

# **Concepts, Definitions, and Inheritance**

Interpreting the atoms of lexical decomposition

by

DAVID ZORNEK

## Conceptual Hierarchy

Concepts are a kind of cognitive representation; but not all cognitive representations are concepts. Concepts have been called “glue that holds our mental world together” [8] and “units of thought, constituents of belief and theories” [4]. Representations of particular objects, e.g. this particular coffee mug on my desk, my grandmother, etc., are not concepts. Going back to Plato, concepts have been regarded as *general* in character. A concept is a cognitive representation of some general category. In general, I will follow Greg Murphy’s convention of using “concept” to refer to mental representations of categories and using “category” to refer to the real-world collections of things.

### 1 The Classical View of Concepts

Under the *classical view* of concepts, a concept is fully analyzable by a genus-differentia definition, which provides a set of necessary and sufficient conditions for membership in a represented category. This view was implicit in Plato, whose *modus operandi* was to understand a concept by searching for a definition, e.g. *Meno* began as a search for a definition of “virtue,” and *Republic* was a search for a definition of “justice.” In both cases, it was thought that understanding (and therefore knowledge) would follow from the formulation of an appropriate definition. In pursuit of knowledge, these stories went, definitions were proposed, only to be rejected on the grounds that the category they picked out was either over- or under-inclusive for what the concept demanded, i.e. when they were found to not provide the right set of necessary and sufficient conditions.

Aristotle followed Plato by making the relation between concepts, definitions and categories explicit. For Aristotle, a definition is “an account which signifies what it is to be for something” [2]. The phrase “what it is to be” implies that Aristotle regarded definitions as something more fundamental than simply stipulating how a word is to be used, which is the function of contemporary dictionary definitions; Aristotle was careful throughout the

*Organon* and *Metaphysics* to use “to be” language only when talking about ontological matters and “to predicate” only when talking about linguistic matters.<sup>1</sup> Moreover, the only things that “be” for Aristotle are what he called *substances*. Primary substances are the particular objects we perceive in the world around us, while secondary substances are the general classes of objects. Primary substances cannot be regarded as concepts, since they are not general in character, leaving only secondary substances to be viably called concepts in Aristotle. However, Aristotle lacked the distinction between concepts and categories that I have made here, so his notion of secondary substance does double duty as an account of both the concept and the category. Only secondary substances, i.e. concepts/categories, have definitions [1] [2], and therefore a definition is an account of what a concept is and/or to what category a given object belongs.

In fact, there may be good reason to reject any “category” reading of Aristotle’s account of definitions, limiting him entirely to saying that definitions are an account of concepts specifically. Substances are composed of *matter* and *essence*. The matter of a substance is its material properties, while the essence is what we grasp about a substance, by means of a process of abstraction. Although Aristotle lacked a clear concept-category distinction, his writing about essences is hauntingly close to the way we talk about concepts today. Aristotle frequently gives an alternate formulation of his theory of definition, in which a definition is said to be a statement of the essence of a secondary substance. (In fact, the essence is defined as the “what it is to be” a particular kind of thing.)

The “what it is to be” language already implies that definitions provide a set of necessary and sufficient conditions; in all cases, it can be read as “what it is for a thing to be qua member of this class.” However, Aristotle’s more detailed analyses of the form a definition must take entails that a definition is an account of a hierarchical relation between concepts.

Classical definitions can be broken down into to components: the *genus* and the *differentia*. The genus is the broader class to which category mem-

---

<sup>1</sup>There is a systematic mapping in Aristotle between “to be” and “to predicate,” but the details of this would take us too far afield from the present subject.

bers belong, while the differentia is a set of properties that distinguishes between category members and other items falling under the genus. For example, take Aristotle’s famous definition: A human is a rational animal. “Animal,” in this definition, is the genus. All humans belong to the broader class of animals, i.e. the category of humans is a subcategory of the category of animals, i.e. “human” is subsumed under “animal.” The differentia here is “rational.” All humans possess a rational faculty and no animals other than humans possess a rational faculty. The differentia “rational” specifies which subcategory of animals we are thinking about when we use the concept “human.” Because “human” is a subcategory of “animal,” any knowledge we gain that holds true of all animals can be inferred to hold true of all humans. This kind of knowledge inheritance is the basis for Aristotle’s theory of the syllogism, which was the fountainhead of all subsequent work in logic.

