Gentner, D. (2001). Exhuming similarity. Commentary on Tenenbaum, Joshua B. & Griffiths, Thomas L. Generalization, similarity, and Bayesian inference. Behavioral and Brain Sciences, 24(4), 669.

Commentary on Tenenbaum, Joshua B. & Griffiths, Thomas L. (2001) 'Generalization, Similarity, and Bayesian Inference'

.....

Abstract: *58* words
Main Text: *789* words
References: *112* words
Total Text: *959* words

Exhuming similarity

Dedre Gentner
Psychology Department
Northwestern University
2029 Sheridan Rd.
Evanston, IL 60208
USA

gentner@northwestern.edu
http://www.psych.northwestern.edu/psych/people/faculty/gentner/

Abstract

Tenenbaum and Griffiths' paper attempts to subsume theories of similarity -- including spatial models, featural models, and structure-mapping models -- into a framework based on Bayesian generalization. But in so doing it misses significant phenomena of comparison. It would be more fruitful to examine how comparison processes suggest hypotheses than to try to derive similarity from Bayesian reasoning.

Tenenbaum and Griffith's paper is large in its vision. It aims to synthesize similarity, concept learning, generalization, and reasoning. Under the rubric of Bayesian generalization, it offers a unification of Shepard's spatial model of similarity with Tversky's set-theoretic model and, for good measure, with structure-mapping accounts of similarity. This is a bold and ambitious idea, but there are some problems.

Getting directionality correct

First consider the unification of Tversky's contrast model with Shepard's spatial model and with generalization. T&G give Tversky's statement of the similarity of a target y to a base x:

$$S(y,x) = \Theta f(Y \cap X) + \alpha f(Y-X) - \beta f(X-Y)$$

To achieve a unification of Tversky's contrast model with their Bayesian model of generalization, T&G first assume the special case of the contrast model in which the measure f is additive. Then they adopt the ratio version of the contrast model, and finally they set $\alpha = 0$ and $\beta = 1$. QED – at this point the similarity equation does indeed resemble the Bayesian equation. But these weightings of α and β are the reverse of Tversky's. In Tversky's model, $\alpha > \beta$; that is, the

distinctive features of the target term (y) count more against the similarity of the pair than the distinctive features of the base (or standard, or referent) term (x). This turns out to matter. This is how Tversky explains the finding that people think that Nepal is more similar to China than the reverse. For most of us, China is the richer concept. When it's in the base position, its distinctive features get a lower weight (β) so the similarity of the pair is greater. The T&G formulation predicts directionality preferences that are the reverse of what's normally found.

Common relations are not scorned

T&G also take up structural models of similarity (Gentner & Markman, 1997; Medin, Goldstone & Gentner, 1993). They note that people in our studies often find common relations more important than common object attributes (primitive features) in similarity judgments. For example, people typically consider AA to be more similar to BB than to AC. T&G propose that the relational preference can be derived from Bayesian principles -- specifically, from the size principle, that people prefer hypotheses that delineate small consequential regions – i.e., that they prefer specific hypotheses to general hypotheses. (We might note in passing that the assumption that specific hypotheses are superior to general ones is not unique to Bayesian theories.) J&T conjecture that the greater saliency of relations over primitive object features stems from their greater specificity. Because relations are more specific than objects, a generalization based on relations is more informative. T&G use this rarity principle to offer an explanation for why *same* relations are more salient than *different* relations.

But the rarity explanation for why people pay attention to relations runs into some immediate difficulties. First, in analogy and similarity, higher-order relations such as causal relations are extremely highly weighted, despite being ubiquitous in human reasoning. Second, in natural language descriptions people tend to use relational terms broadly and object terms specifically: that is, they use a relatively small number of high-frequency relational terms, each very broadly, and a large number of low-frequency object terms, each quite specifically. This suggests that it is *objects* that are specific, not relations. To test the claim that the relational preference in comparison does not depend on rarity or specificity, I asked people for similarity judgments in triads that had low-frequency, specific object terms and high-frequency, rather general relational terms. Given the triad

Blacksmith repairing horseshoe

Blacksmith having lunch (A) Electrician repairing heater (B)

15 out of 19 people chose the relational response (B) as most similar to the standard, despite the fact that *repairing* is broadly applicable and frequently encountered, and *blacksmith* is highly specific and rarely encountered.

Another problem with the general claim that relations are salient relative to object features because of their specificity is that relations *aren't* always more salient. Relations have a salience advantage when comparing two present terms, but object features are more salient than relations in similarity-based memory *retrieval* (Gentner, Rattermann & Forbus, 1993; Ross, 1989). This disassociation between the relative salience of objects and relations during mapping vs. during retrieval cannot be accommodated with an all-purpose salience assignment.

There are other problems with subsuming similarity under generalization, such as that comparison

processes systematically highlight not only commonalities but also certain differences – alignable differences, those connected to the common system. T&G's aim of extending classic models of similarity to structured stimuli is laudable. But the attempt to subsume similarity under Bayesian generalization fails to capture the phenomena. Moreover, a processing account of comparison could provide one of the missing elements in Bayesian theories, an account of how people arrive at their hypothesis spaces.

References

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52,* 45-56.

Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25,* 524-575.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology, 23,* 222-264.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254-278.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(3), 456-468.

Acknowledgements

I thank Doug Medin, Lera Boroditsky and Art Markman for discussions of this paper.