

Chapter 1

Concepts and Cognitive Science

Stephen Laurence and Eric Margolis

1. Introduction: Some Preliminaries

Concepts are the most fundamental constructs in theories of the mind. Given their importance to all aspects of cognition, it's no surprise that concepts raise so many controversies in philosophy and cognitive science. These range from the relatively local

Should concepts be thought of as bundles of features, or do they embody mental theories?

to the most global

Are concepts mental representations, or might they be abstract entities?

Indeed, it's even controversial whether concepts are objects, as opposed to cognitive or behavioral abilities of some sort. Because of the scope of the issues at stake, it's inevitable that some disputes arise from radically different views of what a theory of concepts ought to achieve—differences that can be especially pronounced across disciplinary boundaries. Yet in spite of these differences, there has been a significant amount of interdisciplinary interaction among theorists working on concepts. In this respect, the theory of concepts is one of the great success stories of cognitive science. Psychologists and linguists have borrowed freely from philosophers in developing detailed empirical theories of concepts, drawing inspiration from Wittgenstein's discussions of family resemblance, Frege's distinction between sense and reference, and Kripke's and Putnam's discussions of externalism and essentialism. And philosophers have found psychologists' work on categorization to have powerful implications for a wide range of philosophical debates. The philosopher Stephen Stich (1993) has gone so far as to remark that current empirical models in psychology undermine a traditional approach to philosophy in which philosophers engage in conceptual analyses. As a consequence of this work, Stich and others have come to believe that philosophers have to rethink their approach to topics in areas as diverse as the philosophy of mind and ethics. So even if disciplinary boundaries have generated the appearance of disjoint research, it's hard to deny that significant interaction has taken place.

We hope this volume will underscore some of these achievements and open the way for increased cooperation. In this introduction, we sketch the recent history of theories of concepts. However, our purpose isn't solely one of exposition. We also provide a number of reinterpretations of what have come to be standard arguments in the field and develop a framework that lends more prominence to neglected areas

This paper was fully collaborative; the order of the authors' names is arbitrary.

expression "the morning star" refers to the object it does because this expression has the sense it does.

3. *Senses are the indirect referents of expressions in intensional contexts* Certain linguistic contexts (e.g., "... believes that ..." and other propositional attitude reports) have distinctive and peculiar semantic properties. Outside of these contexts, one can freely substitute coreferential terms without affecting the truth value of the sentence ("the morning star is bright" → "the evening star is bright"), but within these contexts, the same substitutions are not possible ("Sue believes that the morning star is bright" ↔ "Sue believes that the evening star is bright"). Frege's explanation of this type of case is that in such contexts expressions do not refer to their customary referents, but rather to their customary *senses*. Since the expressions have different customary senses, they actually have different referents in these contexts. Thus Frege is able to maintain the principle that coreferential terms can be substituted one for the other without a change in truth value, despite what otherwise may have appeared to be a decisive counterexample to the principle.

Frege's semantic theory, and the phenomena he used to motivate it, have generated a great deal of controversy, and they have had an enormous influence on the development of semantic theories in philosophy and linguistics. For now, though, the important issue is the ontological status of senses. Frege argued that senses, construed in terms of these theoretical roles, cannot be mental entities. Since it's common in philosophy to hold that concepts just are Fregean senses, it would seem that Frege's case against mental entities is especially pertinent. The problem, in his view, is that mental entities are subjective, whereas senses are supposed to be objective. Two people "are not prevented from grasping the same sense; but they cannot have the same idea" (1892/1966, p. 60). (Note that for Frege, ideas are mental entities.)

If this is the argument against the view that concepts are mental representations, however, it isn't the least bit convincing. To see why, one has to be careful about teasing apart several distinctions that can get lumped together as a single contrast between the subjective and the objective. One of these concerns the difference between mental representations, thoughts, and experiences, on the one hand, and extra-mental entities on the other. In this sense, a stone is objective, but a mental representation of a stone is subjective; it's subjective simply because it's mental. Notice, however, that subjectivity of this kind doesn't preclude the sharing of a mental representation, since two people can have the same type of mental representation. What isn't possible is for two people to have the very same *token* representation. This brings us to a second subjective-objective distinction. It can be put this way: Mental representations are subjective in that their tokens are uniquely possessed; they belong to one and only one subject. Their being subjective in this sense, however, doesn't preclude their being shareable in the relevant sense, since, again, two people can have the same representation by each having tokens of the same type. When someone says that two people have the same concept, there is no need to suppose that she is saying that they both possess the same token concept. It would make as much sense to say that two people cannot utter the same sentence because they cannot both produce the same token sentence. Clearly what matters for being able to utter the same sentence, or entertain the same concept, is being able to have tokens of the same type. So while mental representations are subjective in the two senses

we've isolated, this doesn't stop them from being objective in the sense of being shareable.⁵

