Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou and A. Ortony (Eds.), Similarity and analogical reasoning (pp. 179-195). New York: Cambridge University Press.

6

Psychological essentialism

DOUGLAS MEDIN and ANDREW ORTONY

> What is common to them all? – Don't say: "There *must* be something common, or they would not be called 'games' " – but *look and see* whether there is anything common to all. – For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!
>
> Wittgenstein, *Philosophical Investigations*

Wittgenstein's admonition "don't think, but look" has had the important effect of stimulating psychologists to reconsider their common practice of equating concept formation with the learning of simple definitional rules. In the early 1970s, psychologists like Eleanor Rosch (e.g., Rosch, 1973), responding to the difficulty of identifying necessary and sufficient conditions for membership of all kinds of categories, proposed alternative models of category representation based on clusters of correlated features related to the categories only probabilistically. Without denying the importance and impact of this changed view of concepts (reviewed, e.g., by Smith & Medin, 1981), we think that in certain respects the "don't think, but look" advice may have been taken too literally. There are problems with equating concepts with undifferentiated clusters of properties and with abandoning the idea that category membership may depend on intrinsically important, even if relatively inaccessible, features. For example, on the basis of readily accessible properties that can be *seen*, people presumably will not judge whales to be very similar to other mammals. However, if they *think* about the fact that whales are mammals not fish, they will probably acknowledge that with respect to some important, although less accessible property or properties whales *are* similar to other mammals. This observation suggests that restricting oneself to relatively accessible properties may make it difficult to account for the perceived similarity of whales to other mammals. If one

cannot appeal to "hidden" properties, it is difficult to explain the fact that people might recognize such similarities. Thus there might be a price to pay for looking rather than thinking.

The question of how best to conceptualize possible forms of similarity is intimately related to the question of how to conceptualize the nature of the "stuff" to which judgments of similarity are applied. Similarity judgments are always made about presented or represented entities. Because even presented entities are perceived and interpreted in terms of an existing set of concepts, there is a sense in which similarity judgments are always made with respect to *representations* of entities (rather than with respect to the entities themselves). In other words, we shall assume that when people judge two things to be similar those things are (at least temporarily) represented, so that similarity judgments are always made vis-à-vis *representations*. This means that theoretical treatments of representations need to endow them with sufficient richness to allow similarity to perform useful functions. In addition to the perceived similarity that results from attending (only) to highly accessible (typically, so-called perceptual) properties, we need to consider the similarity that results from considering more central (less accessible) conceptual material too, because this deeper aspect of similarity makes an important and sometimes indispensable contribution to cognition. For example, it can account for why people might believe that two things with very different surface properties (e.g., whales and bears) still are instances of the same category and therefore why they might judge them more similar to one another than they would on the basis of surface properties alone.

In this discussion, we consider the implications of the distinction between the more accessible, surface, aspects of representations and the less accessible, deeper, aspects for the nature of similarity and its role in cognition. By surface aspects, we mean the sorts of things people describe when asked to list properties of objects and the sorts of things psychologists have tried to use as the building blocks of concepts. Central to the position that we advocate, which we call *psychological essentialism*, is the idea that these surface features are frequently constrained by, and sometimes generated by, the deeper, more central parts of concepts. Thus there is often a nonarbitrary, constraining relation between deep properties and more superficial ones. We shall argue that, although it can be a powerful heuristic for various cognitive tasks, there are limitations to using similarity with respect only to surface properties and that there are problems with ignoring the relation between surface similarity and deeper properties.

