Complexity Measures and Classification Learning

Andreas D Pape[1,4], Kenneth J Kurtz[2,4], & Hiroki Sayama[3,4]

[1] *Department of Economics, Binghamton University*
[2] *Department of Psychology, Binghamton University*
[3] *Departments of Bioengineering and Systems Science and Industrial Engineering, Binghamton University*
[4] *The Collective Dynamics of Complex Systems Research Group, Binghamton University*

In classification learning experiments, test subjects are presented with objects which they must categorize. The correct categories, which are known to the experimenter, are functions of the characteristics ("dimensions") of the objects, such as size, color, brightness, and saturation. The experiments measure the relative difficulty of learning different categorizations. One major factor which influences the difficulty of learning is whether the dimensions are easily distinguishable by the subjects. *Separable* dimension problems are ones in which humans *can* differentiate dimensions of the objects, e.g. size, color, and shape. *Integral* dimension problems are ones in which humans can not (easily) distinguish the dimensions, e.g. brightness, saturation, and hue. Psychologists reason that when dimensions are separable, humans develop logical rules about which dimensions matter, and when they are integral, humans do not (cannot) develop these rules.

Feldman (2000)[1] connected separable dimension learning to logical complexity. *Logical complexity* is a mathematical metric which ranks problems by how concisely the solution can be represented as a set of logical rules. Feldman showed that a logical complexity metric can accurately predict the relative difficulty of separable dimension learning problems for human subjects.

We find the other half of Feldman's story. In integral dimension problems, when humans cannot differentiate dimensions of objects and cannot form logical rule systems, humans fall back statistical information. *Statistical complexity* is a mathematical metric which evaluates the informational density of a problem statistically. Our conclusion is parallel to Feldman's: While Feldman finds that logical complexity predicts learning difficulty in separable dimension problems, we find that statistical complexity predicts learning difficulty in integral dimension problems.

The Psychology of Classification Learning: Background. 'Categorization' is a core area in the field of cognitive psychology: How do people learn, represent, and use categories to make sense of the world? The dominant research approach has been to investigate how people learn from experience to categorize a novel domain; these are the classification learning experiments discussed above. Researchers seek formal (mathematical or mechanistic) models to account for human performance in supervised classification learning.[2] Mathematical models include metrics like the one proposed in this paper that provide typically closed-form predictions of learning difficulty. Mechanistic models are algorithms that are run as simulations and provide stochastic assessments of learning difficulty.

Some category structures are more difficult to learn (i.e., show a greater error rate during training) than others. The difficulty pattern is the foremost empirical phenomenon researchers seek to explain. A classic demonstration originated with Shepard, Hovland, and Jenkins (1961)[3] who tested human learners on the six possible types of category structures that arise from assigning eight training items (based on all possibilities given three binary-valued dimensions) into two equal-sized classes. Learning is easiest when 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. This categorization is called "Type I." Learning is hardest when the two classes must be memorized, since they cannot be distinguished according to any set of rules or statistical regularity (Type VI). The remaining types (II-V) are intermediate in difficulty. In the traditional ordering, Type II (a logical XOR rule on two dimensions) is learned faster than the rest; however, an updated measure of the ordering reflects the finding that Type II does not differ from Types III-V except under particular instructional conditions.[4] A complete description of the six mappings is found in Table 1.

Mathematical accounts of the SHJ ordering. Feldman[1] offered a mathematical account showing that the Boolean complexity of category structures predicts the relative ease with which they are learned. Subsequent research has extended and improved upon the complexity based account of the SHJ ordering.[5–9] These elaborations are consistent with the updated SHJ ordering.[4] Based on this success, a natural step is to return to the psychological evidence to evaluate the complexity approach more fully. The SHJ ordering depends fundamentally on characteristics of the learner and material being learned. In particular, from SHJ we know that speed of learning of the six types differs when the task is item identification rather than classification. As predicted by stimulus generalization theory, learning slows down (for Types I-V) and while the intermediate types (II-V) remain clustered, they separate out into a distinct ordering ($IV < III < V < II$). Nosofsky and Palmeri (1996) studied SHJ classification learning with stimuli comprised of integral dimensions – for which humans generally cannot pick out, attend to, or generate hypotheses about individual stimulus dimensions. With stimuli varying in terms of brightness, hue, and saturation, rather than color, shape, and size, the $I < IV < III < V < II < VI$ ordering was again observed – as if classification were no different than identification learning.[10] To complete the picture, it has been shown that monkeys[11] and young children[12] also learn the SHJ types (with separable dimension stimuli) in the order predicted by pure stimulus generalization theory (i.e. $I < IV < III < V < II < VI$).