In short, we see no reason why concepts can't be mental representations. And given the role of mental representations in theories of psychological processing, it would be entirely natural to follow psychological usage in calling these representations concepts. Still, this usage isn't meant to preclude a role for the abstracta that Fregeans mean to highlight. To see this, one need only consider the question of whether Frege himself could have it both ways, employing mental representations and senses. The answer, of course, is that he could. On this model, beliefs and other propositional attitudes would involve token mental representations that have other representations—concepts—as their constituents. Senses would come in as the semantic values of these representations. That is, in addition to having worldly objects and properties as their referents, mental representations (like words, on Frege's original account) would have senses too. In this way, senses help to type mental representation; they provide part of the conditions for individuating concepts.

Given this way of combining the more traditional philosophical account of concepts with the representationalism of psychology, it's little more than a terminological debate whether representations or the abstracta should be called concepts. Since we think there needn't be any confusion on this point—and since we are primarily interested in the mental representations—we'll continue to follow standard psychological usage, according to which concepts are representations.⁶

With these preliminaries out of the way, we can now turn to the theories of concepts themselves. We will work through five that figure prominently in discussions in linguistics, philosophy, and psychology. They differ in their motivations and the problems they face, but they aren't nearly as distinct from one another as is often assumed. We'll see, for example, that some problems aren't tied to a single theory; rather they present a general challenge to nearly any theory of concepts. Similarly, some of the resources that trace back to one account of concepts can be enlisted in surprising ways to help other accounts. In general, the theories that we will discuss differ in what they say about the structure of concepts. Along the way, we'll mention a number of respects in which the options regarding conceptual structure can be expanded. In the concluding section (sec. 7), we'll bring some of these strands together by discussing four ways of construing what theories of concepts have to say about the nature of concepts.

2. *The Classical Theory of Concepts*

2.1. *Concepts and Definitions*

In one way or another, most theories of concepts can be seen as reactions to, or developments of, what is known as the *Classical Theory of Concepts*.⁷ The Classical

5. A third sense in which mental entities may be subjective—also suggested by Frege's text—is that they are highly idiosyncratic. Much of Frege's criticism of "ideas" is that they are too variable from one person to the next. "A painter, a horseman, and a zoologist will probably connect different ideas with the name 'Bucephalus'" (59). At best, however, Frege's observation establishes only that ideas aren't likely to be shared, not that they are, in principle, unshareable. Moreover, it's hard to see how the idiosyncrasy of ideas would motivate the claim that concepts are abstracta.

6. For further discussion on this point, see the appendix (sec. 8) and Margolis and Laurence (ms).

7. Also called the *Traditional Theory* or the *Definition View*.

Theory holds that most concepts—especially lexical concepts—have definitional structure. What this means is that most concepts encode necessary and sufficient conditions for their own application.⁸ Consider, for example, the concept BACHELOR. According to the Classical Theory, we can think of this concept as a complex mental representation that specifies necessary and sufficient conditions for something to be a bachelor. So BACHELOR might be composed of a set of representations such as IS NOT MARRIED, IS MALE, and IS AN ADULT. Each of these components specifies a condition that something must meet in order to be a bachelor, and anything that satisfies them all thereby counts as a bachelor. These components, or features, yield a semantic interpretation for the complex representation in accordance with the principles of a compositional semantics.

This conception of concepts has a long history in philosophy. The seventeenth-century philosopher John Locke seems to be assuming a version of the Classical Theory when he gives his account of the concepts SUN and GOLD (1690/1975, pp. 298–299 and p. 317, respectively):

[T]he *Idea* of the *Sun*, What is it, but an aggregate of those several simple *Ideas*, Bright, Hot, Roundish, having a constant regular motion, at a certain distance from us, and, perhaps, some other....

[T]he greatest part of the *Ideas*, that make our complex *Idea* of *Gold*, are Yellowness, great Weight, Ductility, Fusibility, and Solubility, in *Aqua Regia*, etc. all united together in an unknown *Substratum*...⁹

On the Classical Theory, most concepts—including most lexical concepts—are complex representations that are composed of structurally simpler representations. What's more, it's natural to construe their structure in accordance with the Containment Model, where the components of a complex concept are among its proper parts.¹⁰ Some of these components may themselves be complex, as in the case of BACHELOR. But eventually one reaches a level of primitive representations, which are undefined. Traditionally, these primitive representations have been taken to be sensory or perceptual in character, along broadly empiricist lines.

