Towards a law of invariance in human conceptual behavior

Ronaldo Vigo (vigo@ohio.edu)
CFACS-Center for the Advancement of Cognitive Science
Ohio University, Athens, Ohio, 45701, USA

Abstract

Invariance principles underlie many key theories in modern science. They provide the explanatory and predictive framework necessary for the rigorous study of natural phenomena ranging from the structure of crystals, to magnetism, to relativistic mechanics. Vigo (2008, 2009) introduced a new general notion and principle of invariance from which two parameter-free (ratio and exponential) models were derived to account for human conceptual behavior. Here we introduce a new parameterized exponential "law" based on the same invariance principle. The law accurately predicts the subjective degree of difficulty that humans experience when learning different types of concepts. In addition, it precisely fits the data from a large-scale experiment which examined a total of 84 category structures across 10 category families ($R^2 = .97, p < .0001$; r=.98, $p < .0001$). Moreover, it overcomes seven key challenges that had, hitherto, been grave obstacles for theories of concept learning.

Keywords: Concepts; concept learning; categorization; law of invariance; mathematical model; pattern perception; ideotype.

Introduction

One of the long standing goals of psychological science has been to discover the laws that govern human conceptual learning behavior and, in particular, to describe these with the mathematical precision and rigor commonly found in the physical sciences. The two main representational paradigms of concepts -- concepts as exemplars and as prototypes -- and their respective formal models, such as the generalized context model (Nosofksy, 1984, 1986), ALCOVE (Kruschke, 1992), and the multiplicative prototype model (Estes, 1986) have contributed significantly toward this goal (Nosofsky, 1991; Nosofsky et al., 1994; Kruschke, 2006). But these models, mostly of the probabilistic variety, have not been able to account for the learnability of large classes of Boolean category structures (Feldman, 2006). One of the reasons is that they do not fully capture key relational and contextual information in sets of stimuli, and how this information plays a role in determining how hard or easy it is to learn a concept (Gibson, 1966; Garner, 1963, 1970a, 1970b, 1974). For example, an abundance of laboratory experiments have supported the premise that subjects extract rules from perceived patterns in stimulus sets (i.e., sets of objects) from which concepts are learned (Bourne, 1966; Estes, 1994; Murphy, 2002). However, in spite of this understanding, the development of a mathematically precise and elegant relational principle of concept learning that is able to reveal the nature of pattern detection with respect to sets of stimuli, and that is sufficiently general to accurately predict the degree of learning difficulty of a wide range of category structures, remains an open problem. Instead, alternative accounts have emerged which place mediating constructs at their core. One such account (Feldman, 2000) posits that since humans report forming rules when performing laboratory categorization tasks, one can then measure the degree of concept learning difficulty associated with a stimulus set by the length of the shortest logical rule that defines it. This proposal, referred to as "minimization complexity", does not answer two key questions about concept learning as a rule-oriented process: 1) what is the nature of the relational pattern perception process that must precede (and that is necessary for) the formation of efficient rules and heuristics in the first place, and 2) what are the limits of our capacity to detect such relational patterns? We believe that the answers to these two questions are the key to explaining and predicting a wide variety of phenomena associated with classification performance. In other words, rule simplification procedures based on Boolean logic should, but do not, give a deep rationale for why it is easier to form rules about certain sets of stimuli but not about others. Such a rationale is necessary to better understand why categorization performance is often inconsistent with rule-based accounts of concept learning (Vigo, 2006; Lafond, 2007). In this report, we propose an invariance principle and law of invariance (LOI) as the answer.

In what follows, we shall discuss how the data from a large-scale human categorization experiment by the author and data from several classic experiments on concept learning can be directly accounted for by a simple mathematical law. The law is based on the assumption that humans learn concepts by applying a differential (analytic) operator to stimulus sets in order to optimize their classification performance. The operator generates an *ideotype* or higher level memory trace of the essential or "atomic" structural patterns (referred to henceforth as the "structural kernels" or SKS) perceived in the stimulus set. Ideotypes are represented by points in a high-level psychological space whose coordinates are the values of their SKS. Varying sensitivity to these SKS (and to the ideotypes in general) can accounts for individual

