2010. In E. Pothos & A. Wills (Eds.), Formal approaches in categorization (pp. 325-332). Cambridge: Cambridge University Press.

15 Comments on models and categorization theories: the razor's edge

Douglas Medin

If I may make a personal remark, one sign of old age is that people ask you to write commentaries on new(er) work. In the present case the invitation for me to write something may be linked to the publication of the Medin and Schaffer (1978) context theory of categorization model more than 30 years ago and/or the Smith and Medin (1981), *Categories and Concepts* book, almost as old. This ought to provide enough distance to view cumulative progress in this area of research and theory. Of course there was more than a little earlier work by Posner and Keele (1968), Reed (1973) and Smith, Shoben and Rips (1974) relevant to models and by Rosch, Mervis and others (e.g. Rosch, 1973, 1975; Rosch Mervis, 1975; Rosch *et al.*, 1976) laying out basic levels and goodness of example or typicality effects that reverberated through the cognitive sciences. The basic levels work was so important that it now has the status of being presupposed in developmental studies on the interaction of language and conceptual development (e.g., Waxman, 1989, 2002; Waxman & Lidz, 2006).

One way of assessing progress in an area is to evaluate how it is doing with respect to narrowness and insularity versus breadth. Cutting edge research seems like something that is inherently good, but it may be useful to examine what is being cut and how that edge is related to broader configurations. If we take the state of categorization research in 1980 as a benchmark, one could provide the following list of limitations of theory and data on categories and concepts.

1. Although concepts serve multiple functions (categorization, inference, communication, etc.) virtually all attention was directed at the categorization function of concepts.
2. Although there was a body of work on natural language concepts and a body on artificially created concepts (e.g., Posner dot patterns) and similar empirical results, the two literatures had little, if anything, to say to each other.
3. Almost all the adult research was conducted with undergraduate students at major research universities.

4. Almost all of the adult research used tasks that could be completed within an hour and nearly always involved exactly two categories.
5. Virtually all of the empirical work was on supervised categorization.
6. The models of categorization focused primarily on predicting transfer performance to new stimuli given after a category training period.
7. The stimuli themselves tended to be visual figures having little meaning or relevance to research participants.
8. The representation of the stimuli was assumed to be fixed and subject only to attentional weighting (a convenient assumption in comparing different models of category learning). Furthermore, the constituent features or dimensions were assumed to be independent and relational properties were ignored (and researchers did a good job of selecting stimuli where this assumption was not obviously violated).
9. There was relatively little categorization research in the cognitive neurosciences (other than the Wisconsin Card Sort task) and virtually none of it employed categorization models.

Let's start with these earlier limitations and examine the current state of affairs, paying special attention to the chapters in this volume. Progress has been considerable and almost everyone is on the cutting edge (of something). In the next few paragraphs I review some of that progress and then turn to what I take to be serious residual challenges to the field.

1. Past: Although concepts serve multiple functions (categorization, inference, communication, etc.) virtually all attention was directed at the categorization function of concepts.

   Present: Here is an area of clear progress. Brian Ross, Art Markman and others have done a number of studies on the inference function of concepts and there is now enough literature to review (Markman & Ross, 2003) and models to predict results from such studies.

   Under the influence of Rips, Osherson, Smith and others there is also something of a literature on the use of categories in reasoning, also known by the term 'category-based induction' (e.g., see Feeney & Heit, 2007, for a review). Finally, there is some work on conceptual combination (again stimulated by Osherson and Smith) as well as some corresponding computational models. Conceptual combination tends to be neglected because it is so challenging even to predict how novel noun-noun combinations will be interpreted (Gagné & Shoben, 2002; Wisniewski, 1997).

Neither of these latter two lines of work is represented in this volume, perhaps a continuing sign that the natural language and artificial, perceptual stimuli literatures still are not on good speaking terms.

2. Past: Although there was a body of work on natural language concepts and a body on artificially created concepts (e.g. Posner dot patterns) and what appear to be similar empirical results, the two literatures had little, if anything, to say to each other.

   Present: One can point to some nice work by Gert Storms and his colleagues applying models of categorization to natural language stimuli (e.g, Storms, De Boeck, & Ruts, 2000), but aside from this exception the two literatures remain as segregated as North and South Korea.

3. Past: Almost all the adult research was conducted with undergraduate students at major research universities.

   Present: This remains largely true with the exception of work on category-based induction which has included a range of participant populations.

4. Past: Furthermore, almost all of the adult research used tasks that could be completed within an hour and nearly as many involved exactly two categories.

