Kwong Jack (2006) "Why Concepts Can't Be Theories", **Philosophical Explorations**, vol. 9, no. 3, September 2006, pp. 309-325. [ISSN: 1386-9795] The version of record is published by Taylor & Francis available online: http://www.tandfonline.com/10.1080/13869790600815830

Why Concepts Can't Be Theories

Jack M. C. Kwong

ABSTRACT: In this paper, I present an alternative argument for Jerry Fodor's recent conclusion that there are currently no tenable theories of concepts in the cognitive sciences and in the philosophy of mind. Briefly, my approach focuses on the 'theory-theory' of concepts. I argue that the two ways in which cognitive psychologists have formulated this theory lead to serious difficulties, and that there cannot be, in principle, a third way in which it can be reformulated. Insofar as the 'theorytheory' is supposed to replace, and to rectify the problems of, the earlier 'classical' and 'probabilistic' theories, its failure confirms Fodor's original observation. Since my critique does not rest on controversial philosophical assumptions and is readily available from within the cognitive sciences, it is a stronger argument than Fodor's.

Introduction

In recent years, Jerry Fodor (e.g. 1994, 1998, 2004) has argued that most theories of concepts in the cognitive sciences and in the philosophy of mind are untenable. According to him, all of these theories violate one or more of what he calls the five 'non-negotiable' conditions. One difficulty with Fodor's argument is that some of the five conditions he stipulates can be shown to be negotiable. Some philosophers, for example, have rejected the view that concepts must be compositional (e.g. Schiffer 1987). Still others have argued that an appeal to content similarity, as opposed to content identity, is sufficient to explain the publicity of concepts (e.g. Harman 1993). While I agree with the main thrust of Fodor's conclusion, I argue that a simpler and more defensible case can be made for it, one that, contrary to Fodor's argument, does not rest on problematic premises and is readily available from within the cognitive sciences.

My approach takes as its point of departure the 'theory-theory' of concepts.[1] This theory holds that conceptual structure should not, pace the 'classical' and prototype theories, or more generally, 'similarity-based' theories, be understood in terms of lists of properties but in terms of theories (naı¨ve or otherwise) that people hold. In the following, I will examine what it means to say that concepts are theories, or that they have a theoretical structure. In particular, I will review and assess two ways in which cognitive psychologists have interpreted the ‘theory-theory’, which for the sake of convenience, I will call the ‘literal’ and ‘liberal’ interpretations. My contention is that both interpretations lead to serious difficulties. The ‘literal’ approach is undermined by problems similar to those of prototype theories of concepts, problems to which it was originally intended to be a solution. The ‘liberal’ approach fares no better, for it yields an account of conceptual structure that exceeds the resources of cognitive economy. Furthermore, I will argue that there cannot be a third way in which to reformulate the ‘theory-theory’. As such, the problems generated by the ‘literal’ and ‘liberal’ approaches render the ‘theory-theory’ an untenable theory of concepts. I will conclude by examining some possible objections that may be raised to the critique presented here.

Two caveats are in order. First, due to the constraint of space, the review of various psychological theories of concepts in the second section will not be a comprehensive one.[2] Its primary goals are simply to take note of some of the major problems that have plagued some of the earlier theories of concepts, and that have motivated and subsequently led to the adoption of the ‘theory-theory’. Second, the term ‘concept’ will be used here in the same sense as cognitive psychologists use it. This in no way implies that their usage is better but merely that it is in the domain of the cognitive sciences that our critique of the ‘theory-theory’ will take place. While the arguments presented below can be expanded for application to other uses of ‘concept’, this task can only be reserved for another occasion.

Two Shifts in Conceptual Structure

To cognitive psychologists, concepts are understood to be mental representations of categories, the ‘building blocks for human thought and behavior’ (Medin 1989, 1469). ‘Category’ in this sense broadly refers to ‘a partitioning or class to which some assertion or set of assertions might apply’ (Medin 1989, 1469). Under this rather loose definition, the diversity of categories includes what some philosophers call ‘natural kinds’, such as gold and water, fictional entities such as ghosts and unicorns, and entirely arbitrary or ad hoc categories. For example, the substance gold constitutes a category because we can say of it (or its samples) that it, among other things, has the atomic number of 79. Also, the group of things consisting of a bubble gum wrapper, a computer, and a paper clip may count as a category because we can say of these things that they are physical objects that can be found on the desk in front of me. According to cognitive psychologists, in order for us to think, to reason or to learn about anything (or any category), we have to represent them in our minds. That is, we have to form concepts. A major concern of cognitive psychology is therefore to determine what kind of structure concepts possess.

In his classic paper ‘Concepts and conceptual structure’, Douglas Medin (1989) observes that in the 20th century, there have occurred two radical shifts in the way cognitive psychologists have theorized about conceptual structure. The first shift occurred when prototype theories replaced the so-called ‘classical’ theory. In the ‘classical’ theory, a concept is construed as an abstract summary representation of the necessary and sufficient features or properties that all things subsumed under the concept must possess. For instance, the concept BACHELOR is a mental representation which embodies the defining properties ‘is male’, ‘is an adult’ and ‘is unmarried’.[3] Whether or not something is to be categorized as a bachelor depends on whether it possesses these three requisite properties. If it does, then it is a bachelor; if not, then it is not.

A serious difficulty with the ‘classical’ theory is the noticeable lack of bona fide definitions. As philosophers and psychologists have demonstrated, many concepts, especially ordinary, non-scientific ones, simply do not have clear-cut definitions (e.g. Fodor et al. 1981; Wittgenstein 1953). Evidently, if there are no definitions, then these concepts cannot have a definitional structure. Another objection to the ‘classical’ theory is raised by attention to the so-called ‘typicality effects’. Much of the research on categorization by Elenor Rosch and her colleagues (e.g. Rosch and Mervis 1975) suggests that people, contrary to what the 'classical' theory assumes, do not treat all members of a particular concept to be equally good examples. Instead, they look upon some members as more typical or 'better' than others and this disposition in turn shapes the way in which they categorize and reason. For example, people not only judge a sparrow to be a more typical example of a bird than a penguin, but they also categorize the former at a much faster rate. The pervasive influence of such 'typicality effects' has convinced some critics that the 'classical' theory is flawed in its account of conceptual structure that presumes that all members subsumed under a concept are of equal weight.