differences in classification performance. Although, throughout this brief report, we shall offer hints about the process or algorithmic level theory of concepts as ideotypes, our main aim is to propose a goal-oriented high-level mathematical description (what David Marr referred to as a *computational theory*) of conceptual behavior. In Marr's computational theory of the human visual system, the Laplacian differential operator plays a role that is similar to the role played by our own differential operator (the *structural manifold operator*) in facilitating the assumed goals of the human conceptual system (likewise, one might say that the *ideotype* of a stimulus set is a higher level cognition counterpart of the *primal sketch* of a visual stimulus). We shall focus on only two, but important, goals of our conceptual system: the first is to generate and supply key information about the "diagnosticity" and "redundancy" of the recognized dimensions in the stimulus set to a rule-construction subsystem, and the second goal is to classify exemplars from the stimulus set optimally. Note that Vigo (2011) generalized the invariance law proposed in this report in order to account for the learnability of ill-defined concepts. This generalization is achieved under a process account of SK detection featuring notions of high-level similarity assessment and goal-directed attention. Unfortunately, its details are beyond the scope of this brief report. Notwithstanding, the aforementioned theoretical assumptions offer an adequate conceptual sketch for interpreting and predicting individual differences as encoded in the parameters of the invariance law.

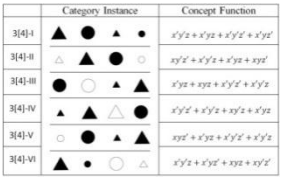

Figure 1 Instances of the 3[4] category types studied by Shepard et al. (1961) where $\boldsymbol{x}$ represents the color dimension, $\boldsymbol{y}$ represents the shape dimension, and $\boldsymbol{z}$ represents the size dimension.

As we shall see, the LOI overcomes seven challenges that have been stumbling blocks for theories of Boolean concept learning: 1) it perfectly predicts the key 3[4] family (Figure 1) learning difficulty ordering as shown in Figure 2 below; 2) it is able to accurately account for the learnability of categories in both up and down parity; 3) it accurately accounts for the learnability of a very large class of category types and families ($R^2 = .97,\ p < .0001$; 4) it does so without the need for free parameters ($R^2 = .70$, p<0.0001), 5) through the use of well-motivated and cognitively meaningful parameters, it can explain individual differences in classification performance; 6) it introduces an original mathematical and deterministic framework for the study of concept learning behavior; 7) it unifies in precise quantitative terms key and ubiquitous constructs in universal science such as symmetry, invariance, and complexity from the perspective of concept research. No other formal model of concept learning behavior has accomplished all of the above.

The Mathematical Law of Invariance

Most investigations pertaining to degree of concept learning difficulty have focused on sets of stimuli that are defined by Boolean algebraic rules (i.e., expressions consisting of disjunctions, conjunctions, and negations of variables that stand for binary dimensions). These algebraic representations of a categorical stimulus (category of objects) or stimulus set are referred to as *concept functions*. Concept functions are useful in spelling out the logical structure of a stimulus set. For example, suppose that x stands for blue, x' stands for red, y stands for round, and y' stands for square, then the two-variable concept function $(x' \cdot y) + (x \cdot y')$ (where "$+$" stands for "or", "$\cdot$" stands for "and", and "x'" stands for "not-x") defines the category of "red and round or blue and square" objects. Clearly, the choice of labels in the expression is arbitrary. Hence, there are many Boolean expressions that define the same category structure. For example, making x stand for red instead of blue yields the structurally equivalent category of "blue and round or red and square" objects, where the relationships between the dimensional values remain the same. These structurally equivalent categories form category types (or distinct structures) and may be represented by a canonical concept function in disjunctive normal form or DNF (informally, a concept function in DNF is simply a function that is a verbatim description of the entire category content just like the function given above). A class of Boolean category types whose category instances are defined by *D* dimensions and contain *p* objects is called a *D*[*p*] family. For instance, the Boolean category described above belongs to the 2[2] family since it is comprised of two objects that are defined by two dimensions (color and shape). Every category family has a fixed number of category types. For example, the 3[4] family has six category types (for a proof see Higonnet et al., 1958). This latter family was studied empirically by Shepard et al. (1961) who observed the following increasing learning difficulty ordering: I < II < [III, IV, V] < VI (with types III, IV, and V of approximately the same degree of difficulty). The degree of learning difficulty of a category type is typically operationalized by the percentage of errors made by a subject while attempting to classify the objects from the stimulus set that is an instance of the type. Figure 1 above illustrates visual instances of the 3[4] family types in the form of simple geometric shapes. This 3[4] family ordering has been empirically replicated numerous times by several

researchers (Shepard et al., 1961; Kruschke, 1992; Nosofsky, 1994; Love & Medin, 1998) but has been difficult to predict quantitatively.

