of invariance of a Boolean category, the more regularity of pattern it has—that is, the simpler it is. Thus, the degree of structural or relational complexity of a Boolean category is indirectly proportional to its degree of invariance. This means that if the degree of invariance is 0, then the function that defines structural complexity must yield a meaningful lowest value for the inverse of 0. Secondly, the function should be monotonic decreasing since any increase in invariance should yield a corresponding decrease in structural complexity.

Moreover, as discussed in the previous section, intuitively, the size of the Boolean category also plays a role in determining its structural complexity. The size or cardinality of a Boolean category is defined by $|\mathcal{D}_{\mathrm{I}}(\mathbf{F})|$ (see Section 4.2). This can also be interpreted as simply the number of positive examples p of the category. Thus, we propose the following formal definition of structural complexity.

Definition 5.1. The degree of structural complexity $\Psi(\mathbf{F})$ of a Boolean category defined by the Boolean expression $\mathbf{F}$ and belonging to the family $D(p)$ is indirectly proportional to its degree of categorical invariance $\Phi(\mathbf{F})$ and directly proportional to its cardinality $|\mathcal{D}_{\mathrm{I}}(\mathbf{F})|$.

$$\Psi(\mathbf{F}) = |\mathcal{D}_{\mathrm{I}}(\mathbf{F})| \times \left[\left[\sum_{i=1}^{D}\left(\left\|\frac{\widehat{\partial}F(x_1\cdots x_D)}{\widehat{\partial}x_i}\right\|^2\right)\right]^{1/2}+1\right]^{-1}.$$

Adding one to $\Phi(\mathbf{F})$ above makes it possible to assign a meaningful quantity to the inverse of 0 without loss of generality. Also, since $|\mathcal{D}_{\mathrm{I}}(\mathbf{F})| = p$,

$$\Psi(\mathbf{F}) = |\mathcal{D}_{\mathrm{I}}(\mathbf{F})| \times [\Phi(\mathbf{F})+1]^{-1} = p\,[\Phi(\mathbf{F})+1]^{-1} = \frac{p}{\Phi(\mathbf{F})+1}.$$

Henceforth, we shall refer to $\Psi(\mathbf{F})$, in the context of Boolean category learnability, as the CIM (concept invariance model) and to its corresponding structural and accompanying computational theory as CIT (concept invariance theory). We shall show in the next section that the CIM is a good predictor of concept learning difficulty. For one, it predicts the empirical SHJ order in that categories of type 3(4)-I are the easiest to learn since they contain the maximal amount of invariance, while categories of type 3(4)-II are more difficult. Categories of type 3(4)-III, 3(4)-IV, and 3(4)-V are of approximately equal difficulty and all more difficult than categories of type 3(4)-II. Finally, categories of type 3(4)-VI are the most difficult of all. In a nutshell, we can describe the difficulty ordering by the following sequence: I(1.66) < II(2) < [III(2.14), IV(2.14), V(2.34)] < VI(4). The table in Fig. 5.1 displays the predictions made by the CIM in respect to the SHJ ordering.

The lack of free parameters in the CIM inclines us to call its relationship to concept learning performance a "law". Readers may try to draw a comparison between the CIM and Shepard's (1987) "law" of generalization. In the CIM we compute the reciprocal of the distance between logical manifolds rather than the negative exponent of the distance between stimuli in some psychological space. In addition, the CIM computes degrees of structural complexity while Shepard's model computes the probability of generalization from a stimulus to another or, in other words, the probability that a certain response learned to stimulus i is made to stimulus j. In Section 5.2 we introduce an exponential version of our model that invites further comparisons to Shepard's model.

Thus far we have concentrated our discussion on the seminal SHJ set of Boolean categories. We have developed a mathematical framework for expressing degrees of invariance in Boolean categories and have introduced a new measure of structural complexity based on the said framework that predicts the critical SHJ learning difficulty ordering. In the following section we aim to find out how well the CIM predicts the degree of learning difficulty for all 41 category types investigated by Feldman (2000) and compare these predictions with those made by the m-complexity model (i.e., minimization complexity) of concept learning. Also, we will discuss how our categorical invariance model compares to Feldman's λ^T spectral decomposition model. We use the same data set that Feldman used in both his paper published in "Nature" (Feldman, 2000) and his paper published in "Journal of Mathematical Psychology" (Feldman, 2006).

