Regaining Composure: A Defense of Prototype Compositionality

Jesse J. Prinz
jesse@subcortex.com

(Draft version for Werning, M., Hinzen, W., & Machery, E. (Eds.).
The Oxford Handbook of Compositionality. Oxford University Press. 2008)

Beginning in the late 1960s, psychologists began to challenge the view the definitional theory of concepts. According to that theory a concept is a mental representation comprising representations of properties (or "features") that are individually necessary and jointly sufficient for membership in a category. In place of the definitional view, psychologists initially put forward the prototype theory of concept, according to which concepts comprise representations of features that are typical, salient, and diagnostic for category membership, but not necessarily necessary. The prototype theory gained considerable support in the 1970s, but came under attack in the 1980s. One objection, most forcefully advanced by Jerry Fodor, is that prototypes do not combine compositionally. Compositionality is said to be an adequacy condition on a theory of concepts. If prototypes don't compose, then prototypes are not concepts. Or so the argument goes.

In this chapter, I will argue that prototypes are sufficiently compositional to overcome the objection. I will not, however, advance the claim that prototypes *are* concepts. Rather, I will say they are very important components of concepts, components that play a privileged role in our mental lives. An adequate theory of how prototypes combine is, therefore, an important part of any adequate theory of thought. I will sketch such a theory, drawing on proposals that I develop in Prinz (2002: chapter 11). In the final section, I will critically evaluate a line of experimental evidence designed to challenge theories of this kind.

1. What Are Prototypes?

Prototype theory emerged out of two main sources. First, research on perceptual category learning suggested that people spontaneously abstract representations of the statistical central tendency when they are exposed to a range of similar images. The abstracted representation corresponds to the average or prototype for a range of training images and can be used to classify future examples. Examples are recognized faster if they are similar to average, even if an average instance has never actually been seen (Posner and Keele, 1968). The second source was philosophical. Wittgenstein (1953) had gained notoriety for railing against the standard approach to philosophical analysis. He rejected the view that ordinary concepts (those expressed by words in ordinary language) have underlying definitions—an assumption that had been central to philosophical practice since Plato. If the definitional theory were right, the entities in the extension of a concept should share a unifying essence. Wittgenstein tried to show that this is not the case. Concepts often refer to sets of things that are unified by family resemblance, not essence; any two items in the set will share some features in common, but the features shared by one pair will not necessarily be the features shared by another. This idea inspired Elanor Rosch and Carolyn Mervis to seek out empirical support for

Wittgenstein's conjecture. Over the following years, they found substantial evidence (Rosch and Mervis, 1975; Rosch, 1978; see also Hampton, 1979; Smith and Medin, 1981). When people list features corresponding to their concepts, the items they come up with are not necessary, but merely typical.

Rosch and others established that typical features are actively used categorization. Category instances that have more of the typical features are rated as better instances than less typical ones (Mervis, Catlin, Rosch, 1976). These typical members are produced faster during category production tasks (Smith, Shoben, and Rips, 1974; Rosch, 1978), and they are learned earlier in development (Rosch, 1973). The categories that are most salient to us are the ones whose instances share many typical features in common and differ in typical features from categories at the same level of analysis (Rosch, Mervis, Gray, Johnson, Boyes-Braehm, 1976). For example, we are more likely to identify something as a DOG than as a POODLE or an ANIMAL, even if it falls under all three categories.

The term "prototype" was introduced to explain results like these. For Rosch (1978), the term refers to the class of behavioral effects, not to an underlying mental structure, but most other theorists have assumed that prototypes are mental representations. On most theories, they are conceptualized as collections of representations corresponding to typical category features. So a bird prototype might include components representing a beak, wings, feathers, flight, and two taloned legs. These features are highly typical (most birds have them), highly salient (they can be seen), and highly diagnostic (something that has one or more of these features is likely to be a bird). But they are not necessary: one could pluck a bird, clip it's beak, and sever it's legs and wings without transforming it into something other than a bird. For many categories, the prototype will include features that are not only contingent, but also far from universal in the category: apples are prototypically red, chairs prototypically have four legs, and dogs prototypically have ears that hang down.

Beyond this general characterization, there are different more specific theories of how prototypes are represented. On some versions, the prototypical features are organized into structured lists, which divide into such subheadings as physical attributes, means of locomotion, and perhaps diet. In a connectionist framework, a prototype might be a collection of weighted connectionists between feature-representing nodes, or, more graphically, points in a multidimensional space, whose dimensions correspond to nodes in the network. On an empiricist approach, prototypical features might be interpreted as components of structured mental images, and imagistic simulations of prototypical activities. A bird protoype might be an image or a bird together with images of how birds move and how they eat. For what follows, the exact format of these representations need not concern us.