The view we propose is more optimistic about the role of (superficial) similarity in cognition than that of Lance Rips but less optimistic than the view proposed by Edward Smith and Daniel Osherson. Rips believes that categorization does not necessarily depend on similarity, although he admits that sometimes it might. He presents arguments and data to support his claim that there are factors that affect judgments of category membership that do not affect judgments of similarity and that there are factors that affect judgments of similarity that do not affect judgments about category membership. If similarity and categorization can vary independently of one another, then neither can determine the other, and, in particular, similarity can be neither necessary nor sufficient for categorization. Smith and Osherson, on the other hand, suggest that similarity has a role not only in categorization but also in decision making. To explain this role they need an approach to similarity that is flexible enough to vary depending on how, whether, and when features are weighted (Tversky, 1977). The other contributors to Part I are more agnostic on the conceptual structure issue as it relates to similarity. Linda Smith argues for a more constrained view of similarity but allows for developmental changes in the aspects of similarity processing that are consciously available. That is, young children may have access only to global, overall perceptual similarity, and may learn only later to identify the components or features that determine it. Ryszard Michalski includes a role for similarity but argues that similarity is constrained by goals (his chapter is covered in the commentaries to Part II). Lawrence Barsalou agrees that similarity is involved in concepts, but a big chunk of the similarity that emerges is apparently context-dependent.

Why are there such divergent views about a matter that one might think should be quite straightforward? As psychologists, we expect a great deal of the construct of similarity. On the one hand, we sometimes treat it as a stable construct. On this view, robins *really are* like sparrows in some absolute observer-independent sense, and it is this objective similarity that underlies our perception of them as similar. In contrast to this strong form of metaphysical realism, at other times psychologists treat similarity as a highly flexible construct grounded not so much in objective reality as in the degree to which shared predicates are judged to be involved. When viewed in this way, similarity becomes more like a dependent variable than an independent variable. If similarity is to be grounded in shared predicates, it is open to the objection that it is not really grounded at all. Indeed, Goodman (1972) and Watanabe (1969) have offered formal proofs of just this point by showing that when similarity is defined in terms of shared

predicates all pairs of entities are equally similar. In addition, there is no guarantee that different individuals share the same beliefs about what properties objects have, and contextual factors can have dramatic effects upon the accessibility and salience of different predicates. This need to conceptualize similarity sometimes as fixed and sometimes as flexible poses a dilemma. We require a fixed notion to account for intuitions such as that robins are more like sparrows than they are like sunglasses, whereas we need a flexible notion to account for a whole host of empirical results of the kind presented by Barsalou, for example, and E. Smith and Osherson.

Can this dilemma be resolved? We think that the framework we propose can take us at least part of the way. A first step is to define similarity not in terms of logically possible shared predicates but in the more restricted sense of shared *represented* predicates. For example, both tennis balls and shoes share the predicate *not having ears*, but it is unlikely that this predicate is part of our representation of either tennis balls or shoes. By restricting ourselves to represented predicates we can restrict the predicates that contribute to the determination of similarity. Of course, this leaves unanswered the question of what determines which predicates *are* part of our mental representations. To address this question, we need to take a second step: We suggest that perceptual similarity based on representations of what appear to be more accessible surface properties provides an initial conceptual structure that will be integrated with and differentiated into the deeper conceptual knowledge that is acquired later. Thus properties associated with a concept are linked both within and between levels to produce coherence. The reason that *not having ears* is not a predicate in our mental representation of tennis balls is that it would be an uninformative, isolated fact, unrelated to the rest of our knowledge about tennis balls. Our basic claim is that the link between surface and deep properties serves two functions: It enables surface similarity to serve as a good heuristic for where to look for deeper properties, *and* it functions as a constraint on the predicates that compose our mental representations. Although this constraint need not necessarily be a tight one, it may be enough to allow us to have a notion of similarity that is flexible without being vacuous.

This all suggests that the general way out of the dilemma is to acknowledge, as has been acknowledged in other areas of psychology, that logical and psychological accounts of certain phenomena need not necessarily be compatible. It is now generally accepted that psychologically plausible accounts of certain phenomena are at odds with purely logical analyses. People are not wetware instantiations of formal

systems, be they logical or statistical, as a wealth of research on judgment under uncertainty has shown (e.g., Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). In the case of similarity, what is needed is a richer account that does more than simply view similarity in terms of lists of matching properties or shared predicates. With this general solution in mind, we shall try to sketch a psychologically plausible view of conceptual structure and then relate it to the chapters that are the subject of our commentary.