5.2. CIM predictions

We compared the learning difficulty predictions made by the CIM for the 41 category types studied by Feldman (2000) to those made by the strong version of Feldman's m-complexity model. A table of the predictions made by both models is seen in Fig. 5.2 below.

Fig. 5.3 shows predictions made by the strong version of the m-complexity model for the 41 category types. The model yields an $R^2 = .42$ accounting for approximately 42% of the variance. The complexity values are computed using a heuristic proposed by Vigo (2006) that is a better approximation of the strong version of the m-complexity model. Feldman's original m-complexity values for the 41 category types (i.e., the weak version of the m-complexity model) yield an $R^2 = .45$. In contrast, our invariance "law" yields and $R^2 = .64$ accounting for approximately 64% of the variance. Fig. 5.4 illustrates this result.

Type	*Boolean Formula*	*Minimization Complexity*	*Structural Complexity*
3(2)-1	((x´y´z´)+(x´y´z))	2	1.00
3(2)-2	((x´y´z´)+(x´yz))	5	2.00
3(2)-3	((x´y´z´)+(xyz))	6	2.00
3(3)-1	((x´y´z´)+(x´y´z)+(x´yz´))	3	1.54
3(3)-2	((x´y´z´)+(x´y´z)+(xyz´))	5	1.80
3(3)-3	((x´y´z´)+(x´yz)+(xy´z))	8	3.00
3(4)-1	((x´y´z´)+(x´y´z)+(x´yz´)+(x´yz))	1	1.66
3(4)-2	((x´y´z´)+(x´y´z)+(xyz´)+(xyz))	4	2.00
3(4)-3	((x´y´z´) + (x´y´z) + (x´yz´) + (x y´z))	6	2.14
3(4)-4	((x´y´z´) + (x´y´z) + (x´yz´) + (x y´z´))	6	2.14
3(4)-5	((x´y´z´) + (x´y´z) + (x´yz´) + (xyz))	6	2.34
3(4)-6	((x´y´z´) + (x´y z) + (x y´z) + (x yz´))	10	4.00
4(2)-1	((x´y´z´w´)+(x´y´z´w))	3	1.00
4(2)-2	((x´y´z´w´)+(x´y´zw))	6	2.00
4(3)-3	((x´y´z´w´)+(x´yzw))	7	2.00
4(4)-4	((x´y´z´w´)+(xyzw))	8	2.00
4(3)-1	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´))	4	1.54
4(3)-2	((x´y´z´w´)+(x´y´z´w)+(x´yzw´))	6	1.80
4(3)-3	((x´y´z´w´)+(x´y´z´w)+(xyzw´))	7	1.80
4(3)-4	((x´y´z´w´)+(x´y´zw)+(x´yz´w))	9	3.00
4(3)-5	((x´y´z´w´)+(x´y´zw)+(xyz´w´))	10	3.00
4(3)-6	((x´y´z´w´)+(x´y´zw)+(xyz´w))	10	3.00
4(4)-1	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´y´zw))	2	1.66
4(4)-2	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yz´w´))	7	2.14
4(4)-3	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yz´w))	7	2.14
4(4)-4	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(x´yzw))	7	2.34
4(4)-5	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyz´w´))	8	2.34
4(4)-6	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyz´w))	8	2.34
4(4)-7	((x´y´z´w´)+(x´y´z´w)+(x´y´zw´)+(xyzw))	8	2.34
4(4)-8	((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(x´yzw))	5	2.00
4(4)-9	((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xy´zw´))	9	2.67
4(4)-10	((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xy´zw))	10	2.67
4(4)-11	((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xyzw´))	6	2.34
4(4)-12	((x´y´z´w´)+(x´y´z´w)+(x´yzw´)+(xyzw))	9	2.67
4(4)-13	((x´y´z´w´)+(x´y´z´w)+(xyzw´)+(xyzw))	6	2.00
4(4)-14	((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(x´yzw´))	11	4.00
4(4)-15	((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xy´z´w))	13	4.00
4(4)-16	((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xy´zw´))	13	4.00
4(4)-17	((x´y´z´w´)+(x´y´zw)+(x´yz´w)+(xyzw´))	13	4.00
4(4)-18	((x´y´z´w´)+(x´y´zw)+(xyz´w´)+(xyzw))	8	4.00
4(4)-19	((x´y´z´w´)+(x´y´zw)+(xyz´w)+(xyzw´))	12	4.00