2. Prototypes And Concepts

Rosch and others found overwhelming evidence that prototypes are used in categorization and other cognitive tasks. They also found evidence that prototypes are pervasive. Almost every public language expression shows prototype effects, suggesting that words are grasped by means of prototypes. This pattern of findings led naturally to the conjecture that concepts are constituted by prototypes. A concept is a mental representation of a category. Concepts are postulated to explain categorization and comprehension of language. Concepts are also presumed to be the building blocks of thoughts. They are the primary representational tools used in cognition. The discovery that prototypes are pervasively used in cognitive tasks can be taken as direct evidence for the view that concepts are prototypes. By the early 1980s, this was the new orthodoxy in psychology.

But doubts began to emerge as well. Some of these doubts had to do with the fact that psychologists were discovering evidence for some other kinds of mental representations that also seemed to play important roles in cognition. Two of these were particularly well demonstrated. First, there is good evidence that people store mental records of previously experienced category exemplars (Brooks, 1978; Medin and Schaffer, 1978; Nosofsky, 1986). Prototypes are representations of average category instances that are acquired by abstracting over encounters with specific objects. But the specific objects are also internally represented and stored, and these records play a role in categorization. For example, if you encounter an unusual chair, you might store an image of it, and use that image to recognize similarly unusual chairs in the future.

Second, evidence accumulated for the view that people construct theories corresponding to the categories they are familiar with (Carey, 1985; Murphy and Medin, 1985; Keil, 1989; Rips, 1989). A theory can be understood as a set of principles corresponding to causal or explanatory relations between observed features, including the postulation of hidden mechanisms that cannot be readily observed. A theory of birds might tell us that wings are used for flying and that digestion is achieved via organs that are hidden from view. Theory theorists showed that we sometimes categorize something on theoretical grounds, even if it does not resemble a category's prototype.

By the mid 1980s, it seemed that theoreticians had a difficult choice to make: they had to decide whether concepts are prototypes, sets of exemplars, or theories. Simply equating concepts with prototypes no longer seemed tenable because there was good evidence for these other kinds of psychological structures. But the assumption that a choice needed to be made was based on a mistake. In reality, there is no need to choose. Each of these psychological entities may contribute to a theory of concepts.

One framework for integration is suggested by Barsalou (1987). He argues that concepts are temporary constructions in working memory, rather than fixed and enduring entities in long-term memory. What we have in long-term memory is a sizable body of knowledge associated with each familiar category, and only small subsets of that knowledge matter for any given task. On any given occasion, we generate an active representation that contains features relevant to task performance. The body of knowledge associated with a category contains prototypes, exemplar representations, and theoretical beliefs. Each of these can contribute depending on context. We might call the stored information conceptual knowledge and the temporary constructions concepts.

Using this terminology, we can see that the theorist need on decide what kinds of mental entities our concepts are. On some occasions concepts may be exemplars, on others, on others, they may include theoretical features, and on others they may be prototypes. Exemplars may be valuable when faced with tasks that require unusual instances of a category, such as "exotic fruit" or "dangerous pets" or "beach shoes." Theoretical features may be most valuable in situations where one must reflect on an unusual application of a category. For example, "a fruit that resist insect attacks" or "a pet that can help with house chores" or "shoes to where when escaping a burning building." Notice that these examples all involve concept combination. That is, when one concept is combined with others a context is created that may influence which aspects of conceptual knowledge we tap into.

If different kinds of representations can contribute to the construction of concepts, then the question about prototypes is not *whether* they are concepts but *when* are they concepts. I think the most plausible answer is that prototypes are our default conceptual representations (see discussion of "default proxytypes" in Prinz, 2002). If we are presented with either no conceptual information, or a typical context, or a context that does not bring to mind any unusual constraints, we will represent a category using prototypical features. The idea is this. Since a prototype corresponds to a typical category instance, it should be the default representation unless we have reason to think things are not typical, and since a prototype contains features that are salient, it should be the default unless context requires us to reflect on features that are not immediately apparent. Indeed the proposal follows almost directly from what prototypes are. Prototypes do not include all typical features: dogs typically have spleens, but having a spleen in not part of the dog prototype. Rather, prototypes comprise features that are typically noticed when we encounter instances of the category. Thus, prototypical features are features that we typically represent. The features we represent typically on encounters will also be the ones that are most strongly encoded and easily accessed. So prototypes are likely to be the default representations of their corresponding categories. So concepts are prototypes by default, and that suggests that concepts are prototypes most of the time.

3. The Compositionality Objection

As we've just seen, the hypothesis that concepts are prototypes has been challenged by appeal to evidence that other kinds of representations can contribute to conceptual tasks. This led me to conclude that concepts are *usually* prototypes. This qualified conjecture is the most defensible version of prototype theory. But it faces another objection that is sometimes considered fatal. The objection stems from the allegation that prototypes do not combine compositionally.

Compositionality can be defined as a property that a system of representations has if the content of a compound representation is determined as a function of the contents of its component parts. For our purposes, we can operationally define a compound as a representation corresponding to a phrase of English and the parts of the compound can be defined as the representations corresponding to the words that make up a phrase. So a phrase with the form ADJECTIVE NOUN will be a compound with two parts, corresponding to the adjective and the noun. A system of concepts is compositional if the content of phrasal concepts (concepts expressed with phrases) is determined as a function of the content of the component lexical concepts (concepts expressed with single words).