What is psychological essentialism?

We consider psychological essentialism to be a psychologically plausible analog of the logically implausible doctrine of metaphysical essentialism. The philosophical problem about essences is this: If one wants to argue that an object has an essence by virtue of which it is that object and not some other object, one has to face the problem, clearly recognized by Aristotle, that what that thing is is not independent of how it is described. The same object might be correctly described as, for example, a piece of rock, a paperweight, or an ashtray. But this means that the same piece of rock under these three different descriptions must have three different essences. In general, the problem is that an object appears to need as many essences as there are possible true descriptions of it.[1] But such a proliferation of essences undermines the very notion of an essence as some unique hidden property of an object by virtue of which it is the object that it is.

Because of observations such as these, the idea that objects might have some (possibly unknown) internal essence that makes them the objects they are is a philosophical orphan, banished to the netherworld of Platonic forms. Justifiable as this exile may be from a philosophical point of view, it is possible, and we think useful, to postulate something that might be called *psychological essentialism*. This would be not the view that *things* have essences, but rather the view that people's *representations* of things might reflect such a belief (erroneous as it may be). Since a major task for cognitive psychology is to characterize knowledge representations, psychological theories about them have to be descriptions of psychological reality, not of metaphysical reality. Thus, if people believe that things have essences, we had better not ignore this fact in our theories of knowledge representation.

We think there is evidence that ordinarily people *do* believe that things have essences. Many people behave as though they believed it, presumably because the assumptions that things have essences is an

effective way of viewing the world and making predictions about it. One reason for supposing that people's concepts often embody an implicit belief that things have essences is provided by the third and fourth experiments described by Rips, in which subjects were unwilling to change the way in which they classified objects even though transformations of certain superficial properties of those objects rendered them more similar to exemplars of some other category. In these experiments subjects were behaving as though they believed that category membership depended upon the possession of some "hidden" (Rips calls them "extrinsic") properties of which observable properties are but typical signs. There is another reason that leads us to believe that people typically endorse, at least implicitly, some sort of essentialism. The nature of a great deal of scientific inquiry appears to be focused on trying to get at the "underlying reality" of phenomena rather than merely describing their observable properties. For example, the idea that things have essences was a guiding principle in the development of modern taxonomy by Linnaeus.

We should emphasize again that we are not claiming that objects have essences or that people necessarily believe that they know what these essences are. The point about psychological essentialism is not that it postulates metaphysical essentialism but rather that it postulates that human cognition may be affected by the fact that people believe in it. In other words, we are claiming only that people find it natural to assume, or act as though, concepts have essences.

Psychological essentialism should not be equated with the classical view that concepts are representations of classes of objects that have singly necessary and jointly sufficient conditions for membership. First of all, on our account people may sometimes believe that necessary and sufficient conditions are a *consequence* of the essential nature of the thing in question, rather than that essential nature itself. Furthermore, the essential nature may not generate necessary and sufficient properties at all. For example, it may be part of the represented essence of *bird* that birds fly, even if it happens that not all birds do fly and that people know this. More generally, we propose that the knowledge representations people have for concepts may contain what might be called an *essence placeholder*. There are several possibilities for what is in such a placeholder. In some cases, but by no means in all, it might be filled with beliefs about what properties are necessary and sufficient for the thing to be what it is. In other cases it might be filled with a more complex, and possibly more inchoate, "theory" of what makes the thing the thing that it is (see Murphy & Medin, 1985). It might, additionally, contain the belief (or a repre-

sentation of the belief) that there are people, experts, who really know what makes the thing the thing that it is, or scholars who are trying to figure out exactly what it is. Just as with theories, what the placeholder contains may change, but the placeholder remains.