What does this portend for complexity-based accounts? Once committed to a particular instantation of logical complexity, such as Feldman's use of Boolean complexity, only one ordering can be predicted. While a logical complexity measure predicts the separable-dimension SHJ ordering, we ask: does a statistical complexity measure predict the SHJ ordering that emerges from a different task (identification), different materials (integral-dimension stimuli), or different learners (monkeys and children)? We find that it does.

Measures of Complexity: Background. While a number of different ways have been proposed for measuring complexity,[13] there are two largely distinct (but related) approaches to complexity characterization. The first approach is *logical complexity*. Logical complexity generally characterizes the length of a (shortest) description of a system, a problem, or a task. This class includes Feldman's Boolean complexity that counts the minimal number of logic gates needed to represent categorical structures, and the Kolmogorov (algorithmic) complexity, i.e., length of the shortest program code to produce a certain output.[14] The second approach is *statistical complexity*, which generally characterizes the amount of information or uncertainty in a system, a problem, or a task, based on a probabilistic framework. This class includes the lower bound of statistical data compression and the well known Shannon information entropy[15] which measures how much information an observer could gain from one observation of a system: the higher the Shannon entopy, the more unpredictable. More concretely, 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 $H(X) = -\sum_i p_i \log_2 p_i$. (The use of base 2 for $\log$ assumes that the amount of information entropy is measured in *bits*, i.e. binary digits.) 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 $H(X) = -(.5 \cdot (-1) + .5 \cdot (-1)) = 1$. 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 metric, based on Shannon's Entropy. In this paper, we propose a new complexity metric for the SHJ classification tasks based on the statistical complexity approach, and show that this new metric correctly predicts the ordering of empirical difficulty of tasks with integral dimensions. It is based on the idea of: 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, would have low average unpredictability, because one dimension resolves all uncertainty (even though the other two resolve no uncertainty), while a Type VI classification would have high average unpredictability, because one dimension resolves very little unpredictability.

Formally, let a classification task be formulated as a binary function $f(x) \to \{A, B\}$, 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

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

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

This metric calculates the average Shannon entropy remaining in the categorical decision $f(x)$ under a condition that the object's properties for n dimensions are fixed to particular binary value assignments. To understand the metric, let n be the number of dimensions which are observable, and the vector $b = (b_1, b_2, \ldots, b_n)$ is the vector of the observed dimensions. (For example, suppose we observed the first two dimensions, which had a value of 0 and 1 respectively. Then $n = 2$ and $(b_1, b_2) = (0, 1)$). We then find the probability that an arbitrary vector x is in some category a, given that n dimensions are fixed to the values given by b. That is the meaning of the expression $p(f(x) = a | x_{i_1} = b_1 \wedge x_{i_2} = b_2 \wedge \ldots \wedge x_{i_n} = b_n)$; literally, this expression 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 category types A and B. The second innermost summation is over possible vectors b, which is of course 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, then, is a weighting scheme so that maximal unpredictability sums to one.

Given this definition, $H(0)$ represents the entropy when no dimensions are observed. Naturally, it yields gives the largest value across values of n. $H(0) = 1$ if two categories are equally present in the stimulus space. On the other extreme, $H(d)$ always gives the smallest value because the least unpredictability remains once the object's properties are fully determined. 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 $H(d) > 0$. That is not the case in this series of psychological experiments, however.)

We then further aggregate this metric over all possible values of n, under the idea that the true unpredictability would be a metric which captures a number of possible observable dimensions. This aggregate metric is:

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

We consider $\hat{H}$ a measure of the overall statistical complexity of a classification task. In particular, if we order categories by $\hat{H}$, we are able to uncover the human difficulty ordering under integral dimensions.

