Complexity Measures and Concept Learning

Andreas D. Pape\textsuperscript{a,d}, Kenneth J. Kurtz\textsuperscript{b,d}, Hiroki Sayama\textsuperscript{c,d}

\textit{a}Corresponding Author. Department of Economics, Binghamton University, PO Box 6000, Binghamton, NY 13905. \texttt{apape@binghamton.edu}. 1-607-777-2660.

\textit{b}Department of Psychology, Binghamton University, PO Box 6000, Binghamton, NY 13905. \texttt{kkurtz@binghamton.edu}

\textit{c}Departments of Bioengineering and Systems Science and Industrial Engineering, Binghamton University, PO Box 6000, Binghamton, NY 13905. \texttt{sayama@binghamton.edu}

\textit{d}Collective Dynamics of Complex Systems Research Group, Binghamton University, PO Box 6000, Binghamton, NY 13905

\textbf{Abstract}

The nature of concept learning is a core question in cognitive science. Theories must account for the relative difficulty of acquiring different concepts by supervised learners. For a canonical set of six category types, two distinct orderings of classification difficulty have been found. One ordering, which we call paradigm-specific, occurs when adult human learners classify objects with easily distinguishable characteristics such as size, shape, and shading. The general order occurs in all other known cases: when adult humans classify objects with characteristics that are not readily distinguished (e.g., brightness, saturation, hue); for children and monkeys; and when categorization difficulty is extrapolated from errors in identification learning. The paradigm-specific order was found to be predictable mathematically by measuring the logical complexity of tasks, i.e., how concisely the solution can be represented by logical rules. However, logical complexity does not explain the general order. Here we show that a new difficulty measurement, i.e., the amount of uncertainty remaining when a subset of the dimensions are specified, can correctly predict the general order. This result contrasts with the logical-complexity-based task ordering because our proposed measurement captures statistical, not logical, complexity. This suggests that, when learners do not/cannot form logical rules about the characteristics of objects, they may be using statistical means in their category learning. It is known in information science that logical (algorithmic) and statistical (information theoretic) complexities are fundamentally linked. Our proposed statistical

\textit{Preprint submitted to Elsevier} July 1, 2014
complexity measurement, therefore, naturally complements the logical complexity one and extends the overall applicability of the complexity-based approach to understanding concept learning.

**Keywords:** Concepts, Induction, Complexity, Learning

1. Introduction

In a canonical classification learning experiment, human learners are tested on the six possible categorizations that assign eight examples (all possibilities of three binary-valued dimensions) to two equal-sized classes (Shepard, Hovland, and Jenkins, 1961). These classification problems, commonly referred to as the SHJ types, have been instrumental in the development and evaluation of theories and models of category learning. Learning is easiest for Type I in which the classes can be distinguished using a simple rule on a single dimension—e.g. all large items are category A and all small items are category B. Learning is most difficult for Type VI in which the two classes cannot be distinguished according to any set of rules or statistical regularities. The remaining types (II–V) are intermediate in difficulty. Table I provides a complete description of the six mappings.

In the original experiments and contemporary replication (Shepard et al., 1961; Nofofsky, Gluck, Palmeri, McKinley, and Glauthier, 1994), adult human learners classified objects with characteristics that can be easily distinguished, such as shading, shape, and size. These experiments yield what we call the paradigm-specific order, in which, among the intermediate types (II–V), Type II (a logical XOR rule on two dimensions) is learned faster than the types III–V, which are learned at the same speed. In an update to the traditional SHJ ordering, it has been shown through a review of the existing literature along with new experiments that Type II does not differ from Types III–V except under instructional conditions that encourage rule formation or attention to particular dimensions (Kurtz, Levering, Romero, Stanton, and Morris, 2012).

In studies of the SHJ types with different circumstances of learning, the paradigm-specific order does not hold. Specifically, the intermediate types separate in a consistent fashion, so $I < IV < III < V < II < VI$, into what we call the general order. There are four cases which yield the general order: First, stimulus generalization theory, which generates a prediction of the ordering of the classification problems based on the frequency of
mistakes (pairwise confusions) in learning unique labels (i.e., identification learning) for each item; (Shepard et al., 1961) second, stimuli comprised of integral dimensions (Garner, 1974) that are difficult for the learner to perceptually analyze and distinguish (such as brightness, hue, and saturation) (Nosofsky & Palmeri, 1996); third, learning by monkeys (Smith, Minda, and Washburn, 2004); and fourth, learning by children (Minda, Desroches, and Church, 2008).