Jerry Fodor has argued vigorously that concepts must be compositional (Fodor, 1981). He has emphasized two motivations for this requirement (Fodor and Pylyshyn, 1988). First, compositionality explains our apparent *productivity*, i.e., ability to think an unbounded number of distinct thoughts despite having a finite conceptual repertoire (this is what Chomsky sometimes calls "creativity" in his work on syntax). You have probably never thought about the category of pink tennis balls silkscreened with portraits of Hungarian clowns, but you have no difficulty grasping what these would be. The concept, PINK TENNIS BALLS SILKSCREENED WITH PORTRAITS OF HUNGARIAN CLOWNS, is perfectly intelligible because we are familiar with its component concepts. We can grasp the compound by combining these. The content of the whole derives from its parts. If the content did not derive from its parts, there would be no explanation of how we understood it. More generally, if compounds were not functions from parts, each compound would have to be learned independently by, for example, being presented with category instances. That would mean we couldn't grasp novel, uninstantiated concepts. It would also mean we couldn't acquire novel thoughts by recombining concepts we already possess. Given the frequency with which we have novel thoughts and the ease with which we grasp novel concepts, it seems overwhelmingly likely that concepts compose.

Fodor's second reason for insistent that concepts compose is that compositionality explains the *systematicity* of thought. If one can entertain a thought of the form aRb, then one can entertain a thought of the form bRa. For example if I can conceive of Obama beating Clinton in an election, I should also be able to conceive of Clinton beating Obama. Compositionality explains this systematicity by saying that that such related formulas can be produced using the same concepts and rules of combination. My concept of electoral victory can be freely combined with my concepts of individuals, allowing me to conceptualize what it would mean for anyone individual to beat any other. It would be bizarre to the point of absurdity to imagine someone who could conceive of one victory without being able to conceive of any other. Some victories may seem more likely or more desirable or more imaginable, but all are conceivable in the sense that we know what it would mean to say, of any person, that she or he won an election. This suggests a compositional system at work.

Fodor and his collaborators argue that prototypes cannot satisfy the compositionality argument. I will focus on the presentation of this objection in Fodor and LePore (1996). Fodor and LePore point out that prototypes of compound concepts are often not derived from the prototypes of their component concepts. A feature that is prototypical for a compound might not be prototypical for the concepts that comprise it. Evidence for such *emergent features* is widespread in the psychological literature (Osherson and Smith, 1981; Murphy, 1988; Medin and Shoben, 1988; Kunda, Miller, and Clair, 1990). For example, people say that PET FISH prototypically LIVE IN BOWLS even though this feature is prototypical for neither PETS not FISH; WOODEN SPOONS are prototypically LARGE, unlike its components; and HARVARD GRADUATED CARPENTERS are judged to be NON-MATERIALISTIC unlike HARVARD GRADUATES or CARPENTERS considered in isolation.

Emergent features come from somewhere other than the prototypes corresponding to the parts of a compound. Thus, the way we acquire the prototype for a compound is not a compositional process. This gives rise to the following argument:

P1. Concepts are compositional
P2. Compound prototypes have emergent features
P3. Prototypes are not compositional (from P2)
C. Therefore, prototypes are not concepts (P1, P3, Leibniz's Law)

On the face of it, this looks like a powerful objection against prototype theory.

4. Compositionality Regained

On closer inspection, however, the foregoing argument is invalid. It turns on a failure to clarify the modality of the compositionality requirement. There are two possibilities to consider. First, consider:

> **Mandatory Compositionality**
> When two concepts are combined, they must *necessarily* combine compositionally.

This seems to be what Fodor and LePore are presupposing. If this were an accurate characterization of the compositionality requirement, the emergent features objection would be successful. However, Mandatory Compositionally can be contrasted with a weaker alternative:

> **Potential Compositionality**
> When two concepts are combined, they must *be capable of* combining compositionally.

The phenomenon of emergent features does not rule out potential compositionality. The fact that compound prototypes sometimes contain features not derived from the prototypes of their components does not show that there is no way to generate a prototype for a compound by a compositional procedure. So the question we need to answer is, must concept necessarily compose or is it enough that they have this potential?

To answer this question, recall that compositionality is postulated to explain productivity and systematicity. Potential Compositionality is all we need to explain these two phenomena. Saying thought is productive means we are *capable of* generating an unbounded number of distinct thoughts. Saying thought is systematic means we are *capable of* forming certain thoughts given that we possess certain others. Notice the modality here. Productivity and systematicity are capabilities. We don't actually generate an unbounded number of thoughts, or entertain every thought that is systematically related to the ones we already possess. But we could in principle. These capabilities require only that we be *capable of* computing novel compounds on the basis of their components; they require only Potential Compositionality. We can be systematic and productive simply by having compositional mechanisms at our disposal *even if we don't generally use those mechanisms or if we regularly supplement them with other methods of combination*.