Problems with the 'classical' theory caused the first shift in the theorizing about conceptual structure. An alternative approach developed in the 1970s to interpret category membership as 'more graded and structured in terms of features that are only typical or characteristic of categories' (Medin, Lynch, and Solomon 2000, 123–24). Like its 'classical' predecessor, this 'probabilistic' or prototype view holds that concepts are to be construed as lists of properties. Unlike the 'classical' theory, the prototype theory does not require these properties to be necessary and sufficient, but merely stereotypical among instances of a concept, where the typicality of a property is determined by the frequency with which it appears among instances of the concept. For example, for the concept BIRD, the property 'flies' is considered more typical than the property 'is white', since more birds fly than birds that are white. Also, the prototype theory differs from the 'classical' in that the determination of whether an entity is subsumed under a concept depends on whether it satisfies a 'sufficient' number, or a 'critical sum', of stereotypical properties. Put differently, something is categorized under the concept if it is similar to the prototypical representation.
For this reason, prototype theories are sometimes described as 'similarity-based' theories.

Prototype theories are not, however, problem-free. Concepts in prototype theories recognize only the central tendency but not information about category size, the variability of the examples, and correlations and embedding relations among properties (Medin 1989, 1472). An example given by Medin is the intuition that many people have of small birds being much more likely to sing than large birds. According to him, this intuition cannot be obtained from the prototype representation of BIRD because features encoded therein are treated as independent and additive. Thus, even though the prototype representation encodes both the properties 'is small' and 'sings', it does not identify any correlation between them.[4] Since prototype representations leave out information to which, as experimental studies have shown, people are sensitive, and which they incorporate in their categorization tasks and other reasoning activities, Medin concludes that prototype theories provide at best an inadequate account of conceptual structure.[5]

Another weakness of prototype theories is that the notion of similarity is too vague. It is too flexible and unconstrained for precise purposes. As mentioned, whether an entity is to be counted as a member depends on how similar it is to the prototype representation, where similarity is measured partly in terms of the number and kinds of properties members share. What properties are listed in the representation is therefore of critical importance for categorization, since different properties will lead to different categorizations. The problem with prototype theories is that they do not provide any explanations as to why certain properties are selected as part of the prototype representation. As Nelson Goodman (1972) has pointed out, there is in principle an infinite number of properties which two objects, however perceptibly dissimilar they are, can share. For example, a beach ball and a computer are both 'not a whale', 'not a pair of scissors', etc. Should the objects then, by virtue of sharing these trivial properties, be considered 'similar' and thus, subsumed under the same category? The answer, it would seem, hinges on the criteria of conceptual representation and the properties that they invoke. A theory of concepts must therefore be capable of explaining why certain properties are chosen to appear in the prototype representation. In simply compiling lists of such properties, prototype theories do not meet this requirement.

‘Concepts Are More Than Lists’

The objections that I have considered so far all suggest that prototype theories are inadequate: they fail to incorporate information crucial for categorization and to explain why certain properties are selected under the prototype representation. In general, prototype theories are deficient in that they ‘seem to discard too much information that can be shown to be relevant to human categorization’ (Medin 1989, 1472). Such a problem, according to Medin, helped trigger the second shift, namely from similarity-based theories to ‘theory-based’ theories, or the ‘theory-theory’.

How does the ‘theory-theory’ differ from similarity-based theories? As a starting point, Medin’s terse statement is instructive: ‘Concepts are more than lists [of properties]’ (Medin 1989, 1476). Such lists are, to be sure, useful but are insufficient in explaining conceptual structure. An adequate theory of conceptual structure must provide more than the listing of all the stereotypical properties shared by members of a concept. At a minimum, concepts must include the information which is ignored by prototype theories, namely information about category size, interproperty relations, member variability, and more generally, anything ‘that can be shown to be relevant to human categorization’. At a first approximation, then, concepts can be said to have a theoretical structure in that they encode lists of properties, and explanatory and underlying principles that account for the aforementioned deficiencies. Is this what cognitive psychologists mean when they claim that concepts have a theoretical structure?

Here, we have to be cautious about what cognitive psychologists mean by the first occurrence of ‘theory’ in ‘theory-theory’. Different interpretations are at play. As Laurence and Margolis note, ‘there is considerable disagreement about how lenient one should be in construing a body of representations as a theory’ (1999, 44). Some employ the term in the sense of the natural sciences and demand that concepts possess the same characteristics as those of scientific theories. Others use it without stringent conditions and construe concepts in a more liberal fashion. Let us now consider in turn two approaches, which I label the ‘literal’ and ‘liberal’ interpretations.

Cognitive psychologists of the ‘literal’ interpretation include Alison Gopnik, Andrew Meltzoff, and Henry Wellman. In ‘The theory theory’, Gopnik and Wellman (1994) identify three characteristics which they insist a theory must have. First, a theory must be abstract in the sense that it posits theoretical entities or constructs in order to provide causal explanations for the empirical phenomenon in question. An example is the atomic theory in chemistry, which posits protons and electrons in order to explain, among other things, why various chemical substances have the properties they do (e.g. boiling point, color, acidity, etc). Second, consequent upon the previous requirement, a theory must introduce a ‘different set of vocabulary’ specific to the subject matter under investigation (1994, 260).

What Gopnik and Wellman mean by this is that prior to the formulation of a theory, the empirical phenomenon to be explained would have to be described in terms of an ‘evidentiary’ vocabulary. Since a theory goes beyond the phenomenon by positing ‘hidden’ theoretical entities, it will need corresponding terms to refer to these entities. That is why theories must have a novel vocabulary. Lastly, the third characteristic is that a theory must be coherent in the following way: ‘The theoretical entities and terms postulated by a theory are closely, “lawfully”, interrelated with one another [and with the evidence]. As a result, changes in one part of a theory have consequences for other parts of the theory’ (1994, 260–61).

By virtue of these three characteristics, Gopnik andWellman claimthat a theory is able to provide explanations, interpretations and predictions of the empirical phenomenon in question. By the same token, anything lacking in any of these characteristics will not qualify as a theory. In fact, Gopnik and Wellman point out that empirical typologies and generalizations are not theories but are merely ‘orderings, partitionings, and glosses of evidence and experience [that] share the same basic vocabulary as the evidence itself’ (1994, 260–61). For example, in classifying plants, we may decide to put all plants with stems in one group, and those without in another. Within the group of stemmed plants, we may further divide them into sub-groups of over or under one-foot tall. This classification of plants will not count as a theory, by Gopnik and Wellman's standards, since it is based solely on observable or evidential properties (e.g. we can see stems). Moreover, the resultant categories of sub-groups do not appeal at all to any theoretical constructs, and thus do not lead to any new vocabulary.