Another reason for not equating psychological essentialism with the classical view is based on one particular reading of, or defense of, the classical view. This reading turns on the classical view's distinction between the *core* of a concept, which brings out its relationship to other concepts, and an associated *identification procedure* for identifying instances of the concept (see Smith & Medin, 1981). For example, the core of a concept like *boy* might contain properties such as *male, young,* and *human* that could be used to understand its relation to other concepts like *girl, colt,* and *man*. The identification procedure might consist of processes employing available information about currently accessible properties like hair length, height, characteristic gait, and so on, that can be used to help determine that some person is likely to be a boy rather than a girl or a man. One defense of the classical view is that the typicality effects used to attack it are based on properties involved in the identification procedure rather than on core properties. Insofar as this defense of the classical view can be upheld, however, one might object that it presupposes too great a dissociation between the core properties and the others. Our view is that the more central properties are best thought of as constraining or even generating the properties that might turn out to be useful in identification (see Smith, Medin, & Rips, 1984, for related arguments). Furthermore, rather than seeing a sharp dichotomy between core properties and properties that constitute the basis for an identification procedure, we conceive of properties as lying on a continuum of centrality ranging from relatively inaccessible, deep properties to more accessible, surface ones.

The notion of a continuum of centrality linking deeper and more superficial properties may provide the basis for some structuring of, or coherence in, family resemblance categories. For example, associated with a person's representation of *male* may be the idea that being male is partly a matter of hormones, which directly influences features such as height and facial hair. In the absence of deeper principles to link more superficial properties, categories constructed only in terms of characteristic properties or family resemblances may not be psychologically coherent. In some recent experiments using artificially constructed category materials, we have clear evidence that providing deeper linkages is sufficient to enable people to find family resemblance categories to be natural or coherent, and suggestive evi-

dence that these linkages may be *necessary* for coherence (Medin, Wattenmaker, & Hampson, 1987). One way to summarize our argument is to say that twins are not twins *because* they are similar; they *are* similar because they are twins. So the second key element in our psychological essentialism is that our mental representations reflect the notion that properties differ in their depth and that deep properties are often intimately linked to the more superficial properties that so often drive our perceptions of and intuitions about similarity. The linkages between surface and deep properties are a function of the theories we have about the deep ones.

So far we have made two main points. First, people act as if their concepts contain essence placeholders that are filled with "theories" about what the corresponding entities are. Second, these theories often provide or embody causal linkages to more superficial properties. Our third tenet is that organisms have evolved in such a way that their perceptual (and conceptual) systems are sensitive to just those kinds of similarity that lead them toward the deeper and more central properties. Thus whales, as mammals that look more like fish than like other mammals, are the exception that proves the rule: Appearances are usually not deceiving. This means that it is quite adaptive for an organism to be tuned to readily accessible, surface properties. Such an organism will not be led far astray because many of these surface properties are constrained by deeper properties. If this view is correct, then the types of category constructions based on global similarities described by Linda Smith will tend to be just those partitionings that will be useful later on as the child acquires more knowledge and begins to develop deeper conceptual representations. In other words, psychological similarity is tuned to those superficial properties that are likely to be causally linked to a deeper level. This is particularly likely to be true with respect to natural kinds.

The question we now want to address is whether or not this way of augmenting the structure of concepts can provide a framework within which to understand the chapters in Part I. We shall suggest that what Rips is actually describing are the kinds of factors that go into *identification* of concepts – his challenge to similarity we take as supporting the general view that similarity of a putative category member to representations of exemplars and prototypes is a fallible heuristic for deciding category membership. We interpret his chapter as showing that the limitations to this heuristic are determined by the degree to which the surface features on the basis of which such judgments are made are constrained by less accessible, deeper features, or psycho-

logical essence. Linda Smith's paper provides a developmental perspective according to which infants and young children may have little else to go on than surface features. In the context of the present discussion, this is a profoundly important observation. The very fact that young children seem to classify only in terms of global similarity, rather than by isolating distinct dimensions, may provide them just the stability needed to make it likely that they will construct appropriate and useful categories. This suggests to us that attention to surface features early in development may be an asset rather than a limitation. Edward Smith and Daniel Osherson also present an account of the role of similarity, this time not in judgments of category membership but in decision making and choice. Rather than questioning the theoretical utility of similarity, Smith and Osherson suggest that it may be able to explain more than was previously thought. We suspect, however, that they will ultimately need to supplement their treatment of similarity in terms of representations involving lists of independent features with a view that includes the notion that the predicates in representations are interrelated and may differ in their centrality. On our account of conceptual structure, linkages between deeper features and more superficial features greatly constrain the contexts in which the assumption of independent features will work.