3[4] Family	Standardized Proportion of Errors	Standardized IL Predictions
Type I	-1.3	-1.3
Type II	-0.9	-0.9
Type III	0.3	0.3
Type IV	0.2	0.2
Type V	0.1	0.1
Type VI	1.6	1.6

Figure 2 LOI standardized predictions for the 3[4] stimulus types using data from the experiment by the author ($\boldsymbol{R^2}$=1, p<.0001) using a single scaling parameter *k* estimated for all six types.

In a more recent and broader study, Feldman (2000) observed an approximate empirical difficulty ordering for 76 category types from the 3[2], 3[3], 3[4], 4[2], 4[3], and 4[4] families along with their "down parity" counterparts. A category is in down parity whenever it has more objects than its complementary category; otherwise, it's in "up parity" (the complement of a category is the set of objects that are also definable by *D* dimensions but that are not in the category). Although a difficulty ordering was observed for the aforementioned 76 types, the classic 3[4] family ordering discussed above was not observed by Feldman. In our study (described briefly under the methods section), we extended these same six families, by adding the 2[1], 2[2], 3[1], and 4[1] families (for a total of 84 category types in up and down parity across 10 families). The 2[1] and 2[2] families were tested because they were studied extensively in the 1960s (Hunt et al., 1960; Welles, 1963; Haygood et al., 1965). In our study, we measured the "subjective degree of learning difficulty" of each category type by computing the average percentage of classification errors made by subjects when attempting to classify members of its instances. As expected, we observed the classic 3[4] family difficulty ordering.

To understand how the LOI accounts for the learnability of the above category structures, consider a simple example. The stimulus set containing a triangle that is black and small and a circle that is black and small and a circle that is white and large which is described by the concept function $xyz + x'yz + x'y'z'$ (note that, for readability, we have eliminated the symbol "·" representing "and"). Let's encode the features of the objects in this category using the digits "1" and "0" so that each object may be representable by a binary string. For example, "111" stands for the first object when *x*=1=*triangular*, *y*=1=*small*, and *z*=1=*black*. Thus, the entire set can be represented by $C = \{111,\ 011,\ 000\}$. If we perturbed this stimulus set with respect to the shape dimension by assigning the opposite shape value to each of the objects in the set, we get the perturbed stimulus set $T_1(C) = \{011,\ 111,\ 100\}$ which indicates a transformation along the first dimension (in general, $T_i(C)$ stands for the category *C* transformed along the *i*-th dimension). Now, if we compare the original set to the perturbed set, they have two objects in common with respect to the dimension of shape. Thus, two out of three objects remain the same. This ratio is a measure of the partial homogeneity of the category with respect to the dimension of shape and can be written more formally as $\Lambda_i(C) = |C \cap T_i(C)| / |C|$. Here, $|C|$ stands for the number of objects in the category and $|C \cap T_i(C)|$ for the number of objects that they share (Vigo, 2009).

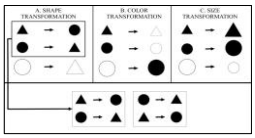

Figure 3 Structural manifold transformations across the dimensions of shape, color, and size for a 3[3] family category instance.

The first pane of Figure 3 illustrates this transformative process. Doing this for each of the dimensions, we can generate the SKS of the stimulus set (represented by the Boolean category C) with the Λ operator (where *D* is the number of dimensions in C):

$$(1.1)\ \Lambda(C) = \left(|C \cap T_1(C)| / |C|, |C \cap T_2(C)| / |C|, \ldots, |C \cap T_D(C)| / |C|\right)$$