This 'literal' view of theory stands in sharp contrast to the less rigorous, 'liberal' interpretation of theory proposed by Murphy and Medin. Here is what they say about the 'theory-theory':[6]

> When we argue that concepts are organized by theories, we use theory to mean any of a host of mental 'explanations,' rather than a complete, organized, scientific account. For example, causal knowledge certainly embodies a theory of certain phenomena; scripts may contain an implicit theory of the entailment relations between mundane events; knowledge of rules embodies a theory of the relations between rule constituents; and book-learned, scientific knowledge certainly contains theories. (1985, 426)

The contrast to be drawn between this interpretation of theory and Gopnik and Wellman's 'literal' approach is not one between a complete, organized scientific theory and an incomplete, unorganized one.[7] While Gopnik and Wellman do impose strict conditions for what qualifies as a theory, they do not require theories to be complete. In fact, the very possibility that incomplete theories can undergo change is what partly motivated them to view conceptual change as analogous to changes in scientific theories. Moreover, to demand that people's concepts embody complete scientific theories is implausible, for many people lack knowledge of such theories.

In short, what qualifies as a theory for Murphy and Medin need not satisfy all three of Gopnik and Wellman's stipulated conditions. To Murphy and Medin, something already counts as a theory if it provides 'underlying principles', such as 'causal connections, script links, and explanatory relations' (1985, 436), to explain any or all of the following: how concepts are internally structured, why certain properties are included in a conceptual representation, why some of these properties are correlated, and/or how concepts are related to one another. In their view, scripts and empirical typologies will count as a theory because they explain why certain things are categorized or grouped together as they are. The implication seems to be that insofar as a principle, generalization, or even a claim, performs any of the aforementioned explanatory functions, it is integral to conceptual representation.
As to how much information a concept should represent, Murphy and Medin offer the following characterization: the theory-based approach 'expands the boundaries of conceptual representation' to include 'all of the relations involving that concept and other concepts that depend on it' (1985, 436). Only by doing so can theories embody knowledge about concepts and their use. This expansive treatment, however, leads to problems of its own. As we shall see, it is in its very attempt to be all-inclusive that Murphy and Medin's 'theory-theory' encounters serious difficulties.

Neither 'Literal' nor 'Liberal'

In this section, I will argue that neither the 'literal' nor the 'liberal' interpretation is tenable. Moreover, I will show that there cannot be a third way in which to reformulate the 'theory-theory'. I begin with the 'literal' interpretation. Recall that the central difficulty with prototype theories is that they 'seem to discard too much information that can be shown to be relevant to human categorization' (Medin 1989, 1472, quoted above). A prototype representation tends to abstract only the 'central' properties from instances of a concept and consequently, omit information which is relevant to the way people categorize, such as information regarding interproperty relations, variability of members, and category size. Such omissions are what led to the demise of the prototype theory and the adoption of the 'theory-theory'.

However, the 'literal' interpretation of the 'theory-theory' also fails to capture information relevant for categorization. Consider the following scenario. Suppose that James acquires the concept MOLD in his biology class and understands mold to be microscopic organisms composed of filaments called hyphae. By producing spores, hyphae can germinate to form a cottony mass called mycelium, commonly found in foods (e.g. bread) that have gone bad. Moreover, James finds out that hyphae not only can cause serious health problems such as skin allergies and blurry vision, but can also cause their hosts (e.g. woods or foods) to decay. By Gopnik and Wellman's criteria, what James has learned about mold qualifies as a theory: it is abstract (the positing of 'hidden' theoretical entities such as filaments to explain colony mass), it has a new vocabulary ('hyphae' and 'filaments'), and it is coherent (e.g. the lawful relationship between hyphae and mycelium).

Now, for the sake of argument, suppose also that James has a grandmother who, for whatever reason, always shrieks when and only when she sees mold. That is, her shrieking is reliably caused by the presence of mold (to make this situation more realistic, we may further assume that his grandmother used to be a micro-biologist who over the years has developed a keen eye for identifying mold but has yet to overcome her fear or disgust of it). Having repeatedly witnessed his grandmother's shrieking upon seeing mold, and remembering no other times in which she has shrieked, James forms the belief, perhaps naturally so, that his grandmother shrieks when and only when mold is present. In this context, whether James' belief is true is unimportant. What is important is that if James comes across some stuff in the future which causes his grandmother to shriek, he will automatically interpret her behavior in terms of his belief and judge that stuff to be mold. In fact, his belief may even be his sole basis for identifying mold, if, for instance, the stuff in question is a kind of mold he has never seen before.[8]

What is crucial to note here is that James' judgment constitutes an act of categorization. According to Smith and Medin, categorization is understood as the act of determining whether 'a specific instance is a member of a concept' (1981, 6). Similarly, categorization establishes 'points of contact between previous situations and the current context' (Medin 1989, 1469). Accordingly, James' judgment counts as an instance of categorization because he is subsuming some new stuff (i.e. that which alarms his grandmother) under the concept MOLD. However, even though James appeals to his belief in categorizing instances of mold, his belief does not count as part of the scientific theory of mold which he has learned in his biology class. There are two reasons for this. First, his belief does not cohere (in Gopnik and Wellman's sense) with the theory. For example, as far as James knows, there exist no 'lawlike' relationships between any theoretical constructs (e.g. hyphae) and his grandmother's shrieks.[9] As a result, there will be no consequences for the theory even if James ceases to believe that mold causes his grandmother to shriek. Second, this belief, qua an empirical generalization, does not import any new vocabulary into the theory, since it is couched purely in evidentiary terms (e.g. 'grandmother', 'shriek', etc). As such, James' belief, by Gopnik and Wellman's standards, cannot be a part of the scientific theory. Nevertheless, his belief is vital and relevant for categorization: grandmother's shrieking is a way in which James can identify mold. In short, James is appealing to some piece of information outside of the theory in order to categorize.[10] For this reason, the 'literal' interpretation of the 'theory-theory' faces the same problem as prototype theories: it fails to incorporate 'information shown to be relevant for categorization purposes'. As such, it must also fail to provide an adequate account of conceptual structure for the same reason.