The existence of emergent features shows only that prototypes are not always combined compositionally. To refute prototype theory, Fodor and LePore would have to demonstrate that prototypes *cannot* be combined compositionally. Not only do they fail to do this, but there is every reason to believe that prototypes *can* combine compositionally. It is easy to come up with a method. The simplest possibility is to simply pool features together. For example, a CHIMP DETECTIVE might be a typical looking chimpanzee in a trench coat solving whodunit murder cases by careful deduction from the evidence. In some cases. compounds may be generated by swapping a prototypical feature for a new value. A PINK TENNIS BALL will be represented as pink and not yellow, but this transformation of the tennis ball prototype can be generated using a compositional procedure. In other cases, a compound prototype might be generated by introducing a relation between two prototypes. An STRENGTH PILL might be conceived as a typical pill (say, a capsule) that improves strength as typically identified (say, lifting weights).

These informal proposals have been fleshed out in the form of more formal theories. The literature on prototypes includes a number of compositional models. For example Smith, Osherson, Rips, and Keane (1988) have developed a model in which prototypes are selectively modified in accordance with principles that are consistent with compositionality: prototypes for compounds are computed on the basis of their component concepts. Hampton (1991) has a model in which prototype features are pooled together and weights on those features are systematically adjusted. Wisniewski (1997) has a model in which hybrid prototypes are formed, in some cases, and relations are introduced in others in accordance with reliable rules. These models demonstrate that prototypes can be integrated compositionally.

There is also empirical support of this conclusion. Research has exposed systematic patterns in the ways we integrate prototypes. For example, Smith et al. show that we increase the diagnosticity rating of an attribute dimension in a nominal concept when it is combined with an adjectival concept corresponding to a value along that dimension. Such predictable patterns suggest that compositional mechanisms are at work.

These considerations expose an equivocation in Fodor and LePore's argument. They invoke compositionality in two premises. P1 says concepts are compositional. As we have seen, this is only true if 'compositional' is interpreted as 'Potentially Compositional.' P3 says prototypes are not compositional, and this is supported by the presence of emergent features. But we have seen that feature emergence only demonstrates that prototypes do not combine compositionally *of necessity*. Therefore, the argument should be reconstructed as follows:

P1. Concepts are Potentially Compositional
P2. Compound prototypes have emergent features
P3. Prototypes are not Mandatorily Compositional (from P2)
C. Therefore, prototypes are not concepts (P1, P3, Leibniz's Law)

This argument is invalid because the properties mentioned in its first and third premises are distinct.

The basic idea I have been advancing is that prototypes *can* be combined compositionally, even if they aren't always combined that way. One can put the point by saying that compositionality is a fallback method of combining concepts. If we are presented with a novel compound, we can generate a prototype from the parts if we need to. But if the compound is familiar, such as PET FISH, there is no need to use a compositional procedure. We can simply use our memories of the pet fish we have encountered to generate a prototype for the compound that is independent of the prototypes of its parts. This leads one to propose that we deploy compositional methods of combination only when we have no knowledge of the things that fall in their extensions. When such *extensional knowledge* exists, we get emergent features; when it does not, we combine compositionally. Call this the Extensional Knowledge Proposal.

4. Is The Extensional Knowledge Proposal Irrational?

Fodor and LePore consider the Extensional Knowledge Proposal and reject it. They think such a method of combining concepts would be irrational. Here is their argument. They begin with the Extensional Knowledge Proposal in order to prove that it leads to an absurd conclusion:

> P1. We combine prototypes compositionally only when we lack extensional knowledge (Proposal)

Fodor and LePore notice that extension knowledge tends to diminish with complexity. Consider the concept BROWN COWS OWNED BY PEOPLE WHOSE NAMES BEGIN WITH 'W'. This is a very complex concept, and, because it combines so many elements, it designates a small and obscure category, one with which we are unlikely to have firsthand experience. Thus,

> P2. The more complex a compound is, the *less* you are likely to have extensional knowledge of it.

But complexity *also* tends to reduce prototypicality. For example, pet fish who live in Armenia and have recently swallowed their owners are unlikely to be prototypical pets. Thus,

> P3. The more complex a compound is, the *less* likely we are to be able to predict its prototypical features on the basis of its components.

P1 and P2 entail:

> P4. The more complex a compound is, the more likely it is to be compositionally combined

But when combined with P3, this leads to the following unhappy conclusion:

> C. The more likely we are to combine prototypes compositionally, the less likely we are to be able to predict its prototypical features on the basis of its components

This conclusion is taken to demonstrate that the assumption on which it is predicated is an irrational method of combining concepts.