Fig. 5.2. Minimization-based complexity vs. invariance-based structural complexity of the Boolean categories from six Boolean category families: 3(2), 3(3), 3(4), 4(2), 4(3), and 4(4).

Applying the Spearman σ test, which measures the correlation between the predicted order of the 41 category types and their empirical order, yields an impressive $-.80$. In contrast, the minimization-complexity hypothesis did not show as strong a correlation, with a Spearman σ of $-.60$ for the 41 category types. Notably, there are no free parameters in the CIM, which explains our inclination to call the relationship between the CIM and concept learnability a mathematical law of cognition. Of course, the Boolean *m*-complexity model does not depend on free parameters either. However, as documented, it is not as successful in predicting degrees of learning difficulty for the 41 Boolean category types as the CIM.

Moreover, the Boolean complexity model proposed by Feldman (2000) does not account for what are known as "parity effects". A Boolean category is in up parity if the number of its positive examples is smaller than the number of its negative examples; likewise, it is in down parity whenever the reverse is true. Clearly, 3(4) type categories (i.e., the SHJ types) have no parity. It turns out that categories in up parity have a constant learnability advantage over categories in down parity. This phenomenon had been observed empirically by Haygood and Bourne (1965) in a classic experiment involving rule-based classification.

For example, they discovered that disjunctive rules (Boolean categories of type 2(3)) were more difficult to learn than conjunctive rules (Boolean categories of type 2(1)). This finding was consistent with a number of previous empirical findings, most notably those of Hunt and Hovland (1960), Welles (1963), and Conant and Trabasso (1964). Feldman (2000) tested the parity phenomenon for an unprecedentedly large number of category types. However, Feldman's Boolean complexity model presented in the same paper does not account for the phenomenon.

In general, the CIM predicts this learnability advantage. For example, for a type 2(3) category, CIM predicts a value of 1.54 while for a type 2(1) category it predicts a value of 1, making the 2(3) type category more structurally complex and hence more difficult to learn. This is consistent with the aforementioned findings. In fact, for all 76 category types (35 in up parity, 35 in down parity, and 6 with no parity) the CIM accounts for 42% of the variance and correlates more significantly with the empirical data (Spearman $\sigma = -.69$). This is higher than the predictive performance attained by Feldman's complexity model (Feldman, 2000) for only the 41 category types (35 in up parity and 6 with no parity).

In contrast, Feldman's λ^T spectral decomposition model of complexity (Feldman, 2006) discussed under section 3.5 accounts

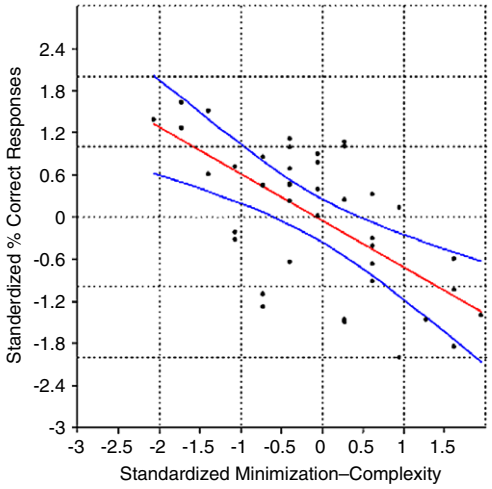


Fig. 5.3. Proportion correct for 41 category types studied by Feldman (2000) plotted against strong Boolean complexity predictions. These include the 3(2), 3(3), 3(4), 4(2), 4(3), and 4(4) families. The center line indicates linear regression: $R^2 = .42$. The flanking curves define the 95% confidence interval.