It is, of course, an oversimplification to speak of *the* Classical Theory of concepts, as though there were just a single, unitary theory to which all classical theorists subscribe. In reality, there is a diverse family of theories centered around the idea that

8. By "application" we mean a semantic relation; that is, a concept encodes the conditions that are singly necessary and jointly sufficient for something to be in its extension. Another sense of the term is to indicate a psychological process in which an object is judged to fall under a concept. We'll try to avoid this ambiguity by always using "application" in the semantic sense, unless the context makes it very clear that the psychological sense is intended. Notice, then, that in the first instance we have characterized the Classical Theory in semantic terms. This doesn't mean, however, that the theory is devoid of psychological import. See the discussion of concept acquisition and categorization, below.

9. Locke's views about natural kind concepts are complicated by the fact that he took natural kinds to have both a nominal and a real essence. For Locke, the real essence of a kind like gold isn't known, but the nominal essence is, and must be, in order to possess the corresponding concept. Arguably, however, he takes the nominal essence to give necessary and sufficient conditions for the application of a kind concept, since he holds that the nominal essence is defined relative to the real essence in such a way that the two track one another.

10. It's natural, but not mandatory. Alternatively, one could think of a classically structured concept as a node that stands in inferential relations to its defining features. The advantage of the Containment Model is that it makes especially clear which associated concepts are its defining features and which are incidental.

concepts have definitional structure. What we call the Classical Theory of concepts is an idealized account that abstracts away from many of their differences. To mention just one point on which classical theorists disagree: Many recent classical theorists have abandoned the strict empiricist view that concepts are ultimately composed of features expressing sensory properties.

It would be difficult to overstate the historical predominance of the Classical Theory. Aspects of the theory date back to antiquity (see Plato 1981 [chapter 2 in this volume]).¹¹ And the first serious challenges to its status weren't until the 1950s in philosophy, and the 1970s in psychology. Why has the Classical Theory been held in such high regard? The theory has powerful explanatory resources, offering unified accounts of concept acquisition, categorization, epistemic justification, analytic entailment, and reference determination, all of which flow directly from its basic commitments (see Fodor, J. A. et al. 1980 [chapter 21]). We will briefly review these accounts, since it helps to flesh out the Classical Theory and its substantial motivations.

Box 1

The Classical Theory

Most concepts (esp. lexical concepts) are structured mental representations that encode a set of necessary and sufficient conditions for their application, if possible, in sensory or perceptual terms.

Concept Acquisition If a concept is a complex representation built out of features that encode necessary and sufficient conditions for its application, then the natural model of concept acquisition is one where the learner acquires a concept by assembling its features. If, in accordance with the empiricist version of the Classical Theory, we add the further stipulation that primitive features are sensory or perceptual, the model we arrive at is something like the following. Through perception, sensory properties are monitored so that their representations are joined in a way that reflects environmental contingencies. Having noticed the way these properties correlate in her environment, the learner assembles a complex concept that incorporates the relevant features in such a way that something falls under the new, complex concept just in case it satisfies those features. In this way, all concepts in the end would be defined in terms of a relatively small stock of sensory concepts. As John Locke put it in *An Essay Concerning Human Understanding* (1690/1975, p. 166),

[E]ven the most abstruse Ideas, how remote soever they may seem from Sense, or from any operation of our own Minds, are yet only such, as the Understanding frames to it self, by repeating and joining together *Ideas*, that it had either from Objects of Sense, or from its own operations about them. . . .

A somewhat more recent advocate of this position is the influential twentieth-century German philosopher Rudolf Carnap. In "The Elimination of Metaphysics through Logical Analysis of Language," Carnap writes (1932/1959, pp. 62–63),

11. When, for the first time, we refer to a chapter that is reprinted in the present volume, we'll indicate this with brackets. Subsequent references will omit the bracketed material.

In the case of many words, specifically in the case of the overwhelming majority of scientific words, it is possible to specify their meaning by reduction to other words ("constitution," definition). E.g., "'arthropodes' are animals with segmented bodies and jointed legs." ... In this way every word of the language is reduced to other words and finally to the words which occur in the so-called "observation sentences" or "protocol sentences."¹²

In the face of repeated failures to analyze everyday concepts in terms of a purely sensory base, contemporary theorists have often relaxed the strong empiricist assumption that all simple concepts must be sensory. For example, Eve Clark (1973) sees the process of acquiring the meaning of a word like "brother" as comprising several stages where semantic components get added to an initial representation. In the earliest stage the representation consists of only two components: +MALE, -ADULT. In subsequent stages, -ADULT is changed to \pm ADULT, +SIBLING is added, and +RECIPROCAL is added. In this way, a representation for "brother" is gradually constructed from its constituent representations, which collectively provide a definition of the word and distinguish it from related words, such as "boy." Though these components may not be primitive, Clark isn't committed to the idea that further decomposition will always lead to purely sensory concepts. In fact, she says that many words, especially relational terms, require possibly irreducible features that encode "functional, social, or cultural factors" (p. 106). Similarly, the linguist and philosopher Jerrold Katz writes (1972 [chapter 4 in this volume], p. 40),