   Present: There is now a modest amount of work looking at expertise either by identifying real-world experts (e.g. Medin *et al.*, 1997) or by dint of training in the lab (e.g. Gauthier & Tarr, 1997; Goldstone, 1998). This research is directly relevant to limitation 8 and has not tended to be addressed by models of categorization (but see by Chapter 10.) McDonnell and Gureckis.

5. Past: Virtually all of the empirical work was on supervised categorization.

   Present: Although the absolute amount of work on unsupervised categorization remains small, it has experienced a relatively large increase from its low base (see Chapters 9 and 10 by Pothos *et al.* and McDonnell and Gureckis). It now seems to be a realistic expectation that models of categorization should account for unsupervised categorization (a.k.a. free sorting).

6. Past: The models of categorization focused primarily on predicting transfer performance to new stimuli given after a category training period.

   Present: There is increasing attention directed to the learning side of category learning. This ranges from predicting the relative difficulty of learning different category partitionings (a tradition stemming from the classic Shepard Hovland & Jenkins,

1961 monograph), to predicting overall learning curves, to predicting learning curves for different stimuli (e.g. see Chapter 8 by Griffiths *et al.*).

7. Past: The stimuli themselves tended to be visual figures having little meaning or relevance to research participants.

   Present: This picture is largely unchanged. There are notable exceptions (see Chapter 12 by Harris and Rehder) by brave souls like Greg Murphy (e.g. Murphy, 2004; Murphy & Allopenna, 1994) aiming to account for the role of prior knowledge and meaningfulness on category learning. I'm still fond of the Wisniewski and Medin (1994) paper showing that meaningful category labels affect what is likely to count as a feature as well as category learning. Ed and I were dying to develop a computational model for this sort of context and the progress represented in the Rehder and Murphy (2003) KRES model (Chapter 12) suggests that it's not impossible.

8. Past: The representation of the stimuli was assumed to be fixed and subject only to attentional weighting (a convenient assumption in comparing different models of category learning). Furthermore, the constituent features or dimensions were assumed to be independent and relational properties were ignored (and researchers did a good job of selecting stimuli where this assumption was not obviously violated).

   Present: See the comment on expertise and feature learning above. Work on feature learning is important but inconvenient for modellers who prefer to model a steady-state, fixed representation. The same may hold for ignoring relational properties, despite the fact that relations are central to computational models of analogy and, arguably, to models of similarity (Markman & Gentner, 2000; Medin, Goldstone, & Genter, 1993).

9. Past: There was relatively little categorization research in the cognitive neurosciences (other than the Wisconsin Card Sort task) and virtually none of it employed categorization models.

   Present: This may be a case where a rising tide boosts all ships. The burgeoning of cognitive neuroscience has been associated with serious and successful efforts to link brain processes with categorization processes (see Chapter 4 by Ashby *et al.*). As they note, one of the large gains is that brain activity becomes an important dependent variable for constraining and testing models of categorization.

Some researchers might quibble with this list of criteria and important issues. It's fairly obvious why models of categorization should care

about knowledge effects, feature inference and unsupervised categorization but perhaps it's more debatable as to how important it is to extend learning procedures to more realistic durations, employ a wider range of stimulus materials or sample a broader set of study populations. My first response to this issue is why would one not want to show that a model has explanatory power beyond the confines of a narrow set of stimuli, procedures and populations? Perhaps more telling than this in principle statement is the in practice fact that generalizations about categorization and inferencing derived from studies with undergraduates do not appear to carry well to other populations (e.g. Medin & Atran, 2004). Indeed, some researchers (Henrich, *et al.*, in press) argue that undergraduates are 'the weirdest people in the world'.

In summary, I think the overall picture is at once mildly encouraging and seriously discouraging. The most positive developments are in cognitive neuroscience and forays into building theory and data on knowledge effects with meaningful materials, as well as bridging between natural language stimuli and the artificial.

A number of the chapters in this volume illustrate encouraging breadth in different ways. John Krusche's chapter (Chapter 6) describes a program of research on attention in learning that nicely bridges with work in associative learning, including research with non-human animals. Livesay and McLaren's treatment (Chapter 7) is more conservative but provides some useful model contrasts. Nosofsky's chapter (Chapter 2) shows the latest, greatest in developing and defending exemplar models of categorization and Minda and Smith's chapter (Chapter 3) does the same for prototype models of categorization. Iba and Langley's chapter (Chapter 11) reminds us that the machine learning area is also an important source of models. I couldn't help feeling some nostalgia because around 1990 there was a great deal of interaction between psychology and machine learning on categories and concepts, something that seems largely absent nowadays. Pothos *et al.* (Chapter 9) show another direction of breadth and generality by pushing the idea of simplicity. Griffith *et al.* (Chapter 8) provide a vision of how models might address more complex stimuli and relational structures. These are all useful contributions.