Note that equation 1.1 does not define the transformation T_i and hence, does not specify how to compute $|C \cap T_i(C)|$ (for any i). In fact, thus far, I have only described the transformation T_i in non-mathematical terms. In Vigo (2009) this "extraction" of the SKS of a stimulus set is achieved with a mathematically precise and generalizable definition of Λ as a partial differential operator on concept functions which, by its very nature, mathematically defines the role and nature of T_i. This is expressed in equation 1.2 below where $\Lambda(F)$ stands for the structural manifold of the concept function *F* and where a "hat" symbol over the partial differentiation symbol indicates discrete differentiation (for an explanation of the equivalence of equations 1.1 and 1.2 below, see Vigo (2009) or the technical appendix of this note).

$$(1.2)\quad \Lambda(F) = \left(\left\|\frac{\hat{\partial}F(x_1,\cdots,x_D)}{\hat{\partial}x_1}\right\|, \left\|\frac{\hat{\partial}F(x_1,\cdots,x_D)}{\hat{\partial}x_2}\right\|, \cdots, \left\|\frac{\hat{\partial}F(x_1,\cdots,x_D)}{\hat{\partial}x_D}\right\|\right)$$

Please note that: 1) 1.2 above is not the gradient operator (see technical appendix) and 2) applying the structural manifold operator to a concept function is *not equivalent* to factoring out variables from the concept function formulae in DNF that define the category structures. To recognize this, note that the variables of the sixth and last concept function in the table of

Figure 1 may be factored out in several ways: yet, the degree of invariance of the concept function is zero. It is also important to recognize that the components of the structural manifold which reveal the patterns of invariance in the stimulus set are partial measures of its homogeneity. Accordingly, the perceived relative degrees of total homogeneity across category types from different families can then be measured by taking the Euclidean distance of each structural manifold (equation 1.1) from the zero structural manifold whose components are all zeros (i.e., **0**=(0,…,0)). Thus, the overall degree of invariance (or homogeneity) Φ of the concept function *F* (and of any stimulus set that it defines) is given by the equation below (where $\vec{F}$ is the stimulus set defined by *F*):

$$(1.3)\quad \Phi(F) = \left[\sum_{i=1}^{D}\left[\left\|\frac{\partial F(x_1,\ldots,x_D)}{\partial x_i}\right\|\right]^2\right]^{1/2} = \left[\sum_{i=1}^{D}\left[\frac{|\vec{F}\cap T_i(\vec{F})|}{|\vec{F}|}\right]^2\right]^{1/2}$$

Using our example from pane one in Figure 3, we showed that the original stimulus set and the perturbed stimulus set have two elements in common (out of the three transformed elements) in respect to the shape dimension; thus, its degree of partial invariance is expressed by the ratio $2/3$. Conducting a similar analysis in respect to the dimensions of color and size, its logical manifold computes to $\left(\frac{2}{3},\frac{0}{3},\frac{0}{3}\right)$ and its degree of categorical invariance is:

$$(1.4)\quad \Phi(xyz + x'yz + x'y'z') = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{0}{3}\right)^2 + \left(\frac{0}{3}\right)^2} = .67$$

But how does invariance help us understand concept learning? The proposed mathematical theory describes the goal of our conceptual system as being that of the extraction or detection of SKS in the stimulus set in ways that optimize classification performance: in particular, in ways that generate information regarding the redundancy and diagnosticity of its dimensions for the purpose of constructing efficient membership rules and for the purpose of assessing degree of homogeneity and degree of learning difficulty. To illustrate, consider the partial symmetry shown in the bottom pane of Figure 3. This symmetry is revealed when the structural manifold operator is applied to the stimulus set in the top pane of Figure 3. Identifying these partial symmetries allows our conceptual system to determine the diagnostic value of each dimension in that the more symmetries that are detected, the less the associated dimension is useful in determining category membership. In other words, the dimensions associated with high invariance do not help us discriminate the perturbed objects from the original objects in terms of category membership. Consequently, these particular dimensions do not carry "diagnostic" information about their associated category; however, they signal the presence of redundant information that is eventually eliminated. Again, note that the ratio between the number of qualitative symmetries and the number of objects in the stimulus with respect to a particular dimension (i.e., the value of the SK) is a measure of the partial homogeneity of the stimulus set. Due to their great utility in forming efficient rules, the SKS that our conceptual system should be most sensitive to are those that have value 0. In-between valued SKS should play a relatively lesser, but important role in determining "in-between rules". We assume that this implicit heuristic drives the classification process. As a consequence, the LOI should be able to predict that, because it has the most instrumental importance in maximizing classification performance (from the standpoint of the assumed goals of the conceptual system), non-redundant information is emphasized by the human conceptual system. Also, we assume that performance gains will be disproportionately smaller as homogeneity detection increases because most of the information needed to classify efficiently is supplied by a relatively few SKS that equal to zero. This is consistent with the trend of the data per category family tested in our current experiment, which indicates that the degree of subjective learning difficulty of a category type (as measured by the proportion of errors in the classification tasks) decays in a non-linear monotonic fashion (likely exponential) as a function of its degree of invariance. Using the above description of the invariance pattern detection process, a simple mathematical law of conceptual behavior emerges: namely, that the degree of subjective learning difficulty ψ of a stimulus set $\vec{F}$ defined by a concept function *F* is directly proportional to its cardinality or size and it is indirectly proportional to the exponent of the degree of invariance of the concept function *F* that defines it. This relationship is expressed formally by the *parameter-free* equation in 1.5 below.