[T]he English noun "chair" can be decomposed into a set of concepts which might be represented by the semantic markers in (4.10):

- (4.10) OBJECT, PHYSICAL, NON-LIVING, ARTIFACT, FURNITURE, PORTABLE, SOMETHING WITH LEGS, SOMETHING WITH A BACK, SOMETHING WITH A SEAT, SEAT FOR ONE.

He adds that these semantic markers—or features—require further analysis, but, like Clark, he isn't committed to a reduction that yields a purely sensory base.

No doubt, a component-by-component model of concept acquisition is compelling even when it is detached from its empiricist roots. The simplicity and power of the model provides considerable motivation for pursuing the Classical Theory.

Categorization The Classical Theory offers an equally compelling model of categorization (i.e., the application of a concept, in the psychological sense; see note 8). In fact, the model of categorization is just the ontogeny run backwards; that is, something is judged to fall under a concept just in case it is judged to fall under the features that compose the concept. So, something might be categorized as falling under the concept CHAIR by noting that it has a seat, back, legs, and so on. Categorization on this model is basically a process of checking to see if the features that are part of a concept are satisfied by the item being categorized. As with the general model of concept acquisition, this model of categorization is powerful and intuitively appealing, and it's a natural extension of the Classical Theory.

12. Throughout we'll ignore certain differences between language and thought, allowing claims about words to stand in for claims about concepts. Carnap's account is about the semantics of linguistic items but otherwise is a useful and explicit version of the Classical Theory.

Epistemic Justification A number of philosophical advocates of the Classical Theory have also emphasized the role it could play as a theory of epistemic justification. The idea is that one would be justified in taking an item to fall under a given concept by determining whether its defining components are satisfied.

The quotation from Carnap (above) is part of a larger passage where he explains that we are justified in taking a thing, *x*, to be an arthropode if a sentence of the form "the thing *x* is an arthropode" is "deducible from premises of the form '*x* is an animal,' '*x* has a segmented body,' '*x* has jointed legs' . . ." (1932/1959, p. 63). Since the components that enter into the concept provide a definition of the concept, verifying that these components are satisfied is tantamount to verifying that the defined concept is satisfied as well. And since it's often assumed that the ultimate constituents of each concept express sensory properties, the verification procedure for a concept's primitive features is supposed to be unproblematic. The result is that justification for abstract or complicated concepts—including the "theoretical" concepts of science—reduces to a series of steps that implicate procedures with little epistemic risk.

Analytic and Analytic Inferences Another important motivation for the Classical Theory is its ability to explain a variety of semantic phenomena, especially analytic inferences. Intuitively, there is a significant difference between the inferences in (1) and (2):

(1) Smith is an unmarried man. So Smith is a man.

(2) Smith is a weight-lifter. So Smith is a man.

In (1), unlike (2), the conclusion that Smith is a man seems to be guaranteed by the premise. Moreover, this guarantee seems to trace back to the meaning of the key phrase in (1), namely, "unmarried man."

Traditionally, analytic inferences have been taken to be inferences that are based on meaning, and a sentence or statement has been taken to be analytic just in case its truth is necessitated by the meanings of its constituent terms. Much of this conception of analyticity is captured in Immanuel Kant's account of analyticity as conceptual containment. "Either the predicate *B* belongs to the subject *A*, as something which is (coveredly) contained in this concept *A*; or *B* lies outside the concept *A*, although it does indeed stand in connection with it. In the one case I entitle the judgment analytic, in the other synthetic" (1787/1965, p. 48). One of the most widely cited examples in the contemporary literature is the concept BACHELOR. Consider (3):

(3) Smith is a bachelor. So Smith is a man.

The inference in (3) is not only correct but seems to be guaranteed by the fact that it is part of the meaning of "bachelor" that bachelors are men. It's not as if one has to do a sociological study. The Classical Theory explains why one needn't look to the world in assessing (3), by claiming that the concept BACHELOR has definitional structure that implicates the concepts MAN, UNMARRIED, and so on. Thus (3) and (1) turn out to be similar, under analysis.

Katz (1972) gives much the same explanation of the validity of the inferences from (4.13)

(4.13) There is a chair in the room.