## 2 The Downfall of the Classical View of Concepts

The classical view, in the form originally put forth by Aristotle, tasked definitions with (1) specifying what is *internal* to a concept, by virtue of their stipulation of necessary and sufficient conditions for category membership, and (2) specifying a hierarchy relation that holds between concepts, which is *external* to what is contained in any particular concept. At some point between Aristotle and Kant, philosophers lost sight of (2) to a large degree and began focusing almost exclusively on (1). Kant’s distinction between analytic and synthetic truths in *Critique of Pure Reason*, under one (not uncontroversial, but not at all implausible) reading, relies heavily on an appeal to (1), but no appeal is made to (2), except to the degree that he endorses and relies on Aristotle’s theory of the syllogism. For Kant, an analytic truth is one in which the predicate is “contained” in—or internal to—the subject concept, and all of his most famous examples of analytic truths (perhaps all of his examples) are cases in which the predicate is a component of or entailed by the definition of the subject concept.

By losing sight of (2), the road was paved for Wittgenstein’s influential

argument that definitions are not a complete account of semantic content, on the grounds that no set of necessary and sufficient conditions could be found for membership in the category “game” [14]. As an alternative, he offers us the view that word reference is determined by “family resemblances,” in which category membership is determined by general similarity, rather than some static list of necessary and sufficient conditions. Wittgenstein’s argument, it should be observed, addresses only (1). He fails to account at all for the task performed by definitions in (2), which leads him to the view that definitions do not convey word meaning. Wittgenstein’s argument against the classical view, although widely regarded as strong and important, had no empirical scientific backing and amounted to little more than “We’re having lots of trouble finding necessary and sufficient conditions, so why should we even suppose they exist?”

Beginning with Eleanor Rosch’s work in the 1970s, Wittgenstein’s family resemblances have amassed a body of empirical support that cannot be ignored. The literature refuting (1) is far too extensive to be surveyed completely here, with literally thousands of experiments spanning a period of over 40 years confirming the ubiquity of the family resemblance model of conceptual representation. Some hangers on in the fringe notwithstanding, the classical view is now widely accepted by cognitive scientists as—at minimum—unjustifiable on both theoretical and empirical grounds. It is my view, however, that only aspect (1) of the classical view has been overturned. Rather than surveying the entire literature, I will focus mostly on Rosch’s work, since it provided the impetus for all later research on this subject, and since her experiments provide data in favor of (2), while simultaneously providing data against (1).

### 3 Rosch vs. Classical Concepts

#### 3.1 Typicality

In [10], and [12], and [11] Rosch argued that the classical view entails that there are no “better” or “worse” members of a given category. If an item is

picked out by the category definition, it is in the category; if it isn't picked out, then it's not in the category. Category membership is all-or-nothing, and under the classical view it doesn't make any sense to talk about some items being *more* "in the category" than others. This is the aspect of the classical view that Rosch challenged, and it clearly amounts to a challenge against (1).

Rosch's work introduced the notion of *typicality*, which is explicitly presented as a development on Wittgenstein's family resemblances. Under the typicality model, category membership is gradient, rather than binary. Categorization is performed on the basis of some sort of overall similarity between members, i.e. a family resemblance, but there need not be any particular feature that holds true of all members of the category. This allows us to hold the view that birds can fly, but some birds such as penguins and ostriches cannot fly. Typicality is a measure of "goodness" of an item as a category member, and has been generally shown to be predictive of whether an item will be included by test subjects in the category.