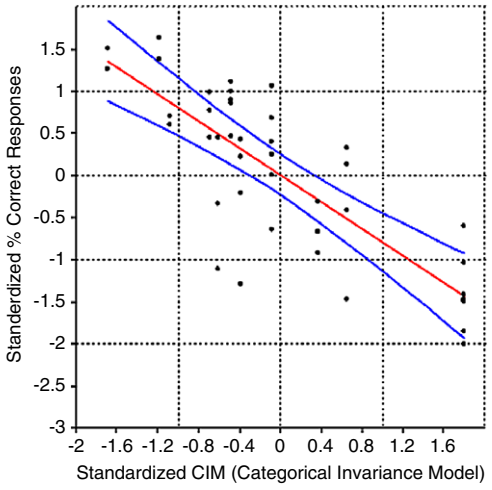


Fig. 5.4. Proportion correct for 41 category types studied by Feldman (2000) plotted against invariance law predictions of the degree of learning difficulty. These include the 3(2), 3(3), 3(4), 4(2), 4(3), and 4(4) families. The center line indicates linear regression: $R^2 = .64$. The flanking curves define the 95% confidence interval.

for approximately 50% of the variance in the data. Both the CIM and the λ^T model are able to predict the SHJ learning difficulty ordering. However, as already explained, the spectral decomposition model depends on weights per level of decomposition to do its work (i.e., for each K, the number of rules in a set of rules of degree K is assigned a weight); on the other hand, the CIM in its basic form does not depend on any kind of weights.

Moreover, Feldman (2006) attempted to fit Krusche's ALCOVE (1992) to the same data set. While ALCOVE does not have a mechanism for dealing with parity (in fact, models like ALCOVE are symmetric in respect to parity), it nevertheless predicts the SHJ difficulty ordering. According to Feldman, ALCOVE yields an R^2 of .21 for all 76 category types.

While the spectral model explains the degree of learning difficulty in terms of the regularities inherent in Boolean categories as measured by the sum of the weighted components of their spectral decomposition, the CIM explains the same regularities in terms of degrees of structural symmetry or invariance and we believe that it does so in a way that intuitively suggests the cognitive process underlying the computation. In this respect, unlike Garner (1974), we believe that invariance or symmetry offers a natural and *direct* way to measure the regularities in a pattern.

Interestingly, both models rely on some type of discretization of concepts found in the calculus: as Feldman (2006) pointed out, the λ^T measure is analogous to a series expansion in analysis. On the other hand, our invariance law is based on a more elaborate mathematical framework involving concepts that are hybrids from analysis, Boolean algebra, and discrete topology.

One interesting aspect of the CIM is that its reliance on both categorical invariance and category size tacitly suggests two distinct processes that may lie at the core of all of cognition: namely, pattern detection and working memory. This is most apparent when considering the parity phenomenon. More specifically, we hypothesize that parity effects may be the result of observers searching for invariances (perhaps using short-term or working memory and selective attention resources) within competing categories and settling for the one from which it is easiest to extract invariances in working memory. This is normally the smaller category unless the bigger category has a comparatively much higher degree of invariance than the smaller.

However, this maxim works only up to a point: for even when there is a higher degree of invariance in a competing category, subjects probably opt to commit to long-term memory the competing category that has only one member, thereby bypassing the implicit analysis of invariance altogether. This tradeoff between memory limitations and pattern perception needs to be further researched empirically. However, a good first theoretical step toward identifying it is to characterize the role that both invariance and category size may be playing in concept formation and this is what we have done thus far through our characterization of structural complexity.

We could also, at the risk of introducing additional qualitative assumptions, describe this relationship in terms of an exponential functional of invariance. Doing so eliminates the need for normalizing $\Phi(\mathbf{F})$ by adding the number one. Under such an interpretation we get Definition 5.2. With this move, a conceptual unity between our model and Shepard's model of generalization emerges. The two models have similar forms, but Shepard's model depends on a similarity metric between stimuli in "psychological space" which determines the probability (or rate) of generalization from a stimulus to different test stimuli, while our model depends on an invariance metric in structural space for sets of stimuli (i. e., categories) which determines the degree of structural complexity of a categorical stimulus explicitly (and the rate of correct classifications implicitly). This conceptual unity is another strength of the CIM.