To elaborate on these empirical findings, the results with integral-dimension stimuli (Nosofsky & Palmeri, 1996) are interpreted as reinforcing Shepard et al.’s (1961) view that stimuli generalization theory predicts ease of learning unless a process of attention or abstraction can be applied by the learner. In the cross-species research (Smith et al., 2004), four rhesus monkeys were tested on a modified version of the SHJ six types. The core finding is that Type II was more difficult for the monkeys to learn than Types III-V (which the authors elect to average over in their reporting). In the developmental work (Minda et al., 2008), the researchers modified the SHJ task to be age-appropriate for children of ages 3, 5, and 8. Only Types I-IV were tested: Type II was the most difficult to learn (consistent with the general rather than the paradigm-specific order). No significant difference between Types III and IV was observed, however it appears that the researchers did not evaluate the interaction between age of children and their performance on Types III and IV. From the mean accuracy data, it can be seen that the children show increasingly good performance on Type III with age and increasingly poor performance with age on Type IV. While we do not have access to statistical support, the available evidence is consistent with the younger children learning Type IV more easily than Type III (as in the general ordering).

There are two classes of explanation in the psychological literature on category learning. Mechanistic models, which are implemented in computational simulations of trial-by-trial learning, have been used to explain the paradigm-specific order (i.e. Love, Medin, and Gureckis, 2004; Kurtz, 2007) and some have been shown to account for both the paradigm-specific and general orders (i.e. Kruschke, 1992; Nosofsky & Palmeri, 1996; Pape & Kurtz, 2013). The other approach is based on the use of formal metrics to measure mathematical (logical) complexity (Feldman, 2000, 2006; Goodman, Tenenbaum, Feldman, and Griffiths, 2008; Goodwin and Johnson-Laird, 2011; Lafond, Lacouture,
These models heretofore account only for the paradigm-specific order. Indeed, if one remains committed to any particular, parameter-free metric of complexity, then only one ordering can be predicted no matter the circumstances of the learning. This suggests that mathematical complexity can never account for both orderings.

We put forth a mathematical complexity metric, statistical complexity, which can account for the general order. We posit that statistical and logical complexity, considered together, provide a complete picture of how human conceptual difficulty reflects underlying mathematical complexity. The complete picture is as follows: on one hand, there are sophisticated learners, who can observe separable dimensions and form something akin to logical rules, for whom problem difficulty is captured by logical complexity; on the other hand, there are unsophisticated learners or learners unable to separate dimensions, who do not/cannot form logical rules, for whom problem difficulty is captured instead by statistical complexity.

2. Theory

Logical complexity characterizes the length of a (shortest) description of a system. It includes Feldman’s Boolean complexity, used to characterize the SHJ tasks, which counts the minimal number of logic gates needed to represent categorical structures, and Kolmogorov (algorithmic) complexity, which is the length of the shortest program to produce a certain output [Li & Vitányi, 2008].

By contrast, statistical complexity characterizes the amount of information or uncertainty in a system. It includes the lower bound of statistical data compression and Shannon information entropy [Shannon, 2001], which measures how much information an observer can gain from one observation of a system: the higher the Shannon entropy, the more unpredictable. In particular, for a probabilistic system \( X = \{x_1, x_2, \ldots, x_n\} \) whose state \( x_i \) arises with probability \( p_i \), the system’s information entropy is given by \( H(X) = -\sum_i p_i \log_2 p_i \). (Base 2 reflects measuring information in binary ‘bits.’) For example, if \( X \) consists of two items that occur with equal probability (like heads and tails of a fair coin), then the system is maximally unpredictable, then the information entropy is \( H(X) = \)

\[ \frac{1}{2} \text{ bits}. \]

---

1 The model presented by Nosofsky, Palmeri, and McKinley (1994) combines elements of both mathematical and mechanistic models and accounts for the paradigm-specific order.
- (.5 \cdot (-1) + .5 \cdot (-1)) = 1. If X' consists of two items that occur with probabilities .75 and .25 (an unfair coin), then the system has lower unpredictability, because the one outcome occurs more often than the other; correspondingly, $H(X') = -( .25 \cdot (-2) + .75 \cdot (-0.41 \ldots)) \approx .81 < 1$.