Whereas the Rips chapter and the Smith and Osherson chapter, intentionally or otherwise, are both concerned with the role that similarity plays in some fairly important *processes*, Lawrence Barsalou focuses more on what similarity might tell us about the underlying conceptual representations in terms of which such judgments are presumably made (what we referred to at the beginning of this chapter as the "stuff to which judgments of similarity are applied"). Whether or not you like Barsalou's message about the instability of concepts depends on who you are. We like it because it is consistent with our view that the more central aspects of our concepts are often quite inchoate and not readily accessible. Ryszard Michalski's two-tiered theory of concept representation is somewhat similar both to Barsalou's distinction between context-dependent and context-independent properties and to the general framework we have been developing. Michalski places greater emphasis on cognitive economy (efficient representation) than we do, and he focused on goal-driven rather than theory-driven representation. We think a continuous gradation of depth is more natural than a dichotomy, but we agree with Michalski that categorization may be more like an inference than a similarity computation.

Commentary

Rips's main point seems to be that in many cases properties that we would consider closely linked to the psychological essence (Rips calls them "extrinsic" features) constitute the criteria for category membership, not superficial features. If, as he suggests, the resemblance approach to categorization is limited to the use of superficial properties, then Rips's observations are quite damaging. In principle, however, we see no reason why the resemblance approach should be constrained in this way. It would seem that category members could be judged to be similar *with respect to deeper features*. In Rips's experiments subjects were encouraged to focus on surface or perceptual features for similarity judgments and on other properties for typicality or class membership judgments. For example, the first experiment was set up in such a way as to lead subjects to consider only a single physical dimension (e.g., diameter) of an unknown object relative to that dimension of two potential categories (e.g., quarters and pizzas). This does not rule out the possibility that category membership judgments were also based on similarity, but with more than a single dimension involved (and not necessarily all readily accessible ones at that). Consider the following variant on Experiment 1: Subjects are told that they should bring to mind some object, *x*, which is 3 inches in diameter, and they are asked, "Is it more likely to be a quarter or a pizza?" They then respond, just as Rips had them respond, presumably by saying that it is more likely to be a pizza. They are then instructed to keep in mind what they imagined *x* to be, and now they are asked whether that *same* object is more similar to a quarter or a pizza. Our point is that it is necessary to know that subjects made their categorization judgments and their similarity judgments using the same instantiation of the unknown object *x*. Once we know what *x* is, the situation with respect to similarity may be radically different. Of course, one would need to be able to explain why people's images are more pizza like than quarter like, so this argument cuts both ways. We agree with Rips that, unless one can specify how similarity is determined, the resemblance approach to similarity is vacuous.

On the other hand, we disagree with Rips's assertion that some of us are committed to the view that (*a*) similarity *determines* the probability that a person will assign some instance to a category and (*b*) that similarity is *responsible* for prototypicality judgments; it is not clear that they are. The assertion that "The probability of classifying exemplar *i* into category *j* is an increasing function of the similarity of exemplar *i* to stored category *j* exemplars and a decreasing function of the

similarity of exemplar *i* to stored exemplars associated with alternative categories" (Medin & Schaffer, 1978) claims no more than the empirical fact that there is indeed a positive correlation, namely, that the more similar *i* seems to stored category *j* exemplars, the more *likely* is *i* to be categorized as a *j*. It doesn't guarantee it; it just makes it more likely. In other words, to say that similarity can play a role in *identification* is not the same as saying that categories (or their corresponding concepts) are *constituted* on the basis of similarity among exemplars. It may be that one heuristic people use for deciding that an *i* is a *j* is similarity to exemplars. Heuristics are not causes, and the fact that a heuristic often works tells us something about the interface between our (selective) perceptual systems and the world. Similarly, Murphy and Medin (1985) suggest that perhaps concepts are like theories in the sense that, if one has a "theory" of what it takes for an *i* to be a *j*, then a decision about a particular case will be based on how well that case seems to "fit" *j*, which is the criterion associated with the theory of *j*-ness. The process of deciding how well it fits may or may not implicate similarity, but it does not *necessarily* do so.