Prediction for the SHJ tasks. Now let us consider the application to the SHJ tasks. Table 1 lists the categories associated with each binary string, for classification type $I - VI$. Table 2 gives the outcome of our metric calculated on each categorization. The number of dimensions in the SHJ problem is $d_{SHJ} = 3$. Correspondingly, the first four rows give the intermediate metric values $H(0)$ through $H(3)$, followed by the aggregate metric $\hat{H}$. Note that the ordering of $\hat{H}$ correctly predicts the observed integral ordering, which is our primary result.

As discussed above, the extreme values $H(0)$ and $H(d_{SHJ} = 3)$ are trivially 0 and 1 respectively.[1] These values have little interpretation.

The more interesting values are $H(1)$ and $H(2)$. $H(1)$ finds $II = VI$ and and $H(2)$ finds $IV = III$, and $V = II$; therefore neither individual metric alone fully captures human learning difficulty. These individual dimension metrics have independent interpretations as well. $H(1)$ 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 O and R for each Types II and VI. Observe that, 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. For example, under Type III, the first dimension is more informative than the second. A similar exercise with $H(2)$, observing any two dimensions, will reveal why $H(2)$ considers Types IV and III equally difficult and V and II equally difficult.

Feldman describes his final conclusion as: "human conceptual difficulty reflects intrinsic mathematical complexity after all" and "subjective conceptual complexity can be numerically predicted[.]" Our finding strengthens this fundamental result: we have a new domain of experiments in which mathematical complexity predicts subjective conceptual complexity. Moreover, our finding deepens the Feldman finding, in that while Feldman only observes 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, Feldman's metric fails in this domain. However, the more fundamental idea that Feldman exposed is indeed true: human conceptual difficulty appears to reflect intrinsic mathematical complexity. Now we know the form of that complexity reflects whether humans can create logical rules about the problem they face.

[1] The first indicates that when no dimension is observed, nothing is known, and the second indicates when all dimensions are observed, there is no uncertainty left, because categorizations are deterministic functions of the object dimensions.

b digits			Category, by Mapping Type					
1	**2**	**3**	*I*	*II*	*III*	*IV*	*V*	*VI*
0	0	0	O	O	O	O	O	O
0	0	1	O	O	O	O	O	R
0	1	0	O	R	O	O	O	R
0	1	1	O	R	R	R	R	O
1	0	0	R	R	R	O	R	R
1	0	1	R	R	O	R	R	O
1	1	0	R	O	R	R	R	O
1	1	1	R	O	R	R	O	R

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

	Category types					
	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 < VI <* III <* V <* II <* VI*					
	**=matches integral classification learning*					

Table 2: Our metric calculated on the SHJ classification learning problems

1. Feldman, J. Minimization of boolean complexity in human concept learning. *Nature* **407**, 630–633 (2000).

2. Pothos, E. & Wills, A. *Formal Approaches in Categorization* (Cambridge University Press, 2011).

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

4. Kurtz, K. J., Levering, K., Romero, J., Stanton, R. D. & Morris, S. N. Human Learning of Elemental Category Structures: Revising the Classic Result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition* Forthcoming (2012).

5. Feldman, J. An algebra of human concept learning. *Journal of mathematical psychology* **50**, 339–368 (2006).

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

7. Lafond, D., Lacouture, Y. & Mineau, G. 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 (2007).

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

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

10. Nosofsky, R. & Palmeri, T. Learning to Classify Integral-Dimension Stimuli. *Psychonomic Bulletin and Review* **3**, 222–226 (1996).

11. Smith, J. D., Minda, J. P. & Washburn, D. A. Category learning in rhesus monkeys: a study of the shepard, hovland, and jenkins (1961) tasks. *Journal of Experimental Psychology: General* **133**, 398 (2004).

12. Minda, J. P., Desroches, A. S. & Church, B. A. Learning rule-described and non-rule-described categories: a comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34**, 1518 (2008).

13. Lloyd, S. Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine* **21**, 7–8 (2001).

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

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

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Andreas D Pape. (email: apape@binghamton.edu).