Making Sense of Typicality: What Makes a Good Example?

Joshua B. Tenenbaum
Dept. of Brain and Cognitive Sciences, MIT E10-120
Cambridge, MA 02139
jbt@psyche.mit.edu

The problem. Across both domains of categorization (Mervis & Rosch, 1981) and reasoning under uncertainty (Kahneman & Tversky, 1972), ratings of "typicality", "representativeness", or "goodness of example" are highly reliable predictors of stimulus recognition, learning, comparison, and frequency estimation behavior. But what makes a good example? Because typicality gradients may have many sources (Barsalou, 1985), both Kahneman & Tversky (1972) and Mervis & Rosch (1981) argue that the concept of typicality should be (or can be) defined only operationally, in terms of subjects' direct judgements. This position has sparked the sharp criticism that intuitive heuristics of typicality are too vague to serve as an explanatory construct in cognitive science (Gigerenzer, 1996).

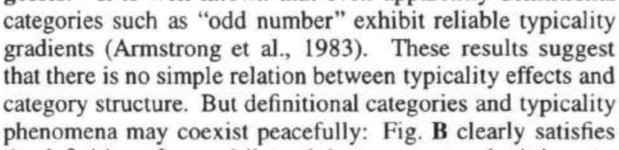
A solution: Bayesian genericity. As one step towards an analysis of typicality, I have developed a formal definition of *genericity*, what it means to be a good example of a process or category in the context of Bayesian inductive inference. (See Feldman (in press) for an alternative non-Bayesian treatment.) Let $\mathcal{H} = \{H_1, \ldots, H_n\}$ denote a set of mutually exclusive hypotheses that might account for an observation D. Then Bayes' rule asserts that D supports H_i (maximizes $p(H_i|D)/p(H_i)$) to the extent that the occurence of D is better explained (more probable) under H_i then under any of the other alternative hypotheses $H_{j \neq i}$, weighted by their priors. Such an observation will be called a *generic* example of H_i. Of the many ways in which an outcome or object may be typical of a process or category, being a good example for the purposes of inductive generalization is surely one of the most natural. This sense of typicality as genericity is not irremediably vague, but follows from normative principles of inductive inference and clarifies a number of puzzling phenomena:

1. Typicality may not imply typical features. Under the standard view, people judge a robin to be a typical bird because it shares many salient features with the bird prototype, or with other birds. But people judge S_1 = HHTHTTTH to be a typical sequence of fair coin flips not because it has certain salient features, but because it does *not* have the features (patterns or biased tendencies) that less typical sequences like S_2 = HTHTHTHT or S_3 = HHTHTHHH do. On a Bayesian analysis, those salient features make S_2 and S_3 less generic, and thus the "featureless" S_1 is correctly identified as the most typical.

2. Typical individuals may not belong to typical subclasses. Robbie the robin would be judged to be a typical or representative bird, and robins are also considered to be a typical *kind* of bird. But consider the two quadrilaterals in Fig. **A** and Fig. **B**. Most people consider **A** to be more typical or representative of quadrilaterals. Yet **B**, as a rectangle, belongs to a typical kind of quadrilateral, while **A** does not. In fact, stimuli belonging to salient subclasses of a category are less generic, and thus less likely to be seen as representative members of the category.

3. Typicality may be compatible with definitional categories. It is well known that even apparently definitional categories such as "odd number" exhibit reliable typicality gradients (Armstrong et al., 1983). These results suggest that there is no simple relation between typicality effects and category structure. But definitional categories and typicality phenomena may coexist peacefully: Fig. **B** clearly satisfies the definition of a quadrilateral, but as a rectangle, it is not a generic quadrilateral like Fig. **A**. Typicality gradients determined by genericity may tell us nothing about the structure of any one category on its own, but a great deal about the structure of the set of categories that comprise the hypothesis space for inductive inference.

4. "A set of typical X's" may not equal "a typical set of X's". In general, categories are best learned from experience with typical members (Mervis & Rosch, 1981). Since category learning usually involves experience with more than one example, the concept of "a typical X" should be naturally extendable to "a typical sample of X's". But this is not trivial: Fig. **A** is a representative quadrilateral, while a sequence of similar shapes (Fig. **C**) is not a representative sample of quadrilaterals. However, under the assumption of independent sampling, the same notion of genericity that picks out typical quadrilaterals (Fig. **A**) and typical sequences (S_1) also distinguishes typical sequences of quadrilaterals (Fig. **D**).



Supported by an HHMI predoctoral fellowship.

References

Armstrong, S., Gleitman, L., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13(3)*, 263-308.

Barsalou, L. (1985). Ideals, central tendencies, and frequency of instantiation as determinants of graded structure in categories. *JEP: LMC, 11(4)*, 629-654.

Feldman, J. (in press). The structure of perceptual categories. *J. Math. Psych.*, in press.

Gigerenzer, G. (1996). On narrow norms and vague heuristics. *Psych. Rev., 103(3)*, 592-596.

Kahneman, D. & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cog. Psych., 3*, 430-454.

Mervis, C. & Rosch, E. (1981). Categories of natural objects. In M. Rosenzweig & L. Porter (Eds.), *Annual Review of Psychology*.