Nevertheless, when all is said and done, we see an important moral in Rips's chapter, although we would give it much more emphasis than he does. We view his arguments and data as supporting the following claim: The criteria for category membership (whatever they are) are not necessarily always apparent on the surface as physical properties of instances of the category. On this reading of Rips, the criteria involve deeper properties that, to varying degrees, may impose constraints on more accessible properties. Sometimes these constraints are strong, although, of course, the issue does not really have to do with categories themselves but has to do with people's representations of them: concepts. So most lay people do not know what the real criteria are for something to be an airplane (although they presumably *do* believe that aeronautical engineers do!). However, even in the absence of such knowledge they assume that these underlying criteria impose strict constraints on some of the accessible features, such as the possession and size of wings, flaps, fins, and other aviation paraphernalia. In cases where the psychological essence imposes strong constraints, similarity to exemplars is likely to be a good heuristic for deciding category membership. Sometimes the constraints are relatively weak, as with many goal-derived categories. Where the constraints are less strong, similarity is likely to be less successful, although in many cases physical properties may be accidentally or indirectly constrained. So, for example, the physical shape of eggs may be indirectly constrained by whatever it is that makes eggs eggs, because

the shape is so well suited to protecting the embryo before, during, and after its passage into the outside world. People presumably believe that eggs are oval because the nature of eggs imposes certain constraints on their physical properties rather than because that is the best shape for fitting them into egg cups!

Our first reaction to the Smith and Osherson paper is that the idea of applying notions of similarity to decision-making and judgment tasks is a good one, and although Kahneman and Tversky did not exactly ignore similarity, they did not undertake the systematic analysis that Smith and Osherson attempt. We think that their simplifying assumption that the values of different features are independent, although a convenient starting point, might pose something of a problem for the general case. Their approach seems to treat features as not being linked in any particular way. We think that it is rare for the value of a feature on one dimension not to affect the likely values on other dimensions. For example, finding out that Linda is a feminist bank teller might not simply change the diagnosticity and votes on the property *politics*, it might also change one's ideas about the style of clothing that Linda might wear, her preferences for different forms of recreation, or even the kinds of food she might enjoy. To be fair to Smith and Osherson, the notion of independent dimensions, as we have said, is a simplifying assumption. Our somewhat pessimistic attitude is driven by the exception that, in practice, this assumption just may not hold often enough for it to constitute an adequate basis upon which to build a general analysis. We know that correlated features violate the independence assumption. For example, people rate small spoons as more typical spoons than large spoons, but they rate small *wooden* spoons as *less* typical than large wooden spoons. If correlated features tend to be the rule rather than the exception (as our view implies), then, in general, it seems unlikely that the effects of a property having a particular value can be confined to a single dimension (Rumelhart & Ortony, 1977), and if they cannot, then the independent-dimensions approach is not going to suffice.

The observations on the cab problem and people's failure to use base rate information are interesting and clever. On the other hand, whether this approach will work in general is not clear. It seems to imply that judgments will be based on similarity computations even when the form of similarity is clearly irrelevant. To give an extreme example, suppose the witness testified with 99 percent reliability that the cab had four wheels. Number of wheels has some diagnosticity for differentiating cabs from trucks and motorcycles, and so we wonder what would prevent the similarity computation from running off

and leading to a continued failure to use base rate information? (The data seem to go the other way.) Similarity may not always be used in a simple, straightforward way because how one interprets similarity data may depend on one's theory of how that similarity was generated. To use an example based on Einhorn and Hogarth (1985), imagine a set of five eyewitnesses testifying as to the apparent speed of the taxicab. Although credibility might generally be expected to increase with interwitness agreement, suppose in some particular case each of five witnesses testifies that the cab was traveling at exactly 58.2 miles per hour. Does this remarkable similarity increase credibility? It seems to undermine it.