Before proceeding, it is appropriate first to consider a possible objection. My argument, as it is presented, hinges on a twofold requirement. First, the presence of mold and James' grandmother's shrieks have to be correlated in a law-like fashion, or else James would not have formed the belief that his grandmother shrieks when and only when she sees mold. Second, law-like correlations involving observables (i.e. empirical generalizations) have to be taken as non-theoretical or else James would not have appealed to something outside of theory when he categorized mold. An objection might thus arise that the latter requirement is simply too restrictive, as it will disqualify many scientific laws from being considered theoretical. For instance, the geological law that rivers erode their banks, qua a law formulated entirely in terms of observables, cannot be regarded as theoretical.

Since this requirement is overly restrictive, the objection continues, it should be abandoned, in which case empirical generalizations and law-like correlations will be accepted as theoretical. By the same token, James' belief must also qualify as theoretical, and in appealing to it to categorize mold, he does not, in fact, step outside the bounds of theory. The 'literal' view may therefore be safe.

In response, two aspects are pertinent. First, the idea that empirical generalizations do not qualify as theories stems from the view of Gopnik and Wellman that anything theoretical must satisfy the three criteria mentioned above. My criticism of their view is premised on a conditional: If theory is construed in this restrictive way, then any appeals to empirical generalizations in categorization would expose this formulation of the 'literal' view as inadequate, for it would fail to incorporate information relevant for categorization in much the same way as prototype theories. Second, I concede that law-like correlations should be regarded as theoretical. On the face of it, such a concession would seem to offer a way out of my earlier criticism, for theory is now expanded beyond Gopnik and Wellman's three criteria to include empirical generalizations and other law-like correlations.[11] Accordingly, James would no longer be going outside of theory when he appeals to his belief to categorize mold.

This alternative formulation and defense of the 'literal' view, however, can also be shown to be deficient. This is because my criticism can be reformulated in such a way that it does not require a law-like correlation between the presence of mold and grandmother's shrieks.[12] The thought experiment can be so designed as to stipulate that James' grandmother only sometimes shrieks at mold.[13] For example, there are times when she sees mold but is too tired or otherwise preoccupied to shriek. Well aware of this, James forms the belief that his grandmother sometimes shriek at the sight of mold. Notice that James' belief is not formed from empirical generalizations or any law-like correlation. From James' perspective, whether his grandmother will shriek at mold or not depends entirely on contingent factors like her mood and her state of mind. Nonetheless, the next time James hears her shriek, he will still come to the conclusion that the stuff in front of his grandmother is mold.[14] In such an event, he will still appeal to a belief outside of theory when he categorizes mold. Consequently, this alternative reformulation of the 'literal' view faces the same problem.

The objection may still persist. If the 'literal' view is supposedly committed to the idea that a concept must include any relevant information for categorization, then why not simply include beliefs of the above kind as part of the theory? Thus, James' peculiar belief about grandmother's shrieks and mold will be incorporated into his theory of mold. With this, another version of the 'literal' view seems to be in the offing, one that satisfies Gopnik and Wellman's three criteria, includes empirical generalizations and lawlike correlations, and contains any belief that is relevant for categorization. Is this third reformulation defensible? On scrutiny, it is not. The inclusion of such beliefs will inevitably inflate conceptual structure. A person can possess numerous beliefs about a particular kind of thing, most of which are relevant for his or her categorization purposes. To pursue the thought experiment, the array of James' beliefs may include that mold is disgusting, that it is to be avoided and if found, eliminated, that it can be found in places like under the kitchen sink, that grandmother typically combats mold with a certain brand of disinfectant, that his friend Mary is not bothered by mold, and so on. All of these beliefs can be used by James (whether singly and jointly) to categorize mold. For example, he can infer that the substance in question is mold after hearing his grandmother describe it as disgusting and/or seeing her reach for the disinfectant.[15] If these beliefs are incorporated into theory, conceptual structure under the 'literal' view will uncontrollably expand.[16]

Consider another example. People often categorize birds by appealing to their stereotypical or superficial properties, all of which are plausibly included in their (naı¨ve) theories of birds. However, they can also do so by appealing to a myriad of other beliefs, many of which may be deemed trivial, accidental, or even false: that city birds are dirty, that some birds are scavengers, that birds are my friend John's favorite creatures, that birds migrate south, that birds sometimes serve as a symbol of freedom in literature, that birds are signs of good or bad luck, and so on. Again, all of these beliefs can be used (either singly or jointly) to identify and categorize birds. To include all of these beliefs in theory would no doubt inflate the conceptual structure for birds endlessly.

Two points merit attention here. My criticism of the 'literal' view, whether the original or derived versions, can be generalized beyond cases in which people have strange grandmothers. As the above examples show, it is quite a common occurrence for people to appeal to extra-theoretical beliefs when they categorize. This being the case, the 'literal' interpretation is just as defective as the prototype theories, which it is intended to replace. Also, if all of the beliefs relevant for identification and categorization are incorporated in a theory, then this modified 'literal' theory will incur problems of manageability, ones that are similar to those that plague the 'liberal view', to which our discussion will presently turn.

The 'liberal' interpretation may at first glance offer an exit out of the difficulty faced by the above formulations of the 'literal' view. After all, its main criterion by which something can be judged to qualify as a theory, or some part thereof, is its ability to explain the aforementioned relations covered under a concept. James' belief, insofar as it helps him categorize mold, will be incorporated as part of the theory in this interpretation. Problems nevertheless persist. Whereas the 'literal' construal of theory restricts the inclusion of information, the 'liberal' construal runs the risk of being excessively broad—so much so that it renders no useful account of conceptual structure. The latter offers an account of conceptual structure so generous that it turns out to be of very little practical use. For instance, it will have to include: (1) lists of properties, necessary, sufficient, and/or stereotypical; (2) all of the previously ignored information in 'classical' and prototype theories regarding interproperty relations, embedding relations, category sizes, and member variability; (3) empirical generalizations, scripts and typologies; (4) theoretical beliefs ('theory' in the 'literal' sense); (5) extra-theoretical beliefs like James' belief about his grandmother's shrieks. Finally, it is obliged to take in (6) information that is trivial. To explain this, it is instructive to revisit, this time in full, an earlier excerpted quote from Murphy and Medin:

> In general, it can be seen that the similarity-based approach requires a minimum of conceptual organization and relations, whereas the theory-based approach emphasizes both. One way to describe this difference is to say that the theory-based approach expands the boundaries of conceptual representation: In order to characterize knowledge about and use of a concept, we must include all of the relations involving that concept and the other concepts that depend on it. To explain conceptual coherence, the processes that operate on a concept must be considered in addition to the information directly stored with it. (1985, 436)

Something in this quote merits close attention. In order to characterize knowledge about a concept and its use, 'all of the relations involving the concept and other concepts that depend on it' (my italics) must be incorporated. This requirement seems somewhat extreme. Consider, again, the concept BACHELOR. It enters into relations with many things, the concepts of which depend on it. For example, it enters into relations with complex concepts of which it is a constituent part: BACHELORS WHO APPEAR ON TV REALITY SHOWS, BACHELORS WHO DON'T LIKE BREAD, BACHELORS WHO OWN BLACK CROWS. The list goes on ad infinitum. If Murphy and Medin are correct, then all of these relations must be represented under the conceptual structure of BACHELOR.

The pressing question is: should all of the above information be represented under a concept? The answer is a resounding 'no'. As Lloyd Komatsu points out, when theorizing about conceptual structure, it is important to keep in sight the trade-off between a concept's economy and its informativeness: the more economical a concept is, the less informative it becomes, and vice versa. Accordingly, the 'classical' and the prototype theories are highly economical but uninformative, since they rely only on 'a single representation [that] serves an entire category, representing all instances for all purposes' (Komatsu 1992, 502). This representation will then be utilized to perform all sorts of cognitive tasks, such as planning, combining concepts, and making inferences. However, since only the central properties (namely, the definitional and the stereotypical) are represented, concepts in such theories suffer ‘a loss in the amount of detail’ that they can transmit (Komatsu 1992, 502). In contrast, ‘theory-based’ theories are highly informative but not very economical. The concepts that they denote are multiplex in nature; even ‘a single word or mental category implies multiple representations’ (Komatsu 1992, 502).[17] Since these multiple representations will be accessed each time the concept is tokened, the heavy burden on cognitive economy will be unbearable.

It should be clear from the above discussion that the problem with the ‘liberal’ interpretation of the ‘theory-theory’ is that if conceptual representation is supposed to include ‘all of the relations involving the concept and the other concepts that depend on it’, plus the information omitted by prototype theories (e.g. interrelation among properties, embedding relations, variability of members), then concepts will not be economical at all; all of the above information will have to be accessed every time the concept is used. This in turn will generate significant problems at the computational level and unduly tax the cognitive resources, such as working memory and processing capacities.[18] Furthermore, according to Komatsu, it will also raise issues related to the ‘frame’ problems and knowledge access prevalent in the field of artificial intelligence (1992, 516). In brief, any account of conceptual structure that results in mental exhaustion cannot be a viable option.

To conclude this section, it suffices to explore briefly whether there is a third way in which the ‘theory-theory’ can be reformulated such that it can steer clear of the foregoing problems. Let us start by considering what this third interpretation should look like. From the preceding discussion, it is clear that conceptual structure should embody more than what the ‘literal’ interpretation (i.e. Gopnik and Wellman’s) permits but less than what the ‘liberal’ interpretation allows. It would appear that the third alternative will be situated somewhere between these two and as a corrective, will comprise only some of the relations stipulated by Murphy and Medin. The issue thus boils down to deciding which parts of the conceptual structure under the ‘liberal’ interpretation to keep and which parts to exclude. However, this is not exactly feasible, for all of the information included under the ‘liberal’ interpretation is assumed to be relevant to categorization. Again, as I have stressed above, this assumption served as the primary motivation behind the adoption of the ‘theory-theory’. If any such information is to be omitted, this refurbished interpretation of the ‘theory-theory’ will backtrack to the beaten path of the prototype and the ‘literal’ interpretation in discarding information needed for categorization.

Notice that such a middle-ground position of the ‘theory-theory’ may well be filled by the two modified reformulations of the ‘literal’ view which I have examined above. Both acknowledge the inadequacy of Gopnik and Wellman’s overly restrictive construal of theory and try to rectify it by incorporating empirical generalizations and/or beliefs into the conceptual structure. Each, however, as explained, turns out to be problematic. One fails to include beliefs relevant for categorization, while the other generates an unwieldy problem of inclusivity, one that is similar to that of the ‘liberal’ view. Other attempts to arrive at a middle-ground position will likewise fail. For instance, Jesse Prinz suggests that ‘a natural proposal for distinguishing theoretical beliefs from other beliefs [is to] say that theories include just beliefs about ontological category membership, essence, and explanatory structure’ (2002, 87). Thus, Medin and Ortony’s psychological essentialism (1989) postulates that people believe that things subsumed under a single category share a common set of essential properties (or an ‘essence’) and that these properties play a crucial role in explanations (e.g. category membership) and in the performance of cognitive tasks such as categorization. In this light, theoretical beliefs are limited to those concerned with such properties, including what Medin and Ortony have called ‘essence placeholders’. Similarly, Keil (1989) demarcates conceptual structure by emphasizing information that is relevant for explaining causal relations and structures as they pertain to ontological matters. For our purposes, there is no need to explore these theories in detail. Suffice it to note that any such attempts to delimit theory will result in the omission of some non-theoretical information that is relevant for categorization, like accidental properties and mistaken beliefs.[19] For instance, James’ belief that mold is disgusting is neither related to the ‘essence’ of mold nor to its ontological status. To conclude, there cannot be a third way in which to reformulate the ‘theory-theory’. The problems with the ‘literal’ and ‘liberal’ approaches, as well as the exhaustion of options, pose serious difficulties for the ‘theory-theory’, making it a problematic view of concepts.[20]

Objections and Implications

One way in which advocates of the ‘liberal’ interpretation of the ‘theory-theory’ can respond to the above objection that its permissive account of conceptual structure violates cognitive economy is to reject the assumption that all the information included under the conceptual representation is accessed every time the concept is used. Instead, they canmaintain that only a subset of such information will be accessed each time the thinker employs the concept. Whichever subset will be involved depends on the specific cognitive task at hand and on its immediate context. As Solomon et al. (1999) point out, experimental studies have shown that the kind of information represented under a concept and thus, available to the thinker, varies with the kind of cognitive task he has to perform, be it categorization, communication, planning, or conceptual combination. In other words, a concept can have multiple representations, each of which corresponds to a specific context and cognitive task. For instance, for the concept ELM, a person may have access to information about the aesthetic properties of elm trees if he has a landscaping project in mind. In contrast, he may have access to the trees’ biological and chemical properties, if he is engaged in a different cognitive task, such as preparing for a botany exam. Moreover, the information available in one context need not be made available in others. In this case, information regarding the molecular make-up of elm trees will typically not be accessed by the landscaper, whose main concern is to select the best way to arrange his elm trees in his garden.