In one experiment, Rosch simply asked test subjects to rate the "goodness" (or *typicality*) of various items as category members. She found that not only were test subject extremely willing to issue such ratings, but that the ratings were fairly consistent across test subjects. Subjects tended to agree that robins and sparrows are highly typical birds, while ostriches and penguins are not, i.e. robins and sparrows are better or more typical birds than ostriches and penguins. In other experiments, Rosch's data showed that typicality went beyond mere willingness of subjects to issue such typicality rankings. Subjects responded do questions such as "Are sparrows birds?" much more quickly than questions like "Are penguins birds?", and they demonstrated greater facility at making inferences that conformed to typicality than those that did not, e.g. inferences about the bird-typical features of penguins were slower than inferences about the bird-typical features of sparrows. Rosch concluded that there are, in fact "better" and "worse" members of categories represented by concepts, and subsequent research has universally agreed with her conclusion.

### 3.2 Hierarchy

Although genus-differentia definition fails to provide necessary and sufficient conditions for category membership, as (1) states, it seems highly implausible that genus-differentia definition has no relation to concepts whatsoever. After all, if an idea pervades for over 2000 years, we should think that it has *something* going for it, until and unless it has been shown to be completely off-track. I do not think this has been shown for genus-differentia definition, and despite its inadequacy, many fruitful results were obtained under the classical view, in particular by Piaget’s influential work from the 1960s [8, pg. 15]. In fact, the same series of experiments that established the falsity of (1) establishes the truth of (2). In failing to remember that (2) is part of the classical view in its original form, researchers have rejected the classical view entirely. In distinguishing clearly between (1) and (2), we will be able to make note of some empirical results regarding the hierarchical organization of concepts, which will help to preserve a portion of the classical view.

Research in cognitive science has established that concepts are organized more-or-less. Rosch’s original experiments showed that, where typicality effects did not confound response times, subjects were more quick to respond to questions that conformed to a hierarchy of the kind expressed in Aristotle’s genus-differentia definitions, and therefore (2), e.g. inferences of the general kind described by Aristotle’s theory of the syllogism. She distinguishes between three hierarchy levels—basic, superordinate, and subordinate—similar to Aristotle, and at times, she refers to these levels as species-levels and genus-levels, as Aristotle does in his writing on definitions. Subjects have a natural tendency to form concepts at the basic level, which is a sort of middle ground between generality of application and specificity of description.<sup>2</sup> Rosch’s results have been reproduced and expanded elsewhere, as in [6], [5], and [3].

A hierarchy is a network in which members are related by the set-inclusion relation. If one concept  $C$  is higher than another concept  $D$  in

---

<sup>2</sup>Rosch’s original metric for measuring basicity was found to be unproductive of actual categorization behavior in certain domains [7], but later research, such as [9], has offered other metrics which do better.

the hierarchy, then the category of which  $D$  is a mental representation is a subset of the category represented by  $C$ ; we say that  $C$  is *superordinate* to  $D$  and  $D$  is *subordinate* to  $C$ . It is not too difficult to see how genus-differentia definition naturally falls out of hierarchy. If we take  $\phi_D(g, d)$  to be a genus-differentia definition for  $D$ , where  $g$  is the genus and  $d$  is the differentia, then we see that the  $C$ -category is a candidate value for  $g$ .  $C$  is superordinate to  $D$ , i.e.  $C$  represents a broader category of which  $D$  represents a subset, i.e.  $C$  is a genus of  $D$ .  $d$  is some property possessed by all members of the  $D$ -category, but no other members of the  $C$ -category. Since hierarchy relations are subset relations, they are reflexive and transitive, but not symmetric. But these are properties of the inheritance relation, which is enough to *suspect* that conceptual hierarchy and inheritance orders have some fundamental relation to each other.

### 3.2.1 Objections to the hierarchy

In [13], Steven Sloman presents a series of experiments which he claims to support the conclusion that “Categories whose natural organization constitutes an inheritance hierarchy are surprisingly rare.” In fact, what he has actually demonstrated is: *Categories whose natural hierarchical organization is not confounded by typicality results in the case of atypical fringe cases are surprisingly rare.*

Sloman’s experiments asked subjects to evaluate quasi-Aristotelian inferences according to their willingness to accept conclusions on the grounds of the premises. I will describe only one experiment here; the differences between experiments are not important for what I will say here. The kinds of results seen, and the kinds of conclusions drawn by Sloman are the same for each experiment.