The problem with this argument is exposed when we notice that there is a tension between the two examples Fodor and LePore invoke to support its major premises. Both BROWN COWS OWNED BY PEOPLE WHOSE NAMES BEGIN WITH 'W' and PET FISH WHO LIVE

IN ARMENIA AND HAVE RECENTLY SWALLOWED THEIR OWNERS are both complex compounds for which we lack extensional knowledge. But there is a difference. The brown cow example lacks emergent features, and the killer pet fish example has them in abundance; killer pet fish are presumably gigantic, viscous, and voracious. This suggests that complexity coupled with ignorance of extension does not always lead to compositional combination. The proposal set out in P1 is a straw man, because it does not completely specify the conditions under which we resort to compositionality.

To see what's missing, we must determine why features emerge in the killer pet fish case. The answer seems to be that we perceive a conflict between its components. A typical pet fish could not possibly swallow its owner. The recognition and resolution of this conflict depends, not on familiarity with killer pet fish, but on basic background knowledge. We reason that a pet fish could only have swallowed its owner if it were gigantic. In the brown cow case, features don't emerge because there are no perceived conflicts between components. Putting this in more general terms, features don't emerge in this case because we lack relevant background knowledge. This suggests a revision in the proposal Fodor and LePore criticize:

> P1.' We combine prototypes compositionally only when we lack extensional knowledge *and relevant background knowledge* (Proposal)

This amendment undermines their argument. Arguably, the more complex a concept is, the more likely we are to have relevant background knowledge. Therefore, complexity tends to promote emergent features. This leads to the right prediction. We are less likely to use compositional mechanisms in cases where those mechanisms are less likely to predict prototypical features.

Fodor and LePore fail to demonstrate that using compositional mechanisms as a backup strategy is irrational. In fact, this policy is paradigmatically rational. When we construct a prototype to represent a compound concept, we often possess relevant exemplar memories and background knowledge that allow us to infer that things falling under it's two component concepts have important properties not shared by things falling under just one of those concepts. When this information is available, we should use it. For example, if we know that red plants are poisonous, we should incorporate this feature into our RED PLANT concept even if it is not possessed by plants or red things in general. Failing to incorporate this knowledge would needlessly place us in harm's way. This reasoning predicts *a priori* what is evidenced empirically: purely compositional combination will only be used when no relevant memories or knowledge is available.

5. The RCA Model

I have mentioned three things that can contribute to the composition process: compositional mechanisms, memories of exemplars, and background knowledge. It worth saying something a bit more specific about how these are coordinated. Elsewhere I propose the following three-stage model of concept combination (Prinz, 2002). When we are given two concepts to combine, we first search memory for relevant knowledge. In some cases, we will have stored concepts corresponding to the compound (these often correspond to lexicalized phrases, e.g., DOG HOUSE, GRAY MATTER, RUSH HOUR). We can also look for stored exemplar representations that can be cross-listed under the two target concepts. WOODEN SPOON and PET FISH might fall under this category. If we find cross-listed exemplars, we can use them to create a prototype for the compound on the fly. I call this the *retrieval stage*.

If the retrieval stage bears no fruit, we move on to a *composition stage*. This is when compositional combination rules kick in to compute a compound prototype. As suggested above, prototypes can be compositionally integrated in a number of ways. The strategy chosen may be dictated by the concepts in question. For example, if one concept refers to a feature of a kind that the other concept contains in it's prototype then a prototype of the feature represented by the first concept replaces to feature in the second prototype. In PINK TENNIS ball, the usual yellow color is replaced by a prototypical pink. If both concepts represent objects and the objects in question are similar in form, then we may simply pool features together. For example, a BEER-BARREL END-TABLE may look like a typical beer barrel and serve the function of a typical end table. In cases, where there are two object concepts that are too dissimilar in form to integrate, a relation may be introduced between them (Wisniewski, 1997). For example a SNEAKER WASHING-MACHINE may be conceptualized as a washing machine for shoes, rather than a washing machine that one where's on one's feet. Wisniewski hypothesizes that different combination strategies may be applied in parallel and generate various competing interpretations. The strategy that yields results fastest or seems less odd may win.

The composition stage is then followed up by an *analysis stage*, in which background information is used to fill gaps, explain relations, and resolve conflicts between the new collection of features. For example, we resolve a perceived conflict between HARVARD GRADUATE and CARPENTER by introducing NON-MATERIALISTIC, and we can resolve the conflict between PET FISH and SWALLOWED OWNER by introducing GIGANTIC. This stage is non-compositional and requires reasoning. In some cases, an emergent feature is almost compulsory because certain solutions to conflicting concepts are particularly obvious. For example, the compound PAPER RAINCOAT must be modified in some way to avoid the fact that ordinary paper cannot function as a raincoat, because it dissolves in water. Features such as LAMINATED or WAXED are likely to emerge, because they are salient examples of water resistant paper.