$$(1.5)\qquad \psi\left(\vec{F}\right) = pe^{-\Phi(F)} = pe^{-\left[\sum_{i=1}^{D}\left[\left\|\frac{\partial F(x_1,\cdots,x_D)}{\partial x_i}\right\|\right]^2\right]^{1/2}}$$

Although the above equation, as seen in column 2 of Figure 4, accurately fits the data, the law may be further generalized with the judicious use of cognitively motivated parameters as shown in equation 1.6 below. While less parsimonious, the parameterized version can account for individual differences in concept learning performance and can further our understanding of the role that invariance pattern information plays in the concept learning process.

$$(1.6)\qquad \boldsymbol{\psi\left(\vec{F}\right) = pe^{-k\left[\sum_{i=1}^{D}\left[\alpha_i\left[\left\|\frac{\partial F(x_1,\cdots,x_D)}{\partial x_i}\right\|\right]\right]^s\right]^{\frac{1}{s}}}}$$

In equation 1.6 above, the scaling parameter α_i stands for the degree of sensitivity to the SK associated with dimension *i*. In

the current study, this is a number in the closed real interval [0, 1] so we assume that sensitivity to each SK is a non-distributed resource. The value of α_i is a function of attention and high-level similarity processes (see Vigo, 2011 for an explanation). From the classification data from our experiment we determined that, in general, the optimal values for α_i are consistent with our hypothesis that humans are most sensitive to SKS that identify the diagnostic dimensions of the stimulus set: in other words, the SKS with value zero. The scaling sensitivity parameter k $(0 \leq k < \infty)$ indicates the overall degree of discriminability between the ideotypes and the standard ideotype represented by **0** (i.e., zero invariance) in the higher order psychological space. Stimulus sets in down parity should result in higher k values due to their corresponding greater variety of ideotypes. Parameter estimates for stimuli in down parity confirm this. Accordingly, k also indicates an increase in stimulus exposure. Indeed, estimates of this parameter using our data for the 3[4] family were higher than those of data from the Feldman experiment (2000) where subjects were exposed to the stimulus set for 25% less time. Finally, s is a parameter that indicates the most appropriate measure of distance as defined by the generalized Euclidean metric (i.e., the Minkowski distance measure). In our investigation, the best predictions are achieved when s=2. Optimal estimates of these free parameters on the aggregate data using the gradient descent method provide a baseline to assess any individual differences encountered in the pattern perception stage of the concept learning process.

Fitness and Robustness

The parameter-free variant of the LOI fits our data very accurately, accounting for about 70% of the variance by removing three outliers ($R^2 = .70$, p<0.0001; r=.84, p<0.0001). The parameterized version, however, can account for individual differences. The parameterized version, with the use of optimal values for k and α_i (as computed by the gradient descent method) accounts for 97% of the variance in the data ($R^2 = .97$, $p < .0001$; r_s=.98, $p < .0001$) when the parameters are estimated on a per family basis and for 95% of the variance using only k. Moreover, it accounts for about 99% of the variance when the parameters are estimated on a per stimulus type basis, and for 87% of the variance when the parameters are estimated across all types (dimensional-level estimates). Figures 4 and 5 above summarize these results. In contrast, with optimal values for all of its parameters, the Generalized Context Model (Nosofsky, 1984) accounts for about 27% of the variance using dimensional-level estimates, and for much less without parameters. Other leading models also tested do not perform nearly as well.