Theory-theorists can therefore overcome the cognitive economy objection by arguing that since only a subset of information will be retrieved for the performance of each specific cognitive task, there will no longer be any economy problems pertaining to memory and computational capacities. Unfortunately, this response will not do. Even if this defense neutralizes the initial objection, it gives rise to a fresh problem. The key idea behind the theory-theorists’ response is that a concept can have multiple representations associated with it in the sense that different information can be represented under different cognitive tasks. Notice, however, that each of these representations is supposed to correspond or belong to the same concept. Returning to our above example, the information regarding the chemical and biological properties, and the information regarding the superficial properties, are both supposed to be of elm trees. The question here for the theory-theorist is: by virtue of what are these various representations representations belonging or corresponding to the same concept? Put another way, if a concept can have different representations associated with it, what ensures that all of these representations are of the same concept?

The ‘liberal’ theory-theorists may be tempted to claim that there is some common element or intrinsic structure that different representations possess such that it makes them correspond to the same concept. In the case of ELM, they may claim that the representations corresponding to the landscaping task and the exam-writing task may both include a commonminimal structure such that itmakes both of themrepresentations of elmtrees. Such a structuremay be considered a ‘core’ part of the concept ELM which is directly responsible for making them both representations of elms. However, this response is clearly not an open option to theory-theorists. In light of our review of the ‘classical’ and prototype theories above, any such attempt to carve out a minimal structure or core part has failed: such a minimal structure will always turn out to be insufficient for the concept, as it will omit information which is relevant for categorization.[21] These are precisely the drawbacks of prevailing theories which the ‘theory-theory’ is formulated to address in the first place.[22]

In short, the ‘liberal’ interpretation cannot overcome the cognitive economy problem. Given that the ‘literal’ interpretation of the ‘theory-theory’ also has its share of problems, and that we have shown that there cannot be, in principle, a third way in which we can construe the ‘theory-theory’, it is clear that the ‘theory-theory’ is an untenable theory of concepts.

What implications does this critique against the ‘theory-theory’ have with respect to theories of concepts in the cognitive sciences and in the philosophy of mind? I suggested earlier that this argument can be used to support Fodor’s conclusion that most theories of concepts in the cognitive sciences and in the philosophy of mind are untenable. The extent to which it does so depends on how one views the status of the ‘theory-theory’. If one accepts Medin’s point that the ‘theory-theory’ is supposed to replace the prototype theory, and that the prototype theory is in turn supposed to replace the ‘classical’ theory, a harsh verdict follows: since the ‘theory-theory’, the latest major solution, is misguided, the cognitive sciences do not have a working theory of concepts. If, on the other hand, one concedes that prototype and/or ‘classical’ theories can be rehabilitated with their problems exposed by theory-theorists corrected, a more lenient verdict will ensue: at least the ‘theory-theory’ is out of the race. In this case, further arguments will be needed to dislodge these revived theories.

There is reason to think that the harsh verdict follows. Attempts to salvage the ‘classical’ and prototype theories have geared mostly towards bridging gaps so that previously ignored information relevant to categorization tasks and reasoning will be admitted. Some, for instance, have argued that the prototype theory may be adjusted to explain correlations among properties by positing ‘labeling’ relations or by combining the correlated properties into a conjunctive one (Smith and Medin 1981). Other efforts at improved revisions also aim to overcome the initial limitations by expanding their scope to incorporate whatever information that critics deem essential. However, as argued above, the problem is a far deeper one. A shared assumption among cognitive scientists insists that concepts must embody all information relevant to categorization. Yet, since there is no definite or cut-and-dry way of determining exactly what or what not to include, the tendency has been to expand coverage as far as possible in order to secure the conceptual structure. Concepts are supposed to encode all that is, or appears, relevant to categorization. Sooner or later, and already in the case of ‘liberal’ theory-theorists, the inevitable outcome is one of straining cognitive economy. Thus, the harsh verdict seems more pertinent.

I also indicated earlier that this critique of the ‘theory-theory’ is much stronger than the arguments advanced by Fodor against psychological theories of concepts. Fodor’s arguments all revolve around what he calls the five ‘non-negotiable’ conditions: (1) concepts are mental particulars; (2) concepts compose; (3) many concepts are learned; (4) concepts are public; and (5) concepts are categories (1998). According to Fodor, any adequate theory of concepts must meet these five conditions. However, as he notes, the ‘classical’, the prototype and the ‘theory-theory’ approaches all fail to do so. For instance, prototype theories violate the compositionality constraint, and thus cannot account for how human thought can be both productive and systematic. Similarly, the ‘theorytheory’, committed as it is to holism, the view that the content of a concept is individuated by the thinker’s complete set of beliefs, violates the publicity constraint. Consequently, Fodor concludes that there are currently no tenable theories of concepts in the cognitive sciences and in the philosophy of mind.

However, Fodor’s five conditions have invited criticisms over the years. Philosophers and psychologists have questioned, in some context or other, whether such conditions are really non-negotiable. For example, not all philosophers agree that compositionality is required to explain the productivity and systematicity of thoughts (e.g. Schiffer 1987). Clearly, if compositionality is not required, then prototype theories could still be in the running as a viable interpretation. Also, some of the specified conditions seem to be tailored to suit Fodor’s need to argue his own theory of concepts (Informational Atomism) and his Representational Theory of Mind (RTM). Thus, if one does not accept his RTM, there is no reason why one must consider the ontological status of concepts in terms of mental particulars and not in terms of epistemic capacities. In contrast, notice that my critique of the ‘theory-theory’ makes no explicit use of Fodor’s five conditions. It appeals only to internal elements of cognitive psychology and does not rely on any controversial philosophical assumptions. As such, it is purely an internal critique, and avoids the vulnerabilities of Fodor’s arguments.