Definition 5.2. The degree of structural complexity $\Psi(\mathbf{F})$ of a Boolean category defined by the Boolean expression $\mathbf{F}$ and belonging to the family $D(p)$ is indirectly proportional to the exponent of its degree of categorical invariance $\Phi(\mathbf{F})$ and directly proportional to its cardinality $|\mathcal{D}_1(\mathbf{F})| = p$.

$$\Psi(\mathbf{F}) = |\mathcal{D}_1(\mathbf{F})| \times e^{-\Phi(\mathbf{F})} = pe^{-\Phi(\mathbf{F})}.$$

This alternative definition of structural complexity performs well with a Spearman σ of $-.65$ and an $R^2 = .58$ for the 41 Boolean category types. As usual, predictions from this exponential model can be improved upon by adding parameters. Note that the ratio between cardinality and invariance in both definitions reflects the tradeoff between memory capacity and pattern perception discussed above.

Thus, now we can offer a possible answer the following question: why should minimization or compressibility of Boolean rules be a meaningful measure of Boolean complexity in the first place? Our answer is that Boolean rules are simply higher-order explicit symbolic approximations of the implicit detection

of invariances intrinsic to stimuli. Admittedly, our analysis above has been limited to 41 Boolean category types in six families. This seemed appropriate since these families are computationally, and more importantly, cognitively tractable—as empirical studies have shown. However, studying a much larger number Boolean category families will likely improve our understanding of the relationship between these two very different ways of measuring complexity.

5.3. Generalization to non-human animals

Todd Gureckis (personal communication) has questioned whether the CIM could successfully model Boolean concept learning data obtained from non-human animals such as monkeys or birds. For example, in one recent study by Smith, Minda, and Washburn (2004), four Rhesus monkeys were trained for 36 days to respond to a monitor displaying visual instantiations of the SHJ category types using a joystick with fruit pellets as rewards. Each solved all six SHJ problems in multiple orders. The resulting difficulty ordering was significantly different from that observed in humans, with concepts associated with type II Boolean categories appearing to be approximately as difficult to learn as concepts associated with type VI Boolean categories for each of the four monkeys.

By extending the CIM with the inclusion of dimensional bias or saliency parameters δ_i ranging over real numbers (where $0 \leq \delta_i \leq 1$ and i is a positive integer representing the particular dimension), we get the following expression which makes structural complexity a function of the degree of invariance, dimensional saliency, and cardinality:

$$\Psi(\mathbf{F}) = |\mathcal{D}_1(\mathbf{F})| \times \left[\left[\sum_{i=1}^{D} \left(\delta_i \cdot \left\| \frac{\widehat{\partial} F(x_1 \ldots x_D)}{\widehat{\partial} x_i} \right\|^2 \right) \right]^{1/2} + 1 \right]^{-1}.$$

For example, letting $\delta_3 = 0$ (and $\delta_1 = \delta_2 = 1$) decreases the invariance contributed by the dimension that is represented by the variable x_3 (or z in the representation of the 3(4) category types in Fig. 2.2 of Section 2) and predicts the order of learning difficulty experienced by the monkeys in Smith et al. (2004). This can be easily confirmed by examining the lexicographical order of the logical manifolds corresponding to the SHJ category types (see Fig. 3.7). Of course, depending on which dimension is discounted, one can generate different orders.

The cognitive motivation behind the CIM extension given in the above definition is that the invariance detection mechanism of some non human animals may not be sufficiently developed to handle more than two dimensions. If there is a high correlation between the limitations of working memory and our ability to detect invariances in categories, then perhaps the reason why non-human animals such as Rhesus monkeys, and even children, do not perform as well as adults is because they do not have the three or four chunk capacity of working memory required to manipulate three or four dimensions implicitly in psychological space.

Indeed, it may be that each dimension requires at least one chunk of working memory capacity in order to be manipulable in working memory. Moreover, it is plausible that Rhesus monkeys are disregarding the one dimension that is most disruptive to the overall invariance of the category. If proven correct, this would seem to indicate that the perceptual and concept learning mechanisms of some non-human animals satisfy some principle of parsimony. In fact, it has been argued that other higher-order cognitive mechanisms in non-humans animals (in particular, deduction and similarity judgments) may also be regulated by some principle of parsimony (Vigo, in press; Vigo & Allen, 2009).