But this picture is also seriously discouraging. I remember reading research on impression formation in the 1970s and 1980s where participants were given a list of traits or behaviours and then asked to make an overall judgment concerning the person being described. Papers in this considerable literature almost always cited Solomon Asch's (1952) book. At one point I finally got around to reading this classic work and I distinctly remember being shocked at how his book was full of ideas that could be explored and also at the fact that impression formation only constituted a tiny fraction of it. The work that followed seemed at best to

be a pale imitation of Asch. The psychology of concepts and categories wasn't driven so much by a book as it was by findings on basic levels and typicality effects by people like Eleanor Rosch and Edward Smith. It may share with social cognition a sense that a lot more could and should be done. So while I congratulate these authors on some very interesting and cutting edge chapters, I also challenge them to pursue greater breadth in participants, paradigms and procedures and to aim for more integrative theories capable of bridging between perceptual stimuli and those carrying knowledge and meaning, between novices and various forms of expertise between categories on the one hand and concepts on the other. It's time to get off the razor's edge.

REFERENCES

Asch, S. (1952). *Social Psychology*. Englewood Ciffs, NJ: Prentice Hall.

Burnett, R., Medin, D., Ross, N., & Blok, S. (2005). Ideal is typical. *Canadian Journal of Experimental Psychology*, **59** (1), 5–10

Feeney, A., & Heit, E. (eds.) (2007). *Inductive Reasoning*, New York: Cambridge University Press.

Gagné, C. L., & Shoben, E. J. (2002). Priming relations in ambiguous noun-noun combinations. *Memory & Cognition*, **30**, 637–646.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Research*, **37**, 1673–1682.

Goldstone, R. L. (1998). Perceptual learning, *Annual Review of Psychology*, **49**, 585–612.

Henrich, J., Heine, S. J. & Norenzayan, A. (in press). The weirdest people in the world? *Behavior and Brain Sciences*.

Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *American Journal of Psychology*, **113** (4), 501–538.

Medin, D. L., & Atran, S. (2004). The native mind: biological categorization and reasoning in development and across cultures. *Psychological Review*, **111**, 960–983.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, **100**, 254–278.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: do all roads lead to Rome? *Cognitive Psychology*, **32**, 49–96.

Medin, D. L., & Schaffer M. M. (1978). A context theory of classification learning. *Psychological Review*, **85**, 207–238.

Murphy, G. L. (2004). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Murphy, G. L. & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 904–919.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77** (3), 353–363.

Reed, S. K. (1973). *Psychological Processes in Pattern Recognition*. New York: Academic Press.
Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review*, **10**, 759–784.
Rosch, E. (1973). Natural categories. *Cognitive Psychology*, **4**, 328–350.
(1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192–233.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573–605.
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382–439.
Ross, N., Medin, D. L., Coley, J. D., & Atran, S. (2003). Cultural and experiential differences in the development of folkbiological induction. *Cognitive Development*, **18**, 25–47.
Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75** (13).
Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.
Smith, E. E., Shoben, E. J., & Rips, L. J., (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review*, **1**, 214–241.
Storms, G., De Boeck, P., & Rits, W. (2000). Prototype and exemplar-based information on natural language categories. *Journal of Memory and Language*, **42**, 51–73.
Waxman, S. R. (1989). Linking language and conceptual development: linguistic cues and the construction of conceptual hierarchies. *Genetic Epistemologist*, **17** (2), 13–20.
(2002). Links between object categorization and naming: origins and emergence in human infants. In D. H. Rakison & L. M. Oakes (eds.), *Early category and Concept Development: Making Sense of the Blooming, Buzzing Confusion*. New York: Oxford University Press.
Waxman, S. R., & Lidz, J. (2006). Early word learning. In D. Kuhn & R. Siegler (eds.), *Handbook of Child Psychology*, (6th Edition, Vol. 2, pp. 299–335). Hoboken NJ: Wiley.
Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, **4**, 167–183.
Wisniewski, E. J., & Medin, D. L. (1994). The fiction and nonfiction of features. In R. S. Michalski & G. D. Tecuci (eds.), *Machine Learning: A Multistrategy Approach* (Vol. 4, pp. 63–84). San Mateo, CA: Morgan Kaufman.