

# **Concepts, Definitions, and Inheritance**

Interpreting the atoms of lexical decomposition

by

DAVID ZORNEK

## Conceptual Hierarchy

In this chapter, we will become clearer on the relationship between concepts, genus-differentia definitions, and typicality (which will be defined below). In Chapter II, we saw that inheritance networks appear in all theories of lexical decomposition. In Chapter 4, we will see that inheritance networks provide a model of genus-differentia relations between atoms. The discussion in this chapter will serve as a bridge between concepts and inheritance networks, by arguing that, although the empirical research is conclusive that definitions do not provide a full analysis of concept, they are an adequate reflection of some aspects of concepts.

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” [12] 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, and by giving a more detailed account of what a definition is. For Aristotle, all and only general categories admit of definition, and a definition is a complete account of what a concept is and/or to what category a given object belongs [1] [2]. Classical definitions can be broken down into two components: the *genus* and the *differentia*. The genus is the broader class to which category members 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. Another way of stating this is in the form of an all-and-only condition for category membership: *all* humans are animals, and humans are the *only* animals that are rational.

From another angle, definitions provide sub- and super-category relations between concepts. 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. Concepts, and therefore knowledge, are organized hierarchically.

## 2 The Downfall of the Classical View of Concepts

We have now identified two aspects of the classical view that will be important for what follows. Definitions are tasked 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.

The distinction between (1) and (2) is loosely analogous to Carnap's distinction between internal and external questions. According to Carnap [5], when we ask logical or empirical questions, they are asked *within* a particular framework. Logical questions are questions pertaining to the logical relations between terms and logical rules defined within the framework; empirical questions pertain to the application of the framework in a particular domain. External questions, which are no less important, are questions asked *about* the frameworks, in particular those questions we might ask about the utility of the framework, i.e. whether and how the framework is a useful tool for meeting our ends, in relation to other possible frameworks. Likewise, we can ask internal and external questions about concepts. We might ask the internal question of whether a given concept applies to a particular object, e.g. "Is this object a human?" Or, we might ask the equally important external question of "How does the concept **human** relate to the concepts of **rational** and **animal**?"

What will follow is an argument that the empirical data reject the viability of (1) as an account of how to go about answering internal questions; it is incorrect to equate the question "Is this a human?" with the question "Is this a rational animal?" However, philosophers and cognitive psychologists have been too hasty in rejecting the classical view entirely, since other empirical research does support (2), which is the claim that definitions can answer external questions about how concepts relate to one another.

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" [20]. 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 a static definition. 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. Rosch's work in particular is useful here, 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), but it has been corroborated at length.

### 3 Typicality vs. Classical Concepts

#### 3.1 Typicality

In [15], and [18], and [17] 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.

Rosch observed a variety of different typicality effects. Each of these effects have been observed in subsequent research, which focuses more closely on specific results that did Rosch’s original experiments.<sup>1</sup>

**3.1.1. Test subjects *do* regard category members as “better” or “worse.”** This is perhaps the most fundamental observation of typicality. In [16, Experiment 1], test subjects were presented with members of a category and asked to issue a numerical ranking (from 1 to 7) of the extent to

---

<sup>1</sup>Often subsequent research was intended as a response to certain problems that were later uncovered in Rosch’s conclusions. These problems are not important here. Our purpose is simply to point out the ubiquity of typicality effects in empirical research on concepts.

which they regarded the item as representative of the category as a whole. She found that not only were test subject extremely willing to issue such ranking, but that the rankings were fairly consistent across test subjects. Subjects tended to agree that robins and sparrows are highly typical birds, while turkeys and penguins. This result was reproduced by Rips, Shoben, and Smith [14], who agreed with Rosch that category membership is determined by overall similarity, but who argued contra Rosch that typicality effects could be reconciled with essentialism (the view that there are some properties essential to category membership).

**3.1.2. Typicality, reaction time, and learning speed.** Rosch and Mervis [18, Experiment 5] set up a sorting task in which subjects were presented with strings of letters and numbers, which experimenters had previously sorted into two categories according to typicality. Subjects were asked to make judgments about which category each string belonged to, and were issued negative feedback when their judgment did not conform with the categories predetermined by experimenters. They found that subjects were more quick to make judgments about more typical members of each category, and that they learned to categorize typical strings more quickly than atypical strings (where speed of learning was measured in terms of number of errorneous judgments made before a successful judgment). In [18, Experiment 6], they repeated the same experiment, except that natural categories were used, rather than artifical categories of strings of letters. Results were the same.

Gastgeb, et al. [8], repeated Rosch and Mervis's sorting tasks for populations with different modes of cognitive functioning. They were able to reproduce the results of Rosch and Mervis in a normally functioning population, which was taken as a control group, and showed that high-functioning autistic children not only displayed the same typicality effects, but that typicality effects on reaction time and learning speed were more pronounced in autistics.