In sum, the model has three stages: Retrieval, Composition, and Analysis. I call it the RCA model. It is not intended as a competitor for other accounts in the prototype literature. Rather, it is a way of capturing what any model that allows for emergent features should include. The main thing to notice in the present context is that, if relevant memories and background knowledge are unavailable, this model predicts that we will fall back on purely compositional combination. Since such information generally *is* available, this will happen only rarely. But to explain productivity and systematicity, the mere possibility of compositional combination is all we need.

Before closing this section, it is worth nothing that all stages of the RCA model can be said to draw on conceptual knowledge, according to the approach to concepts that I endorsed earlier. If concepts were always to be identified prototypes, then the retrieval stage and the analysis stage might be described as drawing on non-conceptual knowledge. But I suggested that all knowledge we have of a category, including stored exemplars and theoretical beliefs, count as conceptual. Thus, there is a further sense in which the present model is compositional. Emergent features are not typically features that are drawn from cognitive resources that rely outside our knowledge of the categories in question. They are just drawn from components of conceptual knowledge that we don't use by default. In other words, the model concedes that prototypes do not always combine compositionally (a compound prototype often has features not contained in the prototypes of its parts) but there is a sense in which it generally preserves the idea that concept combine compositionally: the content of a compound is generally derived from the conceptual knowledge associated with words corresponding to each component concept. It is possible that, under some circumstances, we transcend those bodies of conceptual knowledge during the analysis stage. The point is simply that, on the account of concepts endorsed above, many emergent features derive from resources that can be characterized as conceptual. This is a further sense in which we should not see the phenomenon of emergence as a major threat to compositionality. Let me now put this point to one side and consider one final objection from Fodor and his colleagues.

6. Is the Fallback Proposal Empirically False?

Early I considered one objection to the view suggestion that prototypes combine compositionally as a fallback strategy: the objection that such a strategy would be irrational. I argued that this objection is mistaken; it is paradigmatically rational to depart from compositional procedures when we have relevant background knowledge or relevant exemplar knowledge. In this final section, I want to consider another objection: Connolly, Fodor, Gleitman, and Gleitman (2007) have argued that the view I am defending is empirically false. They have data that they take to show that we do not resort to compositional methods as a fallback plan in cases where we lack relevant knowledge. I will summarize their findings here and explain why they do not pose a threat on the RCA model (for more critical discussion, see Jönsson and Hampton, 2008).

In the study, all subjects were given large lists of sentences, and asked to rate on a scale of 1-10, how likely it is that each sentence is true (0 = "very unlikely" and 10 = "very likely"). The sentence included four different kinds of cases, in random order. Some sentences contained familiar nouns without modifiers and asked about prototypical features (e.g., "Squirrels eat nuts"), some added prototypical modifiers (e.g., "Tree dwelling squirrels eat nuts"), some added non-typical modifies (e.g., "Nicaraguan squirrels eat nuts"), and some added pairs of non-typical modifiers (e.g., "Black Nicaraguan squirrels eat nuts"). These non-typical modifiers were chosen because they are not associated with familiar exemplars and they do not create conflicts that require deployment of background knowledge. They are precisely the kinds of modifiers that should promote a compositional strategy for concept combination if the RCA model is right. Without background knowledge or exemplar knowledge, people should rely on the composition stage, and there should be no emergent features. Connolly et al. (2007) say that the RCA models and others like it make the following prediction: the each sentence time should be judged to be equally likely to be true. If prototypes are used in all cases, then the modifiers should not diminish likely truth. There should be a null effect. But, Connolly at al. did not get null effects. Instead they found that assessments of likely truth diminished significantly for each of the four sentence types just mentioned: unmodified sentences where given the highest ratings of likely truth, followed by sentences with prototypical modifies, then came sentences with atypical modifies, followed by sentences with who atypical modifiers.

Connolly et al. explain the results as follows. They say that when categories become less familiar we should withhold judgment of what their instances are like. If we've never seen a Nicaraguan squirrel, we shouldn't assume it's like a North American squirrel. We should withhold judgment. We should recognize that we really don't have a clue what Nicaraguan squirrels are typically like. Thus, our confidence about their diet should diminish. Implicit in this empirical argument is a further objection to prototype theory, which was already forecast in the argument for irrationality above. The authors underlying conjecture that we don't bother to generate prototypes for unfamiliar compounds, because such prototypes would be of little value. This harks back to an objection that Fodor first advanced years ago. Fodor (1981) argues that, when we encounter complex compounds that refer to unfamiliar categories, we don't generate prototypes at all. He would probably say there is no prototype for the concept BLACK NICARAGUAN SQUIRRELS and this all the more so for the concept BROWN DUCKS THAT LIVE IN BANGKOK AND EAT SPOTTED EELS. This is another way in which prototypes are not strictly compositional, and it cannot be explained on the RCA model.

The Connolly et al. experiment looks like a direct empirical refutation of the RCA model, but closer analysis tells otherwise. I think the study suffers from several individually fatal flaws.