6. Conclusion

In the introduction to this paper, we specified two clearly distinct aims. The first was to develop a simple and natural mathematical framework with which to precisely describe and measure the degree of invariance inherent to Boolean categories: these invariances were expected to account for fundamental aspects of Boolean concept learning behavior. The general plan was to translate topological notions of invariance into discrete logical notions of invariance.

To do this, we introduced the notion of a logical manifold as an ordered set of the L-norms of the partial derivatives of the Boolean category (i.e., a measure of the Boolean category gradient). We then identified the invariance space $\mathcal{L}$ of a Boolean category family $D(p)$ as the logical manifolds of the category types in the family. In addition, we proposed a measure of relative degrees of invariance based on an Euclidean metric on the logical manifolds of these category types. Combined, we call these ideas "logical manifold theory". This is, as far as we know, the first time this type of logico–analytic framework has been used to model cognition.

The second aim was to developed a theory of human concept learning based on the idea that the subjective degree of learning difficulty of a category is determined by an invariance-based measure of the structural complexity of the categorical stimulus. More specifically, we defined the structural complexity of a categorical stimulus as being indirectly proportional to its intrinsic degree of invariance and directly proportional to its cardinality or size. The end result is that human concept learning difficulty is ultimately dependent and best explained, not by complexity, but by invariance principles.

It is interesting to compare this structural account with exemplar accounts of concept learning (Medin & Schaffer, 1978; Nosofsky, 1986). In these accounts, a novel stimulus is classified by determining how similar it is to the stored exemplars of a category and those of a contrasting category. In particular, in Nosofsky's GCM (1986) model of categorization, categorization decisions are based on the similarity of novel stimuli to each of the exemplars in each of the competing categories.

The ratio of the sum of these similarities is a good predictor of the final classification outcome. Selective attention plays an important role in this model as does long-term memory. In the categorical invariance model, however, it is the low degree of symmetry or invariance of the categorical stimulus, along with the greater demands placed by such a stimulus on the agent's invariance detection capacity, that ultimately make a category hard to learn and a category instance hard to classify.

In other words, the cognitive plausibility of our theory relied on the idea that it may be the case that subjects *implicitly* detect patterns (among instances of categories) via a higher-order differentiating mechanism in respect to particular dimensions. This interpretation is consistent with our modeling proposal that the degrees of partial invariance intrinsic to a Boolean category are revealed by its logical manifold. We then conjectured that two aspects of the logical manifold of a Boolean category could be used to determine an invariance order among category types. The first is the maximal invariance value in the logical manifold of the category type, the second the frequency of this value. This ordinal characterization paved the way towards our quantitative measure of invariance.

We used the invariance measure along with information about the size of a category to build the CIM (concept invariance model). The CIM was successful in predicting the SHJ ordering as observed in Shepard et al. (1961) and Nosofsky et al. (1994). More specifically, according to the CIM, categories of type 3(4)-I are the easiest to learn since they contain the maximal amount of invariance, while categories of type 3(4)-II are a more difficult.

Categories of type 3(4)-III, 3(4)-IV, and 3(4)-V are of approximately equal difficulty and all more difficult than categories of type 3(4)-II. Finally, categories of type 3(4)-VI are the most difficult of all for they lack the most invariance. In a nutshell, we can describe the difficulty ordering predicted by the CIM by the following sequence: I(1.66) < II(2) < [III(2.14)], IV(2.14), V(2.34)] < VI(4).

More importantly, the CIM was able to predict the approximate difficulty ordering for all 41 category types studied by Feldman (2000) without the need for free parameters. In fact, the CIM was able to account for approximately 64% of the variance ($R^2 = .64$) in the data. Category size alone only accounts for 27% ($R^2 = .27$) of the variance. Applying the Spearman σ test, which measures the correlation between the predicted order of the 41 category types and their empirical order, yields an impressive $-.80$. In contrast, the minimization-complexity model did not show as strong a correlation, with a Spearman σ of $-.60$ for the 41 category types.