Overall, this incursion of similarity into judgment and decision making is intriguing. Our only reservation is that we think Smith and Osherson are going to need a richer form of knowledge representation and a correspondingly more powerful theory of similarity in order to get it to do what they want it to do.

Linda Smith's analysis reveals that there are nuances associated with modes of similarity processing that undermine the very oversimplified account of development in which young children categorize in terms of superficial similarity whereas older children's categorizations are constrained by deeper aspects of similarity. That is, there appear to be major shifts in how surface similarity is processed at different stages of development. Smith's work appears to conflict with research that shows that young children develop theories that constrain their conceptual behavior (e.g., Carey, 1982; Keil, 1981). Smith's work, however, is with children younger than those used by Carey and by Keil, so that it is not easy to determine the extent of the conflict. There may prove to be a natural integration involving a transition from similarity-based to theory-based categories. Alternatively, it may be that theories are constraining the concepts of even the youngest children and that the similarity-based account of conceptual development is incomplete for all ages.

We agree with Barsalou that there is a great deal of concept instability. However, we think care has to be taken not to equate instability in outputs or behaviors with underlying or internal instability. Might it be that our underlying concepts are in fact stable (whatever that might mean) and that the apparent instability is an artifact of the processes that operate on these stable representations? Given our framework, we would argue that the deeper one goes the more stability one ought to find. So, if we were to ask 200 people on 10 different occasions in 10 different contexts whether dogs are more likely to give birth to puppies than to kittens, we would find remarkable sta-

bility. (Of course, Barsalou does not dispute this.) Michalski would probably argue that we should expect stability only for properties that are part of what he refers to as the base concept representation. One might claim that Barsalou is exploring only the fringes of conceptual use while ignoring the huge quantities of knowledge that are so stable we are scarcely aware of them. We have in mind knowledge of the kind that dogs, but not rocks, can take predicates like *sleep* or *eat* (see Keil, 1981). Still, Barsalou convinces us that there *is some* instability, and we think a view of the kind we have proposed is also committed to this conclusion. Clearly, intraperson instability can arise as a result of context-based priming effects. For instance, in making some judgment about some concept, the context in which the judgment is elicited might prime some of the concept's surface features so that the judgment would be more likely to reflect the causal linkages between the psychological essence and such primed features than between the essence and other surface features.

Barsalou's assault on concept stability leads one to ask *where* stability is and what it means to try to measure it. People do not ordinarily walk around making judgments of typicality or providing dictionary definitions, so it is at least relevant to ask how people normally learn about, update, and use concepts, and what stability or instability is associated with *these* functions. We think that Barsalou is correct in his important observation that concept representations often are constructed on the fly. Only in this way can knowledge be tuned to particular contexts. From the traditional contextless view of concepts, the results on context dependence are depressing for what they tell us about concept stability. However, if one thinks of concept representations as frequently being computed, then contexts serve to fix meaning and provide stability *in that context*. So, if one is talking about a "bird on a Thanksgiving platter" the referent is heavily constrained (Roth & Shoben, 1983). Indeed, much of the stability that one might hope for may depend very much on particular contexts. Consider a concept such as *redneck*.[2] If one is asked to list attributes of this concept the potential list is long, and there might well be considerable instability in just what is listed. As more context is added, the situation becomes more like supplying background information needed to generate specific predictions from a theory. Therefore, if people are asked to judge how likely a redneck is to encourage his 17-year-old daughter to date a 39-year-old man of a different race, one can anticipate considerable agreement both across judges and for the same judge at different times. So the notion of deeper properties is perfectly consistent with a considerable amount of instability and a considerable

amount of context dependence. In the case of the *redneck* example, the key to successful communication lies not so much in the ability to supply the same definition as it is does in the ability to generate the same predictions in specified contexts. Finally, another aspect of context dependence is that different concepts may interact so as to change the importance of some of their constituent properties (e.g., Barsalou, 1982; Ortony, Vondruska, Foss, & Jones, 1985). This all means that Barsalou's results on concept instability can be taken as evidence for flexibility that paradoxically allows for both *accuracy* and (a fair amount of) stability in particular contexts.