ACKNOWLEDGEMENTS

I am grateful to William Seager, Ronald de Sousa, Gurpreet Rattan, and Jennifer Harker for their help and advice in writing this paper. I would also like to thank two anonymous referees of this journal for their insightful comments and suggestions.

NOTES

1. The ‘theory-theory’ is also sometimes referred to as a ‘theory-based’ or ‘explanation-based’ theory of concepts. It is important to distinguish this sense of ‘theory-theory’ from what some philosophers, such as Nichols and Stich (2003), have also labeled as the ‘theorytheory’, which primarily relates to self-awareness.

2. For a detailed overview of these theories, see Komatsu (1992), Laurence and Margolis (1999), Medin (1989), Murphy (2002), and Smith and Medin (1981).

3. Concepts will be denoted by SMALL CAPITALS.

4. To see this, consider the following: For the concept BIRD, the following stereotypical properties may be represented under the prototype: ‘flies’, ‘sings’, ‘is brown’, ‘has a beak’, ‘is small,’ ‘has webbed feet’, etc. If correlations among properties can be represented simply by virtue of listing each of the correlated properties in the representation, then many correlations will result: small birds tend to sing, small birds tend to be brown, brown birds tend to sing, flying birds tend to sing, etc. However, many of these correlations are clearly false. Thus, representing correlations among properties requires something more than the mere listing of properties.

5. For instance, recent evidence from developmental psychology has shown that children’s concepts turn out to be rather complex. Many of their concepts contain information about ontology, functions, causation, intentions, and ‘hidden’ and ‘non-obvious’ properties (such as an organism’s innards or an object’s ‘essence’), to which children appeal when they categorize (Carey 1985; Gelman and Koenig 2003; Gelman and Wellman 1998; Keil 1989; Medin and Ortony 1989).

6. Murphy and Medin explicitly state that they are not proposing a new model of conceptual representation (1985, 426). Our purpose here is merely to summarize what they mean by theory in order to arrive at a manageable construal of the ‘theory-theory’.

7. Another contrast that can be drawn here is that Murphy and Medin are primarily interested in adult concepts, whereas Gopnik, Meltzoff, and Wellman are mainly concerned with concepts held by children. This contrast, however, is of minor importance in the context of our current discussion because our central purpose is merely to come up with possible ways in which to construe the ‘theory-theory’.

8. In what follows, I will assume that grandmother’s shrieks serve as the only way in which James identifies mold. Of course, there will be occasions in which her shrieks will serve as additional evidence for identifying mold and still others, in which they do not come into play. Notice that it does not matter much to the argument just which role grandmother’s shrieks play. So long as James’ belief about grandmother’s shrieks can be used on occasion to categorize mold, it would have to be included as part of the conceptual structure. After all, critics of prototype theories have insisted that correlations be included as part of the concept even though such relations only sometimes play a role in categorization. Moreover, features listed under a prototype do not always factor in the performance of categorization tasks; that is why they are probabilistic.

9. It may be tempting to refer to the reliable relationship between mold and grandmother’s shrieking as a ‘lawlike’ relationship. However, recall that Gopnik and Wellman intend such

relationships to involve a theoretical construct or term. Since the relationship between mold and grandmother's shrieking only pertains to observable phenomena, it is best described as an empirical generalization. Also, it is possible that there could be a lawlike relationship which exists between hyphae and his grandmother's shrieks. However, as far as James knows, this is not something entailed by his theory of mold. As such, his belief must be considered outside of theory.

10. Some may object that my notion of categorization is not the same as that employed by cognitive psychologists. For instance, their notion pertains to spontaneous, quick, or 'intuitive' acts of subsuming things under a category, whereas mine concerns a much broader phenomenon, including categorization which results from deliberation or inferences. In response, I point out that there is no reason to think that James' categorization judgments involving his belief about his grandmother's shrieks cannot be of the spontaneous or the intuitive kind; in fact, James' belief may have been an implicit one all along, never enjoying a moment of conscious reflection. Moreover, it has become 'second nature' for him to associate grandmother's cries with mold. If this is true, then my argument against the 'literal' interpretation holds: he would still be appealing to something outside of theory to categorize (in the same sense of categorization employed by cognitive psychologists).

11. For example, Keil (1989), Murphy and Medin (1989), and Rips (1989) have suggested formulations of theory that incorporate, among other things, empirical generalizations.

12. Incidentally, this avoids another objection: an empirical generalization could still be regarded as theoretical because the requirement for theory is not that the generalization itself has to satisfy the three criteria set forth by Gopnik and Wellman, but that the overall theory, of which the generalization is a part, satisfies them. Thus, the empirical generalization could be incorporated under theory solely on the basis of it being a law-like correlation. In my view, this objection fails because my reformulated thought experiment no longer relies on such a correlation.

13. Here is another way in which the thought experiment can be modified. Suppose that James has witnessed only one occasion in which grandmother shrieked at the presence of mold. However, it was such a memorable and vivid experience (e.g. he has never seen his grandmother shriek like that before) that there is a strong association in his mind between her shrieks and the presence of mold. So, the next time he hears her scream in a similar fashion, he cannot help but to think that she is in the presence of mold.

14. There are two points to note here. First, as with the original thought experiment, whether or not James is correct to arrive at this conclusion is not important; he is perfectly willing to acknowledge the fallibility of his belief. Second, that the belief itself is based on a probabilistic relationship between shrieks and mold should not disqualify it from being used for categorization. After all, prototype theories made use of features and properties that were probabilistic.

15. At this point, one could object that in such cases, James would not be categorizing mold; instead, he would merely be appealing to these beliefs to identify it. However, it is unclear whether the 'literal' view is committed to the idea that a concept must include all information that might be relevant for identifying its members. If it is not, then it could not be faulted for excluding information pertinent to identification. In response, I would like to point out that identification is (a type of) categorization: to identify x as y is simply to apply the concept Y to x. It is unclear how a person can be said to identify something as mold if she had not already applied the concept MOLD to the entity appearing before her. If this is correct, then the 'literal' view would be committed to the idea that a concept has to include any information relevant for identifying its members, for it is already committed to the idea that a concept has to include information relevant for categorization. As we have already seen, this latter commitment was a primary motivation behind the adoption of the 'theory-theory'.