Our complexity metric for the SHJ tasks is based on Shannon’s entropy. The metric encodes: If one observes some subset of the dimensions of the object, how much unpredictability is left in the categorization? A Type I classification, in which one dimension completely determines the category, has low unpredictability, because one dimension resolves all uncertainty (even though the other two resolve no uncertainty), while a Type VI classification has high unpredictability, because no one dimension resolves much unpredictability.

Formally, let a classification task be formulated as a binary function $f(x) \rightarrow \{A, B\}$, where $A$ and $B$ are two categories and $x$ is a multidimensional binary vector with $d$ dimensions, so $x = (x_1, x_2, \ldots, x_d)$. Then, we define the following metric

$$H(n) = \left(2^n \binom{d}{n}\right)^{-1} \sum_{\{i_1, i_2, \ldots, i_n\} \subseteq \{1, 2, \ldots, d\}} \sum_{\{b_1, b_2, \ldots, b_n\} \in \{0, 1\}^n} \sum_{a \in \{A, B\}} h(p(f(x) = a|x_{i_1} = b_1 \land x_{i_2} = b_2 \land \ldots \land x_{i_n} = b_n))$$

(1)

with $h(p) = -p \log_2 p$.

This metric calculates the average Shannon entropy remaining in the categorical decision $f(x)$ given that $n$ characteristics are observed. To understand the metric, let $n$ be the number of dimensions which are observable, and the vector $b = (b_1, b_2, \ldots, b_n)$ be the vector of the observed dimensions. (For example, suppose we observe that the first two dimensions are 0 and 1. Then $n = 2$ and $(b_1, b_2) = (0, 1)$). Next, the expression $p(f(x) = a|x_{i_1} = b_1 \land x_{i_2} = b_2 \land \ldots \land x_{i_n} = b_n)$ gives the probability that the category of $x$ is $a$, given that the $i_1^{th}$ element is $b_1$, and so on. The innermost summation is over categories $A$ and $B$. The second innermost summation is over possible vectors $b$, which is determined by the number of observable dimensions $n$. The outermost summation is over possible subsets of dimensions chosen: for example, if $n = 2$ and $d = 3$, then the possible sets of indices are $\{1, 2\}$, $\{2, 3\}$, and $\{1, 3\}$. The first term is chosen so that maximal unpredictability sums to one.

Given this definition, $H(0)$ represents the entropy when no dimensions are observed, so $H(0)$ must equal or exceed $H(n)$ for all $n > 0$. $H(0) = 1$
if the two categories are equally present in the stimulus space. On the other extreme, $H(d)$ gives the smallest value because the least unpredictability remains when all dimensions are observed. In the SHJ series of experiments, observing all dimensions uniquely defines the category, so $H(d) = 0$. (In principle, one could consider categorization learning in which the categories have some unresolvable uncertainty, in which case $H(d) > 0$.)

We aggregate this metric over all possible values of $n$, i.e. all possible observable dimensions. This aggregate metric is:

$$\hat{H} = \sum_{n=0}^{d} H(n),$$  \hspace{1cm} (2)

We consider $\hat{H}$ to be a measure of the overall statistical complexity of a classification task.

3. Results and Discussion

Consider the application to the SHJ tasks. Table 2 gives the outcome of our metric calculated for each categorization. Because the SHJ stimuli have three dimensions, the first four rows give the intermediate metric values $H(0)$ through $H(3)$, followed by the aggregate metric $\hat{H}$. $\hat{H}$ correctly predicts the general order. To understand the application to SHJ, consider $H(1)$ and $H(2)$. $H(1)$ finds $II = VI$ and and $H(2)$ finds $IV = III$ and $V = II$; therefore neither metric alone fully captures human learning difficulty in the general order. $H(1)$ is the metric of uncertainty left after one dimension is observed. Consider Table 1. Observe that, when the first dimension is 0 and the other dimensions are arbitrary, there is an even mix of $A$ and $B$ for each Types $II$ and $VI$. This holds for the second dimension and the third dimension as well. So, from the point of view of $H(1)$, Types $II$ and $VI$ are equally ‘difficult,’ in that learning a single dimension resolves no uncertainty. Furthermore, note that for all other Types ($I$, $III$, $IV$, and $V$), the previous exercise does not work, so $H(1)$ treats them differently.