First, the study may merely reflect pragmatic effects. If I ask you to tell me Nicaraguan squirrels, I conversationally imply that they may be different from the squirrels with which you are more familiar. Why else would I ask? If the answer were obvious, the question would be foolish. If I were to ask you whether Baltic whales are mammals, I imply that they might not be ordinary whales—they might not be whales at all. So you should modulate your guesses about them accordingly. In fact, asking any seemingly obvious question tips a listener off that the answer may not be obvious. If I ask about whether sea-dwelling whales are mammals, you might think it's a trick question is some subtle way. This would explain why subjects in the experiment were not as confident about tree dwelling squirrels eating nuts, even though squirrels typically live in trees. The experimental results reveal that pragmatic factors are at work. A better design would explicitly eliminate such effects by telling subjects that they should not assume that the categories described are unusual even if they are found in unfamiliar places.

Second, Connolly et al. use an anachronistic measure of prototype structure when they ask subjects to report on "likely truth." The vast literature on prototypes contains various measures for prototypical structure: feature listing, reaction times, typicality judgments, and so on. "Likely truth" is not among these measure, and there is good reason for this. Prototype features are not necessarily true, they are just typical, and there is no straightforward inference from typicality to likely truth. If it's typical of bees that they sting, and then I ask you about some specific bee, it doesn't follow in any systematic way that it stings. Asked whether it is likely to sting, you may have no reliable way to come up with an answer. The RCA model makes predictions about how we represent a category—what features we include and how we weigh them. It does not make any direct predictions about our beliefs about these categories. To see whether Nicaraguan squirrels are represented using prototypical features, we would need another kind of test.

One option is to use the standard feature listing or typicality tests (controlling for pragmatic implicatures). Another option would be to look for implicit measures. If asked whether Hungarian bees sting, I might say I have no clue, but I encountered a bee in Hungary, I would surely avoid it. Likewise, if told to find squirrels in Nicaragua I might show facilitation effects for recognition of squirrel-typical features, such as fluffy tails and nuts. In sum, Connolly chose a bad measure with no solid track record of testing for prototype structure.

Third, Connolly et al. tried to pick adjectives that would not promote theoretical analysis on the part of their subjects, but they may not have succeeded. Take the squirrel case. One thing we know about speciation is that geography makes a difference. Squirrels in one region often differ biologically from squirrels in another. So, when we hear about Nicaraguan squirrels, there is reason to think they may be a different subspecies. Likewise for color. If a squirrel is black, there is reason to think it's a different subspecies. Subjects may subject these compounds to the analysis stage. They may explicitly reason that squirrels with different morphological features and habitats may have different diets. The majority of Connolly et al.'s example suffer from this problem; the majority are natural kind concepts with adjectives that could be interpreted an indicating membership in separate subspecies. Similar problems confound their artifact concepts. For examples, subjects are asked how likely it is that "Handmade saxophones are made out of brass." They may reason that brass is difficult to craft by hand. In some case, they may also have relevant extensional knowledge. Subjections are asked whether "Commercial refrigerators are used for storing food." If they recall that some commercial refrigerators are used in hospitals, they may judge that this sentence is less likely to be true than the sentence "Refrigerators are used for storing food."

Fourth, the data are actually consistent with the hypotheses that people use prototypes for unfamiliar cases. All the judgments about "likely truth" were well above the midline. Subjects were not given the option to say "I don't have a clue, so I won't guess." They seem to have no trouble guessing, and the always think the prototypical feature is preserved in unfamiliar compounds. Their certainty goes down a bit, but this is unsurprising, for reasons I have mentioned.

Fifth, reduction in certainty is actually predicted by some models of prototype combination. Hampton's feature pooling model and Smith et al.'s selective modification model both assume that feature weights are adjust systematically when prototypes are combined (see Jönsson and Hampton, 2008). A model could even build in an algorithm for reducing feature weights when atypical adjectives are applied. Or perhaps it's an attentional effect. When an adjective is introduce it draws attention towards one dimension of the category away from others, and this impacts access to the other features. Connolly et al. respond to a similar suggestion, saying it is a departure from the very proposal that prototype theorists are trying to defend, namely the principle the compounds inherit prototypes from their parts. But it is no departure. Models that systematically adjust feature weights still qualify as compositional, because compound prototypes are generated as a function of component prototypes. The main point, as far as the RCA, Hampton, Smith models are concerned, is that features of the compound are inherited.

In sum, the Connolly et al. study is inconclusive at best. It is not well designed to test for prototype structure and it actually lends support to the view that prototype features are inherited. Future studies may provide more decisive evidence (see Jönsson and Hampton, under review; Sabo and Prinz, in progress). Moreover, to revisit an earlier theme, the inheritance models make good practical sense. If I send you to Nicaragua to find squirrels, it's overwhelmingly likely your squirrel prototype will be the primary tool (probably the only tool) you use in your quest to find them. That prototype will serve as a template that can be used to recognize novel cases. Until you've seen the novel cases, the existing prototype can serve as a reliable means for picking out the category. Of course, you can't be sure that Nicaraguan squirrels are typical, and this uncertainty might be expressed in judgments about likely truth, but if the question is, how will you imagine a Nicaraguan squirrel prior to seeing one, the answer is obviously that you will draw on your prototype.

