11/2: Conceptual Knowledge (Brian)

Murphy, G. I. and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

- Theories are the glue that bind together the web of concepts in our mind.
    - A theory is a mental explanation, which need not be scientific. i.e., causal knowledge, scripts, rules and book-derived knowledge, that indicates which attributes are relevant for categorization.
    - Theories and concepts have bidirectional influences.
    - Quine (1977) first suggested a developmental progression from conceptual coherence based on perceptual similarity to that based on scientific theories. Our theories rarely reach scientific rigor, but Quine was on the right track.
- *Coherence* is not the same as *naturalness* (Keil, 1981). i.e., living things is a natural category, whereas bouncers and consciousness are not. The latter requires a theory to bind them together, making it an *unnatural* or *goal-derived* category.
- *Attributes* are a.k.a features and *underlying principles* are causal connections, script links and explanatory relations underlying theories.
- *Mental chemistry* vs. *Mental composition.* The former emphasizes how the relations, operations and transformations on the features of concepts gives rise to coherence, whereas the latter focuses on the features as independent entities.
- *Illusory correlation* is when a theory predicts a correlation, but the correlation exists only in the person's mind and not in reality, i.e., religion.
- The perceptual system defines concepts based on bottom-up features. However, these can be overridden or complemented by theories (Johnson-Laird, 1976).
- The most useful concepts are neither too broad nor too narrow (Rosch et al, 1976).
- Critiques of similarity-based measures:
    - *Similarity* is insufficient to causally explain conceptual coherence. Concepts which are coherent due to a theory may then seem more similar, confounding interpretation (Goodman, 1972).
        - Further, how does one define the set of features and their weights to be clustered? This problem is ill-posed (Tversky, 1977). Should we include all possible features, ie, every neuron in the brain? (Murphy, 1982a)
    - *Cue-validity* (Rosch, 1976) , i.e., the conditional probability that a thing is in a category based on the presence of a cue. Wrongly predicts that superordinate categories are more coherent.
    - *Category-validity* (Medin, 1983) predicts the opposite, i.e., the conditional probability that a thing has a cue based on its category. Incorrectly predicts that the most specific categories are the most coherent.
    - *Correlated-attributes* (Rosch, 1978), predicts that attributes appear in clusters which divide the world up into natural categories. i.e., not all things have all attributes. This leads to the formation of high within-category correlation and low between-category correlation. This problem is on the right track but still ill-posed - it's not clear which attribute correlations we should pay attention to.
    - *Categorization theories* (Smith & Miden, 1981)
        - *Classical view*: Categories are defined by singly necessary and jointly sufficient features. The flaw is that this does not constrain the coherence of the category.
        - *Probabilistic view.* Concepts are represented in terms of typical features.

This implies that we should be able to separate them using a linear regression, and that separable categories are easier to learn. The data does not support this view (Medin et al, 1981).

- *Exemplar view*. The same as the probabilistic view, with the stronger claim that any given member of the category is a sufficient representation of it. This does not constrain category membership in any reasonable way.

- Theories
    - *Attribute listing* is having a pool of subjects list attributes. These attributes tend to be generalizable. However, subjects may fail to assign the same attributes to subordinates as they do to superordinates, possibly because the attribute is not diagnostic at that level of analysis or in that context, or because we don't understand how subjects generate them.
    - *Correlated attributes* - why are categories with feature correlations more coherent than those without? Arbitrary attribute associations are hard to remember, whereas those that are linked via an explanation encoded in memory reduce interference (Bower et al, 1978).
    - *Concept use*
        - People may form illusory correlations based on theories and then think or act on them.
        - People may not be able to create a theory based on a perceived correlation.
        - People's expectations (theories) may cause them to overlook actual feature correlations and see what they want to see instead (Crocker, 1981).
        - Adding labels to features may aid in categorization as it facilitates the creation of theories (Adelman, 1981).
        - Linear separability may be employed if a theory suggests it is suitable.
        - *Idealized cognitive model* are relationships between concepts and exemplars that are akin to the relationship between theory and data. Both have a tradeoff between parsimony and explanatory power. When we encounter ambiguity we are aware of this tradeoff and revisit our theory to see if we can increase our explanatory power at the expense of parsimony.
        - *Experts* have finer categorical distinctions and their theories allow them to draw more sophisticated conclusions. However, their categories are also more overlapping in terms of attribute membership, as their abstracted view of a subject allows them to see underlying similarities that novices overlook.