	$pe^{-\Phi(F)}$	$pe^{-k\Phi(F)}$	$pe^{-k\Phi_{\alpha_i}(F)}$
84V-F	.70/.84	.95/.98	.97/.99
76F-F	.50/.71	.76/.87	.82/.91
84V-T	.70/.84	.98/.99	.99/.99
76F-T	.50/.71	.96/.98	.96/.98
84V-D	.70/.84	.70/.84	.87/.93
76F-D	.50/.71	.70/.84	.76/.87

Figure 4 Approximate $\boldsymbol{R^2}$s and correlations ($\boldsymbol{R^2/r}$) for the LOI using data from the author's study (84V) and Feldman's study (76F). The first variant of the law has no parameters, the second uses only *k*, the third uses *k* plus alphas. F, T, and D stand for family-level, type-level, and dimensional-level estimates.

In this note, we have argued that, as in the physical sciences, a mathematically precise and general invariance principle can be useful in understanding the nature and limits of human cognition. That such a simple principle can, in deterministic terms, serve as the basis for a mathematical law that explains and predicts key aspects of our concept learning behavior is testimony to the parsimony and structural unity of all natural phenomena, whether physical or mental in nature.

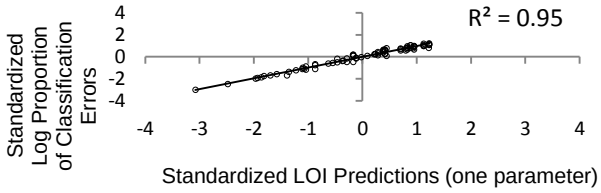

Figure 5 Classification performance predictions (for the 84 types tested) made by the exponential law of invariance using only the scaling parameter *k* (*k* was estimated on a per family basis).

Methods Sketch

Instances of category structures from the 10 tested category families were displayed as sets of one to four flasks (i.e., flat bottles) with two (color and shape), three (color, shape, and size), or four dimensions (color, size, shape, and neck width). Each target set of flasks and its complement (the set of flasks not in the target set) were displayed on a computer screen above and below a line (respectively) for a period of 20 seconds. After this training period, subjects were presented once with each flask (one at a time and at random) from the two sets combined. Subjects were given three seconds to press either a button labeled "yes" or a button labeled "no" indicating whether or not the displayed flask belonged in the target category. After each block of classification trials, a new category type from the tested families was generated and displayed by the program at random. The following 10 families, along with their down parity counterparts, were tested: 2[1], 2[2], 3[1], 3[2], 3[3], 3[4], 4[1], 4[2], 4[3], and 4[4] (a total of 84 types represented by no less than 4 instances each). For testing purposes, the families were grouped as follows: (2[1], 2[2]), (3[1], 3[2]), (3[3], 3[4]), (4[1], 4[2]), (4[3]). The 4[4] family was divided into 3 subgroups. Thirty subjects were used to test each group. This grouping helped in limiting each experimental section to about an hour, thereby reducing noisy data due to subject

fatigue and to the confusability introduced by stimulus sets of mixed dimensions. The program recorded the percentage of classification errors per block of trials.

Bourne, L. E. (1966). Human conceptual behavior. Boston: Allyn and Bacon.
Estes, W. K. (1994). Classification and Cognition. Oxford Psychology Series, 22, Oxford University Press, Oxford.
Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. Nature, 407, 630-633.
Garner, W. R. (1970). Good patterns have few alternatives. American Scientist, 58, 34-42.
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.
Haygood, R. C., & Bourne, L. E., Jr. (1965). Attribute-and-rule learning aspects of conceptual behavior. Psychological Review, 72, 175-195.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, 99, 22-44.
Love, B. C., and Medin, D. L. (1998). SUSTAIN: A model of human category learning. Proceedings of the Fifteenth National Conference on Artificial Intelligence, 15, 671-676.
Murphy, G. L. (2002). The big book of concepts. MIT Press.
Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10(1), 104-114.
Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. Memory and Cognition, 19(2), 131-150.
Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. Psychological Monographs: General and Applied, 75(13), 1-42.
Vigo, R. (2006). A note on the complexity of Boolean concepts. Journal of Mathematical Psychology, 50(5), 1-10.
Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. Journal of Mathematical Psychology, Vol. 53, 203-221.
Vigo, R. (2011). Towards a General Law of Human Conceptual Behavior (under review, contact the author for a copy)

Acknowledgements

I would like to thank John Kruschke, Andrew Halsey, and Derek Zeigler for their helpful comments.