Moreover, when considering all 76 category types (35 in down parity, 35 in up parity, and 6 with no parity) studied by Feldman (2000) the CIM accounted for 42% of the variance in the data. This time, category size alone accounts for 24% of the variance ($R^2 = .24$). In contrast, the strong Boolean complexity account of learning difficulty (based on Boolean minimization) accounts for 42% of the variance in the data for categories in up parity only (the weak complexity model accounts for 45% of the variance), while Feldman's λ^T spectral decomposition model of complexity accounts for about 50% of the variance using multiple weights.

We also discussed how the more elegant exponential version of the CIM (Definition 5.2) is also effective and suggests a conceptual unification between our invariance measure and Shepard's model of generalization. More specifically, if we construe invariance as a higher-order contextual similarity measure, it is not unreasonable to think that there may be a connection between a contextual law of generalization and our notion of structural complexity. It remains to be seen which of these two versions of the CIM will provide a better fit to data from future empirical studies.

In addition, we showed how the CIM may be extended in a simple and natural fashion (by adding a dimensional saliency variable) in order to account for the learning difficulty ordering of the SHJ category types as observed in Rhesus monkeys (Smith et al., 2004). This extension is far from arbitrary as it is motivated by the limits of working memory in different types of non-human animals and in small children. However, it introduces parameters into the model.

Admittedly, the account of human concept learning put forth above involves only categorical stimuli consisting of an arbitrary number of binary-valued dimensions. However, one of the virtues of the current mathematical framework is its flexibility: that is, in principle, it can be extended in the direction of an arbitrary number of multivalued dimensions by increasing the power of definability of the Boolean algebra via an expansion of its carrier set. One of the future research challenges for the current work is to develop such an extension.

Finally, we hope that the modeling techniques and theory proposed in this paper will stimulate further research in the fields of conceptual behavior and concept learning. Also, that they will encourage the utilization of invariance concepts as unificatory principles in other areas of cognitive research. Such efforts can only help lift cognitive science to the same level of systematicness and rigor found in the physical sciences.

Acknowledgments

I would like to thank John Kruschke, Colin Allen, James Townsend, Robert Nosofsky, and Rob Goldstone for their suggestions and encouragement while preparing this manuscript. This work is dedicated to the memory of my advisor and friend Alonzo Church. The truth tables in this article were generated by Truth Table Constructor 3.0, an Internet applet by Brian S. Borowski. This work was supported in part by the National Institute of Mental Health Grant R01MH48494.

Appendix

Proposition A.1.

$$\frac{\left|\mathcal{D}_{\mathrm{I}}(\mathbf{F}) \cap \mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F})\right|}{|\mathcal{D}_{\mathrm{I}}(\mathbf{F})|} = \frac{\left|\mathbf{F}_{\mathbb{O}} \cap \mathbf{F}_{\mathbb{P}}\right|}{\left|\mathbf{F}_{\mathbb{O}}\right|}.$$

Proof. We show that $\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \left[\mathcal{A}_0(\widehat{\partial}_{x_i}\mathbf{F}) = \mathcal{A}_0(\mathbf{F}_{x_i} \oplus \mathbf{F}_{x_i'})\right] = \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})$ since by definition $\mathbf{F}_{\mathbb{O}} = \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})$ and $\mathbf{F}_{\mathbb{P}} = \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})$. $\mathcal{A}_0(\mathbf{F}_{x_i} \oplus \mathbf{F}_{x_i'}) = [\mathcal{A}_1(\mathbf{F}_{x_i}) \cap \mathcal{A}_1(\mathbf{F}_{x_i'})] \cup [\mathcal{A}_0(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i'})]$. But $\mathcal{A}_1(\mathbf{F}_{x_i}) = \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})$ and $\mathcal{A}_1(\mathbf{F}_{x_i'}) = \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})$. Thus, $\mathcal{A}_0(\mathbf{F}_{x_i} \oplus \mathbf{F}_{x_i'}) = [\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})] \cup [\mathcal{A}_0(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i'})]$. Then,

$$\begin{aligned}&[[\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})] \cup [\mathcal{A}_0(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i'})]] \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})\\ &\quad = [[\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})] \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})]\\ &\qquad \cup [[\mathcal{A}_0(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i'})] \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})].\end{aligned}$$

