Universal generalization and universal inter-item confusability

Nick Chater,[a] Paul M. B. Vitányi,[b] and Neil Stewart[c]

[a]*Institute for Applied Cognitive Science, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom;* [b]*Centrum voor Wiskunde en Informatica, 1098 SJ, Amsterdam, The Netherlands;* [c]*Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom.* **nick.chater@warwick.ac.uk   paulv@cwi.nl   neil.stewart@warwick.ac.uk**
http://www.warwick.ac.uk/fac/sci/Psychology/staff/academic.html#NC/
http://www.cwi.nl/~paulv/
http://www.warwick.ac.uk/fac/sci/Psychology/staff/postgraduates.html#NS/

Abstract: We argue that confusability between items should be distinguished from generalization between items. Shepard's data concern confusability, but the theories proposed by Shepard and by Tenenbaum & Griffiths concern generalization, indicating a gap between theory and data. We consider the empirical and theoretical work involved in bridging this gap. [**shepard; tenenbaum & griffiths**]

shepard shows a robust psychological law that relates the distance between a pair of items in psychological space and the probability that they will be confused with each other. Specifically, the probability of confusion is a negative exponential function of the distance between the pair of items. In experimental contexts, items are assumed to be mentally represented as points in a multidimensional Euclidean space, and confusability is assumed to be determined according to the distance between items in that underlying mental space. The array of data that **shepard** amasses for the universal law has impressive range and scope.

Although intended to have broader application, the law is primarily associated with a specific experimental paradigm – the identification paradigm. In this paradigm, human or animal agents are repeatedly presented with stimuli concerning a (typically small) number of items. We denote the items themselves as a, b, . . . , corresponding stimuli as S(A), S(B), . . . , and the corresponding responses as R(A), R(B), ... People have to learn to associate a specific, and distinct, response with each item – a response that can be viewed as "identifying" the item concerned.

How does a law concerning confusability in the identification paradigm relate to the question of generalization? We suggest that there is no direct relationship. Generalization from item A to item B in the sense discussed by **shepard**, involves deciding that an item b has property f, because item a has property f. This is an inductive inference: f(A), therefore f(B). By contrast, confusing item A with B means misidentifying item A as being item B. Generalization typically does not involve any such misidentification: on learning that a person has a spleen, I may suspect that a goldfish has a spleen – but there is no need to misidentify or mix up people and goldfish.

These observations suggest that there may be a gap between

SHEPARD's theoretical analysis, which considers the question of generalization, and his empirical data-base, which concerns confusability. This points up two distinct research projects, attempting to reconnect theory and data.

The first project attempts to connect theory to data. This requires gathering empirical data concerning generalization, to see to what extent generalization does have the negative exponential form predicted from **SHEPARD**'s theoretical analysis. This project is, to a limited degree, taken up in **TENENBAUM & GRIFFITHS**' empirical studies of generalization from single and multiple instances. These preliminary results suggest that the generalization function appears to be concave, which also fits with their Bayesian theoretical analysis. Whether the data have an exponential form, and whether there is a universal pattern of data across many different classes of stimuli, must await further empirical work. But some of our own results have suggested that generalization may be surprisingly variable, both between individuals and across trials, even with remarkably simple stimuli.

Stewart and Chater (submitted) investigated generalization to novel stimuli intermediate between two categories that differ in variability. The effect of the variability of the categories differed greatly between participants – some participants classified intermediate stimuli into the more similar, less variable category; others classified the intermediate stimuli into the less similar, more variable category. Further, altering the variability of the training categories had large effects on individual participants' generalization. When the difference in variability between the two categories was increased, some people increased generalization to the more variable category, and some increased generalization to the less variable category. Extant exemplar (e.g., Nosofsky 1986) and parametric/distributional (e.g., Ashby & Townsend 1986) models of generalization in categorization cannot predict the large variation between participants. This individual variation in performance suggests that there may be no single law governing human generalization, and therefore, that performance may not fit into **SHEPARD**'s theoretical analysis, although it is too early to draw firm conclusions on this issue.

The second project arises from the apparent gap between theoretical analysis and empirical data in **SHEPARD**'s program concerns connecting data to theory. **SHEPARD** has provided a strong evidence that confusability is an inverse exponential function of distance in an internal multidimensional space. How can this result be explained theoretically? The rest of this commentary develops a possible approach.

To begin with, we note that the view of psychological distance as Euclidean distance in an internal multidimensional space may be too restrictive to be applicable to many aspects of cognition. It is typically assumed that the cognitive representation formed of a visually presented object, a sentence or a story, will involve structured representations. Structured representations can describe an object not just as a set of features, or as a set of numerical values along various dimensions, but in terms of parts and their interrelations, and properties that attach to those parts. For example, in describing a bird, it is important to specify not just the presence of a beak, eyes, claws, and feathers, but the way in which they are spatially and functionally related to each other. Equally, it is important to be able to specify that the beak is yellow, the claws orange, and the feathers white – to tie attributes to specific parts of an object. Thus, describing a bird, a line of Shakespeare, or the plot of Hamlet as a point in a Euclidean multidimensional space appears to require using too weak a system of representation. This line of argument raises the possibility that the Universal Law may be restricted in scope to stimuli which are sufficiently simple to have a simple multidimensional representation – perhaps those that have no psychologically salient part-whole structure. We shall argue, however, that the Universal Law is applicable quite generally, since all these aspects are taken into account by the algorithmic information theory approach. This leads to a more generalized form of the Universal Law.

In particular, we measure the distance between arbitrary rep-

resentations (whether representations of points in space, of scripts, sentences, or whatever), by the complexity of the process of "distorting" each representation to the other. Specifically, the distance between two representations, A and B, is defined to be the *sum* of the lengths of the shortest computer program that maps from A to B *and* the length of the shortest computer program that maps from B to A. This is known as sum-distance (Li & Vitányi 1997). Sum-distance measure is attractive not only because it has some theoretical and empirical support as a measure of similarity (Chater & Hahn 1997; Hahn et al., submitted), but also because it connects with the theoretical notion of information distance, developed in the mathematical theory of Kolmogorov complexity (Li & Vitányi 1997). (See Chater 1999, for an informal introduction in the context of psychology.) The intuition behind this definition is that similar representations can be "distorted" into each other by simple processes, whereas highly dissimilar representations can only be distorted into each other by complex processes; the complexity of a process is then measured in terms of the shortest computer program that codes for that process.

SHEPARD uses a specific function, G(A, B), as a measure of the confusability between two items. It turns out that – using only the assumption that the mapping between the input stimuli and the identification responses is computable – it can be shown that G(A, B) is proportional to the negative exponential of the sum-distance between A and B. That is, if distance is measured in terms of the complexity of the mapping between the representations A and B, then **SHEPARD**'s universal law, when applied to confusability, follows automatically (Chater & Vitányi, submitted).

We have suggested that this result is attractive, because it applies in such a general setting – it does not presuppose that items correspond to points in an internal multidimensional psychological space. This observation suggests a further line of empirical research: to determine whether the Universal Law does indeed hold in these more general circumstances.