Technical Appendix

Vigo (2008, 2009, 2011) introduced an original mathematical framework for cognitive research referred to as *logical* (or *structural*) *manifold theory*. The portion of the framework discussed here involves discrete partial derivatives. Discrete partial derivatives are completely analogous to continuous partial derivatives in Calculus. Loosely speaking, in Calculus, the partial derivative of an *n* variable function $f(x_1, \ldots, x_n)$ is defined as how much the function value changes relative to how much the input value(s) change as seen below:

$$\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1 \ldots, x_i+\Delta x_i, \ldots x_n) - f(x_1,\ldots,x_n)}{(x_i+\Delta x_i) - x_i}$$

On the other hand, the discrete partial derivative, defined by the equation below (where $x_i' = 1 - x_i$ with $x_i \in \{0,1\}$) is totally analogous to the continuous partial derivative except that there is no limit taken because the values of x_i can be only 0 or 1.

$$\frac{\hat{\partial} F(x_1, \ldots, x_n)}{\hat{\partial} x_i} = \frac{F(x_1 \ldots, x_i', \ldots x_n) - F(x_1, \ldots, x_n)}{x_i' - x_i}$$

The value of the derivative is ± 1 if the function assignment changes when x_i changes, and the value of the derivative is 0 if the function assignment does not change when x_i changes. Notice that the value of the derivative depends on the entire vector $(x_1, \ldots, x_n)$ (abbreviated as $\vec{x}$ in this note) and not just on x_i. As an example, consider the concept function AND, denoted as $F(\vec{x}) = x_1 \cdot x_2$. Also, consider the particular point $\vec{x} = (0,0)$. At that point, the derivative of the concept function AND with respect to x_1 is 0 because the value of the concept function does not change when the stimulus changes from $(0,0)$ to $(1,0)$. If instead we consider the point $(0,1)$, the derivative of AND with respect to x_1 is 1 because the value of the concept function does change when the stimulus changes from $(0,1)$ to $(1,1)$. Using the discrete partial derivative we can define a logical manifold $\Lambda(F)$ of a Boolean function *F* as follows:

$$\Lambda(F) = \left(\left\| \frac{\hat{\partial} F(x_1,\cdots,x_D)}{\hat{\partial} x_1} \right\|, \left\| \frac{\hat{\partial} F(x_1,\cdots,x_D)}{\hat{\partial} x_2} \right\|, \cdots, \left\| \frac{\hat{\partial} F(x_1,\cdots,x_D)}{\hat{\partial} x_D} \right\| \right) \tag{1.7}$$

Accordingly, the *i*-th component of the manifold of the Boolean concept function F is defined as follows:

$$\Lambda_i(F) = \left\| \frac{\hat{\partial} F(x_1, \cdots, x_D)}{\hat{\partial} x_i} \right\| = 1 - \left[\frac{1}{p} \sum_{\vec{x}_j \in \vec{F}} \left| \frac{\hat{\partial} F(\vec{x}_j)}{\hat{\partial} x_i} \right| \right] \tag{1.8}$$

In the above definition, $\vec{x}$ stands for an object defined by *D* dimensional values $(x_1, \ldots, x_D)$. The general summation symbol represents the sum of the partial derivatives evaluated at each object $\vec{x}_j$ from the Boolean category $\vec{F}$ (the set bracket over the *F* indicates that this is the category defined by the concept function *F*). The partial derivative transforms each object $\vec{x}_j$ in respect to its *i*-th dimension and evaluates to 0 if, after the transformation, the object is still in $\vec{F}$ (it evaluates to 1 otherwise). Thus, to compute the proportion of objects that remain in $\vec{F}$ after changing the value of their *i*-th dimension, we need to divide the sum of the partial derivatives evaluated at each object $\vec{x}_j$ by p (the number of objects in $\vec{F}$) and subtract the result from 1. The absolute value symbol is placed around the partial derivative to avoid a value of negative 1 (for a detailed explanation, see Vigo, 2009).