But $[[\mathcal{A}_0(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i'})] \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i})] = \varnothing$, since $\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{A}_0(\mathbf{F}_{x_i}) = \varnothing$. Thus, $[\mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i}) \cap \mathcal{D}_{\mathrm{I}}(\mathbf{F}_{x_i'})]$. □

References

Aiken, H. H. (1951). The Staff of the Computation Laboratory at Harvard University. In *Synthesis of electronic computing and control circuits*. Cambridge: Harvard University Press.
Bourne, L. E. (1966). *Human conceptual behavior*. Boston: Allyn and Bacon.
Conant, M. B, & Trabasso, T. (1964). Conjunctive and disjunctive concept formation under equal-information conditions. *Journal of experimental Pychology*, *67*, 250–255.
Church, A. (1956). *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.
Feldman, J. (2003a). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, *47*(1), 98–112.
Feldman, J. (2003b). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*(6), 227–233.
Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology*, *50*, 339–368.
Garner, W. R. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior*, *2*, 446–452.
Garner, W. R. (1970). Good patterns have few alternatives. *American Scientist*, *58*, 34–42.
Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225–241.
Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32.
Haygood, R. C., & Bourne, L. E., Jr. (1965). Attribute-and-rule learning aspects of conceptual behavior. *Psychological Review*, *72*, 175–195.
Higonnet, R. A., & Grea, R. A. (1958). *Logical design of electrical circuits*. New York: McGraw-Hill.
Hunt, E. B., & Hovland, C. I. (1960). Order of consideration of different types of concepts. *Journal of Experimental Psychology*, *59*, 220–225.
Imai, S. (1977). Pattern similarity and cognitive transformations. *Acta Psychologica*, *41*, 433–447.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.
Kruschke, J. K. (2006). Models of categorization. In R. Sun (Ed.). *Handbook on computational cognitive modeling*.
Lee, Michael D., & Navarro, Daniel J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, *9*(1), 43–58.
Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.
Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York: Wiley.
Marr, D. (1982). *Vision: A computational approach*. San Francisco: Freeman & Co.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–223.

Miller, G. (1956). The magical number seven, plus of minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
Leyton, M. (1992). *Symmetry, causality, mind*. The MIT Press.
Mendelson, E. (1970). *Boolean algebra and switching circuits*. New York: McGraw-Hill.
Mendelson, E. (1979). *Introduction to mathematical logic*. New York: D. Van Nostrand Company.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. G. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition*, *22*(3), 352–369.
Poggio, Tomaso (1981). *Marr's approach to vision. A.I. Memo 645*. Massachusetts Institute of Technology Artificial Intelligence Laboratory.
Reed, I. S. (1954). A class of multiple error-correcting codes and the decoding scheme. *IRE Trans. on Information Theory*, *vol. IT-4*, 38–49.
Roberts, F. S. (1979). *Measurement theory, with applications to decisionmaking, utility, and the social sciences*. Addison-Wesley, Reading, MA: Addison-Wesley.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structures of categories. *Cognitive Psychology*, *7*, 573–605.
Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.
Shepard, R. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*, 125–140.
Shepard, R. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, *39*, 373–421.
Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins tasks. *Journal of Experimental Psychology: General*, *133*, 398–414.
Thayse, A. (1981). Boolean calculus of differences. In *Lecture notes in computer science*: *Vol. 101*. New York: Springer-Verlag.
Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology*, *50*(5), 501–510.
Vigo, R. (2008). Modal similarity. *Journal of Experimental and Theoretical Artificial Intelligence* (in press).
Vigo, R., & Allen, C. (2009). How to reason without words: Inference as categorization. *Cognitive Processing*, *10*(1), 77–88.
Wang, Y., & Shao, J. (2003). A new measure of software complexity based on cognitive weights. *Canadian Journal of Electrical and Computer Engineering*, *28*(2), 69–74.
Welles, H. (1963). Effects of transfer and problem structure in disjunctive concept formation. *Journal of experimental Pychology*, *65*, 63–69.