In sum, the aggregated statistical complexity metric predicts the general ordering—and makes no observable connection with the paradigm-specific ordering. Further, we note that the complexity values for each type correspond not only to the qualitative ordering in the most comprehensive behavioral data set available (Nosofsky & Palmeri 1996), but also match a more subtle pattern in the data; namely, that the learning curves for Types $II$ and $V$, and
| Dim. Values | Category (A or B) By SHJ Type (I – VI) |
|-------------|--------------------------------------|
| I II III IV V VI |                                        |
| 0 0 0       | A A A A A A                             |
| 0 0 1       | A A A A A B                             |
| 0 1 0       | A B A A A B                             |
| 0 1 1       | A B B B B A                             |
| 1 0 0       | B B B A B B                             |
| 1 0 1       | B B A B B A                             |
| 1 1 0       | B A B B A A                             |
| 1 1 1       | B A B B A B                             |

Paradigm-Specific Order: \( I < II \leq III, IV, V < VI \)
General Order: \( I < IV < III < V < II < VI \)

Table 1: The six mappings of three-digit binary strings to categories, Type I-VI.

for Types III and IV, are grouped together. As can be seen in Table 2, the statistical complexity metric captures the degree of differentiation among the six types in addition to their pure ordering.

4. Conclusions

The existing complexity metric literature concludes that “human conceptual difficulty reflects intrinsic mathematical complexity[.]” (Feldman, 2000) Our finding strengthens and deepens this fundamental result: we identify a new domain of experiments in which mathematical complexity predicts human conceptual difficulty. While the existing metrics only explain a domain in which logical rules are possible and used, we extend the fundamental idea, of measuring mathematical complexity, into a domain where logical rules are impossible. Narrowly defined, logical complexity fails in this domain. However, the more fundamental idea is indeed true: human conceptual difficulty reflects intrinsic mathematical complexity. Now we know the form of that complexity reflects whether learners can create logical rules about the problem they face.
Table 2: Statistical metric calculated on the SHJ classification learning problems.

|       | SHJ Types | Order              |
|-------|-----------|--------------------|
|       | $I$  | $II$ | $III$ | $IV$ | $V$ | $VI$ |               |
| $H(0)$| 1    | 1    | 1    | 1    | 1   | 1    | $I = IV = III = V = II = VI$ |
| $H(1)$| 0.67 | 1.00 | 0.87 | 0.81 | 0.94| 1.00 | $I <^* IV <^* III <^* V <^* II = VI$ |
| $H(2)$| 0.33 | 0.67 | 0.50 | 0.50 | 0.67| 1.00 | $I <^* IV = III <^* V = II <^* VI$ |
| $H(3)$| 0    | 0    | 0    | 0    | 0   | 0    | $I = IV = III = V = II = VI$ |
| $\hat{H}$| 2.00| 2.67 | 2.37 | 2.31 | 2.60| 3.00 | $I <^* IV <^* III <^* V <^* II <^* VI$ |

$^* = matches \ the \ general \ order$
References
Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633.

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology, 50*, 339–368.

Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac, MD: Lawrence Erlbaum.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32*.

Goodwin, G. P., & Johnson-Laird, P. (2011). Mental models of Boolean concepts. *Cognitive Psychology, 63*, 34–59.

Kruschke, J. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22.

Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. *Psychonomic Bulletin & Review, 14*, 560–576.

Kurtz, K. J., Levering, K., Romero, J., Stanton, R. D., & Morris, S. N. (2012). Human learning of elemental category structures: Revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 552–572.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology, 51*, 57–74.

Li, M., & Vitányi, P. (2008). *An introduction to Kolmogorov complexity and its applications, Third Edition*. Springer.

Love, B., Medin, D., & Gureckis, T. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1518.
Nosofsky, R., Gluck, M., Palmeri, T., McKinley, S., & Glauthier, P. (1994a). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition, 22*, 352–352.

Nosofsky, R., & Palmeri, T. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin and Review, 3*, 222–226.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994b). Rule-plus-exception model of classification learning. *Psychological Review, 101*.

Pape, A. D., & Kurtz, K. J. (2013). Evaluating case-based decision theory: Predicting empirical patterns of human classification learning. *Games and Economic Behavior, 82*, 52–65.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*, 3–55. Originally published in 1948.

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*, 1–41.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General, 133*, 398–414.

Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology, 50*, 501–510.

Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology, 53*, 203–221.

Vigo, R. (2013). The GIST of concepts. *Cognition, 129*, 138–162.