7. Conclusion

I have argued that prototypes are the default representations we use when thinking about categories; they serve as out concepts most of the time. Concept combination sometimes requires us to use other sources of conceptual knowledge, especially theories and exemplars. But, when such knowledge is not relevant, we can use prototypes to generate compound representations compositionally. That's all the compositionality we need to explain the productivity and systematicity of thought.

References

Barsalou, L. W. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, edited by U. Neisser. Cambridge: Cambridge University Press.

Brooks, L. R. (1978). Nonanalytic Concept Formation and Memory for Instances. In *Cognition and Categorization*, edited by E. Rosch and B. B. Lloyd. Hillsdale, NJ: Lawrence Erlbaum Associates.

Carey, S. (1988). Conceptual Differences between Children and Adults. *Mind & Language* 3:167-181.

Connolly, A. C., Fodor, J. A., Gleitman, L. R., and Gleitman, H. (2007). Why Stereotypes Don't Even Make Good Defaults. *Cognition*, 103, 1-22.

Estes, W. K. (1994). *Classification and Cognition*. Oxford: Oxford University Press.

Fodor, J. A. (1981). The Current Status of the Innateness Controversy. In *Representations*. Cambridge, MA: MIT Press.

Fodor, J. A., and Z. Pylyshyn. (1988). Connectionism and Cognitive Architecture: A Critical Analysis. In *Connections and Symbols*, edited by S. Pinker and J. Mehler. Cambridge, MA: MIT Press.

Fodor, J. A., and E. Lepore. (1996). The Red Herring and the Pet Fish: Why Concepts Still Can't be Prototypes. *Cognition* 58:253-270.

Hampton, J. A. (1979). Polymorphous Concepts in Semantic Memory. *Journal of Verbal Learning and Verbal Behavior* 18:441-461.

Hampton, J. A. (1991). The Combination of Prototype Concepts. In *The Psychology of*

Word Meaning, edited by Schwanenflugel. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jönsson, M. L., and Hampton, J. A. (2008). On Prototypes as Defaults (Comment on Connolly, Fodor, Gleitman and Gleitman, 2007). *Cognition*, 106, 913-923.

Jönsson, M. L., and Hampton, J. A. (under review). The Modifier Effect in Within-Category induction: Default inheritance in complex noun phrases. City University, London.

Keil, F. C. (1989a). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Kunda, Z., D. Miller, and T. Claire (1990). Combining Social Concepts: The Role of Causal Reasoning. *Cognitive Science,* 14, 11-46.

Medin, D. L., and M. M Schaffer. (1978). Context Theory of Classification Learning. *Psychological Review* 85:207-238.

Medin, D. L., and E. Shoben. (1988). Context and Structure in Conceptual Combination. *Cognitive Psychology* 20:158-190.

Mervis, C. B., J. Catlin, and E. Rosch. (1976). Relationships among Goodness-of-Example, Category Norms and Word Frequency. *Bulletin of the Psychnomic Society* 7:268-284.

Murphy, G. L. (1988). Comprehending Complex Concepts. *Cognitive Science*, 12, 529-562.

Murphy, G. L., and D. L. Medin. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review* 92:289-316.

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General* 115:39-57.

Osherson, D. N. and E. E. Smith (1981). On the Adequacy of Prototype Theory as a Theory of Concepts. *Cognition,* 9, 35-58.

Posner, M. I., and S. W. Keele. (1968). On the Genesis of Abstract Ideas. *Journal of Experimental Psychology* 77:353-363.

Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.

Rips, L. J. (1989). Similarity, Typicality, and Categorization. In *Similarity and Analogical Reasoning*, edited by V. S. and O. A. Cambridge: Cambridge University Press.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language.* New York: Academic Press.

Rosch, E. (1978). Principles of Categorization. In *Cognition and Categorization*, edited by E. Rosch and B. B. Lloyd. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., and C. Mervis. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology* 7:573-605.

Rosch, E., C. B. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem. (1976). Basic Objects in Natural Categories. *Cognitive Psychology* 22:460-492.

Sabo, W. D., and Prinz, J. J. (in progress). When Are Prototypes Compositional? University of North Carolina, Chapel Hill.

Smith, E. E., and D. L. Medin. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Smith, E. E., D. L. Medin, L. J. Rips, and M. Keane. (1988). Combining Prototypes: A Selective Modification Model. *Cognitive Science* 12:485-527.

Smith, E. E., E. Shoben, and L. Rips (1974). Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions. *Psychological Review*, 81, 214-241.

Wisniewski, E. J. (1997). When Concepts Combine. *Psychonomic Bulletin & Review*. 4:167-183.

Wittgenstein, L. (1953). *Philosophical Investigations*, G. E. M. Anscombe, tans. New York: Macmillan.