**3.1.3. Memory and recall.** Rosch [16, Experiment 6] primed subjects with a series of 54 objects belonging to the same category. At the end of a priming session, subjects were asked to write down all of the category members they had seen. It was observed that subjects had far greater recall facility for typical items than they did for atypical ones. Loftus [10] raised some concerns that Rosch failed to control for a tendency of subjects to categorize atypical category members as belonging to a different category; since subjects were cued to perform a task involving a specific category, differences in categorization could introduce a confounding uncontrolled variable. In later research, Ciss and Heth [6] introduced greater controls and were able to produce Rosch’s original result.

**3.1.4. The ubiquity of typicality effects.** This is only a small sample of the research supporting typicality effects. Rosch endorsed a specific model for explaining typicality, i.e. the *prototype theory* of concepts, which states that typicality is a measure of overall similarity to an abstract list of features, called a *prototype* or “best example” of a category. The competing *exemplar theory* holds that typicality is a measure of overall similarity between a list of particular instances of a category that have been encountered by concept-holders in the past. Although I will endorse the ubiquity of Rosch’s typicality effects, I remain mostly agnostic as to whether her prototype theory is the correct model of typicality. The point that is relevant for the present project is that typicality effects have been observed in a variety of cognitive domains and across a number of populations. The abundance of empirical support for typicality, and the relative scarcity of evidence against typicality presents conclusive evidence against aspect (1) of the classical view. However, other research seems to give strong support for aspect (2).

## 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.



(2) seems to be almost trivially true, and despite its inadequacy, many fruitful results were obtained under the classical view, in particular by Piaget’s influential work from the 1960s [12, 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 hierarchically. 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 [9], [7], and [3].

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

---

<sup>2</sup>Rosch’s original metric for measuring basicity was found to be unproductive of actual categorization behavior in certain domains [11], but later research, such as [13], has offered other metrics which do better.

*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 [19], 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 Aristotelian syllogisms that 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;

---

<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.

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. 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.

If concepts are organized hierarchically, and if “platinum” and “iron” are both subordinate to “metal” in the hierarchy, then they are subordinate regardless of typicality. There is no apparent reason why the atypicality of platinum as a metal should inhibit subjects from supplying the middle premise “Platinum is a metal.” But this seems to be exactly what is observed, and therefore, Sloman concludes, concepts are not organized hierarchically (except in a few special domains, e.g. biological concepts).

There is an alternate explanation of Sloman’s observations, which does not threaten hierarchy to the same degree. Greenberg and Bjorklund [?] performed a series of free recall experiments in which subjects were more willing to recall typical subcategories to a superordinate category than they were to recall atypical subcategories. They have hypothesized that this may be due to a tendency toward poor organization or mis-organization of atypical categories. This hypothesis is a variant the very complaints raised by Loftus [10] against Rosch’s original experiments on free recall; free recall might be inhibited by miscategorization of atypical items. If their hypothesis is correct, then Sloman’s observations are consistent with hierarchy and show that people can make mistakes in the way they position concepts within the hierarchy, and that typicality effects have a non-negligible effect on the prevalence of such errors. Under this hypothesis, Sloman’s objections against hierarchy don’t carry quite the same weight.

## 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 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] R. Carnap. Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*, 4:40–50, 1950.
- [6] D. Cisse and D.C. Heth. An evaluation of differential encoding and feature overlap accounts of typicality effects in free recall. *Canadian Journal of Psychology*, 43:359–368, 1989.
- [7] A.M. Collins and M.R. Quillian. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:432–438, 1969.
- [8] H. Gastgeb, M. Strauss, and N. Minshew. Do individuals with autism process categories differently? the effect of typicality and development. *Child Development*, 77:1717–1729, 2006.
- [9] B. Inhelder and J. Piaget. *The early growth of logic in the child: classification and seriation*. Routledge, London, 1964.
- [10] E. Loftus. Spreading activation within semantic categories: Comments on rosch’s “cognitive representation of semantic categories”. *Journal of Experimental Psychology: General*, 104:234–230, 1975.
- [11] G. L. Murphy. Cue validity and levels of categorization. *Psychological Bulletin*, 91, 1982.
- [12] G. L. Murphy. *The Big Book of Concepts*. The MIT Press, Cambridge, MA, 2002.

- [13] 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.
- [14] L. Rips, E. Shoben, and Smith E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12:1–20, 1973.
- [15] E. Rosch. Natural categories. *Cognitive Psychology*, 4:328–350, 1973.
- [16] E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104:192–233, 1975.
- [17] E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*. Erlbaum, Hillsdale, NJ, 1978.
- [18] E. Rosch and C. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.
- [19] Steve Sloman. Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, 35:1–33, 1998.
- [20] Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.