Consider the following arguments:

$$\frac{\text{All metals are pentavalent.}}{\text{Iron is pentavalent.}}$$

$$\frac{\text{All metals are pentavalent.}}{\text{Platinum is pentavalent.}}$$



Both arguments are quasi-Aristotelian syllogisms in that they would be syllogisms if they did not omit the middle premise “ $X$  is a metal,” where  $X$  stands in for either “iron” or “platinum.”<sup>3</sup> The omission of this middle premise is important in order to ensure that hierarchical organization, rather than ability to perform explicit logical inference, is being tested; subjects are, in effect, being asked to make use of their conceptual hierarchy to supply the middle premise.

Subjects were first tested to determine a typicality ranking for iron and platinum as different kinds of metals, and it was observed that iron was regarded as a more typical metal than platinum. Next, subjects were asked to evaluate their willingness to accept each conclusion on the grounds that all metals are pentavalent. Unsurprisingly, it was observed that subjects were more willing to accept the conclusion that iron is pentavalent than they were to accept that platinum is pentavalent. Sloman is explicit in ascribing this result to typicality effects. He writes, “Participants’ own ratings of similarity were used to determine which argument included the more similar categories in each pair of arguments . . . [M]ost participants chose the arguments with the more similar categories [as most convincing].” Later, when summarizing the series of experiments as a whole, he writes “Categorical inclusion relations are often neglected during reasoning, even in a categorical reasoning task where their relevance and application appears to be transparent. Furthermore, *similarity relations play a role in reasoning even when logic says they should be dominated by categorical ones*” (emphasis added). By his own admission, what is being observed is that typicality rankings can trump hierarchy relations in the atypical cases that lie on the fringe of the category, which is already demonstrated in Rosch’s (and later) work. I am left puzzled, then, as to why he ultimately draws the stronger conclusion that concepts are only very rarely organized hierarchically.

---

<sup>3</sup>Readers who are well-versed in Aristotelian logic will recognize the syllogisms as Barbara, from the first figure, with the minor premise omitted. Aristotle regarded such inferences as syllogisms, but *imperfect* syllogisms, due to the omitted premise.

## 4 Summary, or: back to the main point

The real takeaway points of this chapter are as follows:

- (i) On the classical view, a concept is fully analyzable by a genus-differentia definition:
  - (a) Genus-differentia definitions supply necessary and sufficient conditions for category membership (i.e., what is internal to a concept).
  - (b) Genus-differentia definitions describe hierarchy relations within a system of concepts (i.e., what is internal to the entire body of concepts, but external to any individual concept).
- (ii) At some point, researchers began focusing heavily on the concept-internal role of genus-differentia definitions in the classical view, while ignoring the concept-external role they play in the hierarchy.
- (iii) In showing that genus-differentia definitions do not account for what is internal to a concept, researchers have come to reject the classical view entirely.
- (iv) But, the empirical result that concepts are organized hierarchically lends credence to the part of the classical view that claims genus-differentia definitions describe can describe relations between concepts.

## References

- [1] Aristotle. *Categories and De Interpretatione*. Clarendon Press, 1975.
- [2] Aristotle. *Metaphysics*. Green Lion Press, 1999.
- [3] B. Berlin. *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton University Press, Princeton, NJ, 1992.
- [4] Susan Carey. *The Origin of Concepts*. Oxford University Press, Oxford, 2009.
- [5] A.M. Collins and M.R. Quillian. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:432–438, 1969.
- [6] B. Inhelder and J. Piaget. *The early growth of logic in the child: classification and seriation*. Routledge, London, 1964.
- [7] G. L. Murphy. Cue validity and levels of categorization. *Psychological Bulletin*, 91, 1982.
- [8] G. L. Murphy. *The Big Book of Concepts*. The MIT Press, Cambridge, MA, 2002.
- [9] G. L. Murphy and Brownell H. H. Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 1985.
- [10] E. Rosch. Natural categories. *Cognitive Psychology*, 4:328–350, 1973.
- [11] E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*. Erlbaum, Hillsdale, NJ, 1978.
- [12] E. Rosch and C.B. Mervis. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

- [13] Steve Sloman. Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, 35:1—33, 1998.
- [14] Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.