16. For instance, this construal of the 'literal' view not only would encompass items (1)–(5) (see the discussion on the 'liberal' view's account of conceptual structure in the third paragraph below in the text), but also some 'trivial' information (provided that it is relevant for

categorization purposes).

17. Just how much information conceptual structure will need depends, of course, on the theory itself (whether the 'liberal' or 'literal' interpretation).

18. For instance, Prinz argues that if concepts were construed as 'long-term-memory networks', then they could not be activated by our finite and limited working memory (2002, chaps 4, 6).

19. This is perhaps unsurprising given that the only way to ensure that all such information would be included in the theory is to subscribe to the 'liberal' view.

20. I have hitherto been assuming that the 'theory-theory' is a theory of conceptual structure which holds for all kinds of concepts. Thus, the argument presented here shows, at the very least, that there are certain concepts of which the 'theory-theory' cannot be true, and therefore, that the 'theory-theory' is untenable. Not all psychologists, however, make this assumption about the 'theory-theory'. For instance, Gopnik and Nazzi hold that the 'theory-theory' 'does not seem to be as naturally applicable to other types of knowledge, for example, purely spatial knowledge, syntactic or phonological knowledge, musical knowledge or mathematical knowledge' (2003, 307). Instead, they (along with other cognitive psychologists) maintain that the 'theory-theory' holds at least for natural object concepts, concepts like FISH, GRASS, SKUNK, etc. The important point to note here is that our critique of the 'theory-theory' can still be applied to such concepts.

21. It is worth emphasizing that it is not my requirement that concepts should embody all information that is relevant for categorization. As I have noted, this requirement was set forth by cognitive psychologists when they abandoned similarity-based theories in favor of the 'theory-theory'. My point is that such a requirement would rule out the possibility of a theory that posits a minimal structure (see Note 22).

22. We may, of course, want to devise a criterion other than categorization for what to include under conceptual structure. However, adopting this path will essentially undermine the rationale for subscribing to the 'theory-theory'.

REFERENCES

CAREY, S. 1985. Conceptual change in childhood. Cambridge, Mass.: MIT Press.

FODOR, J. 1994. Concepts: A potboiler. Cognition 50: 95–113.
———. 1998. Concepts: Where cognitive science went wrong. Oxford: Oxford University Press.
———. 2004. Having concepts: A brief refutation of the twentieth century. Mind and Language 19: 29–47.

FODOR, J., M. F. GARRETT, E. C. T. WALKER, and C. H. PARKES. 1981. Against definitions. In Concepts: Core readings, edited by Eric Margolis and Stephen Laurence. Cambridge, Mass.: MIT Press.

GELMAN, S. A., and M. A. KOENIG. 2003. Theory-based categorization in early childhood. In Early category and concept development, edited by David H. Rakson and Lisa M. Oakes. Oxford: Oxford University Press.

GELMAN, S., and H. WELLMAN. 1998. Insides and essences: Early understandings of the non-obvious. In Concepts: Core readings, edited by Eric Margolis and Stephen Laurence. Cambridge, Mass.: MIT Press.

GOODMAN, N. 1972. Seven strictures of similarity. In Problems and projects. Indianapolis, Ind.: Bobbs-Merrill.

GOPNIK, A., and T. NAZZI. 2003. Words, kinds, and causal powers: A theory theory perspective on early naming and categorization. In Early category and concept development, edited by David H. Rakson and Lisa M. Oakes. Oxford: Oxford University Press.

GOPNIK, A., and H. WELLMAN. 1994. The theory theory. In Mapping the mind: Domain specificity in cognition and culture, edited by Susan A. Gelman and Lawrence A. Hirschfeld. Cambridge:

Cambridge University Press.

HARMAN, G. 1993. Meaning holism defended. In Holism: A consumer update, edited by Jerry Fodor and Ernest Lepore. Amsterdam: Rodopi.

KEIL, F. C. 1989. Concepts, kinds, and cognitive development. Cambridge, Mass.: MIT Press.

KOMATSU, L. 1992. Recent views of conceptual structure. Psychological Bulletin 112: 500–26.

LAURENCE, S., and E. MARGOLIS. 1999. Concepts and cognitive science. In Concepts: Core readings, edited by Eric Margolis and Stephen Laurence. Cambridge, Mass.: MIT Press.

MEDIN, D. L. 1989. Concepts and conceptual structure. American Psychologist 44: 1469–81.

MEDIN, D. L., E. B. LYNCH, and K. O. SOLOMON. 2000. Are there kinds of concepts? Annual Review of Psychology 51: 121–47.

MEDIN, D. L., and A. ORTONY. 1989. Psychological essentialism. In Similarity and analogical reasoning, edited by S. Vosniadou and A. Ortony. Cambridge, Mass.: Cambridge University Press.

MURPHY, G. 2002. The big book of concepts. Cambridge, Mass.: MIT Press.

MURPHY, G., and D. L. MEDIN. 1985. The role of theories in conceptual coherence. In Concepts: Core readings, edited by Eric Margolis and Stephen Laurence. Cambridge, Mass.: MIT Press.

NICHOLS, S., and S. P. STICH. 2003. Mindreading. Oxford: Oxford University Press.

PRINZ, J. 2002. Furnishing the mind. Cambridge, Mass.: MIT Press.

RIPS, L. J. 1989. Similarity, typicality, and categorization. In Similarity and analogical reasoning, edited by S. Vosniadou and A. Ortony. Cambridge, Mass.: Cambridge University Press.

ROSCH, E., and C. MERVIS. 1975. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology 7: 573–605.

SCHIFFER, S. 1987. Remnants of meaning. Cambridge, Mass.: MIT Press.

SMITH, E., and D. MEDIN. 1981. Categories and concepts. Cambridge, Mass.: Harvard University Press.

SOLOMON, K. O., D. L. MEDIN, and E. LYNCH. 1999. Concepts do more than categorize. Trends in Cognitive Science 3: 99–104.

WITTGENSTEIN, L. 1953. Philosophical investigations. Oxford: Blackwell Publishers.