Conclusion

We have proposed that there is often a nonarbitrary relationship between the represented deep properties and the represented surface properties of concepts. This relationship can vary from a strong causal one to a weaker constraining one, but in either case the use of similarity with respect only to relatively accessible surface properties, because they *are* constrained, can serve as a powerful, although fallible, heuristic for various cognitive tasks. It is in this sense that we think Rips perhaps attributes too little power to surface similarity. Knowledge representations have to be construed as having sufficient richness to allow similarity to perform useful functions. Our reservations about the Smith and Osherson chapter hinge on just this point. The model they propose seems to work quite well if one accepts their simplifying assumptions that schemata are unstructured property lists. But if, as they must be, schemata are complex, multivariate representations involving many represented or inferable interdependencies, it is not at all clear that their relatively straightforward analysis can survive.

We started this chapter by suggesting that Wittgenstein's admonition to look rather than to think may have led psychologists to focus too much on superficial similarity with respect to concept representations incorporating only superficial, accessible properties. There are really two issues, the second of which presupposes the first. The first is that the question is a question about *representations*, not about the things that are represented. As such, the issue is not whether birds possess some feature or features by virtue of which they are birds but whether people's *representations* of birds include some such component, explicit or otherwise. Once this point is established, the second point is that we have to examine how the linkages between superficial and deeper properties serve to provide structure to concept representations. We have suggested that if one does this one may discover

that there is a meaningful and useful role that can be played by psychological essentialism.

NOTES

Douglas Medin was supported in part by grants from the National Science Foundation, NSF84–19756, and the National Library of Medicine, LM04375. We wish to thank Robert McCauley, Brian Ross, and Ed Shoben for their helpful comments.

1 Locke argued that a distinction between real and nominal essence is needed. The nominal essence for Locke was the abstract idea that constitutes the basis of classification. The real essence he explained as that which some suppose is "the real internal, but generally (in substances) unknown constitution of things, whereon their discoverable qualities depend." Thus both Aristotle and Locke, while accepting some form of essentialism, were forced to deny that the essence of a thing lay in that thing independently of the way it is classified. For Aristotle, the essence of something was bound to the way it was described or conceptualized, and for Locke it had to be bound to a corresponding abstract idea, at least in part because he believed that concepts corresponding to nonexistent entities like unicorns and mermaids had perfectly good essences.
2 We use the term *redneck* to make a purely scientific observation and share Barsalou's opinion regarding the inaccuracy and the prejudice fostered by the use of stereotypes.

REFERENCES

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10*, 82–93.

Carey, S. (1982). Semantic development: The state of the art. In E. Wanner and L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 347–389). Cambridge: Cambridge University Press.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92*, 433–461.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects*. New York: Bobbs-Merrill.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Keil, F. C. (1981). Constraints on knowledge and cognitive development. *Psychological Review, 88*, 197–227.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. (1987). Family resemblance, concept cohesiveness, and category construction. *Cognitive Psychology, 19*, 242–279.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliff, NJ: Prentice-Hall.

Ortony, A., Vondruska, R. J., Foss, M. A., & Jones, L. E. (1985). Salience, similes, and the asymmetry of similarity. *Journal of Memory and Language, 24*, 569–594.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15*, 346–378.

Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99–135). Hillsdale, NJ: Erlbaum.

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

Smith, E. E., Medin, D. L, & Rips, L. J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and stereotypes." *Cognition, 17*, 265–274.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Watanabe, S. (1969). *Knowing and guessing: A formal and quantitative study.* New York: Wiley.