many proponents of the theory approach, this seems like a natural move. But it is not the one Machery advocates. In fact, Machery cautions that "philosophical accounts of scientific explanation would probably be useless for spelling out the psychological notion of theory" (2009, p. 102).

I want to suggest that Machery is wrong to dismiss the psychological value of theories of explanation from the philosophy of science and to neglect recent advances that move the psychology of explanation beyond "folk understanding." Scientists are, after all, psychological creatures, and there is every reason to expect the aspects of human cognition that shape everyday explanations to play a role in science. Scientists and everyday cognizers also face similar problems and have similar goals: They confront limited data, and from this they must construct a representation of the world that supports relevant predictions and interventions.

But there is another reason to expect a close correspondence between philosophical and psychological accounts of explanation, one that stems from the philosophical methods typically employed. Here, in uncharitable caricature, is how theory development often proceeds: Philosopher P_1 proposes theory T_1 of explanation; philosopher P_2 quickly generates putative counterexample C, a specific case in which T_1 makes one prediction about what is explanatory, but philosopher P_2's intuition demurs. The philosophical community pronounces one way or the other, based largely on shared intuitions about C, so T_1 stands (for now) or gives way to a new theory. This is not the most efficient way to collect data, and it would not pass muster for an experimental psychologist; but to the extent philosophers are like everyday folk, one would expect convergence between philosophical theories and descriptively adequate accounts of everyday intuitions.

In fact, a growing body of experimental work suggests that theories of explanation from philosophy *can* usefully inform the psychology of explanation and bear a close correspondence to everyday judgments (for reviews, see Keil 2006; Lombrozo 2006). While there is no consensus on a theory of explanation in philosophy, different strands of theorizing seem to capture different aspects of the psychology of explanation. For example, some studies on the role of explanation in category learning have drawn on subsumption and unification accounts of explanation (e.g., Williams & Lombrozo, in press), while others on categorization and inference are consistent with causal theories (e.g., Rehder 2003b; 2006). Empirical research on the cognitive significance and consequences of different kinds of explanations – specifically, functional versus mechanistic explanations (Kelemen 1999; Lombrozo 2009; under review; Lombrozo & Carey 2006; Lombrozo et al. 2007) – has its roots in Aristotle, but can trace a path to contemporary philosophers such as Daniel Dennett.

One reason to appreciate this richer, philosophically informed psychology of explanation is because it has implications for Machery's heterogeneity hypothesis. In particular, the two distinct summary representations that Machery recognizes – theories and prototypes – can be understood as embodying different kinds of (potentially) explanatory structure. Machery recognizes this point, and in fact rejects philosophical accounts of explanation, such as Salmon's statistical relevance model, in part because allowing statistical relationships to play a role in explanation would "blur the distinction" between theories and prototypes (Machery 2009, p. 102). But perhaps the fact that explanations are sensitive to causal *and* statistical relationships is a reason to *endorse* such accounts. Evidence suggests that explanations are sensitive to multiple kinds of knowledge – about causal structure and functional relationships (Lombrozo & Carey 2006), about statistical regularities (Hilton & Slugoski 1986), and about principled generic knowledge (Prasada & Dillingham 2006). These are precisely the kinds of knowledge that Machery suggests concepts contain.

Recognizing the "heterogeneity" of explanatory structure does not eliminate the heterogeneity of concepts, but it does suggest a path to unifying concepts by appeal to explanation. It also pushes back Machery's concerns about natural kinds and elimination from concepts to explanation: What are the distinct kinds of explanatory relations, and do they as a class support relevant generalizations that suggest "explanation" is a natural kind and a valuable theoretical term for a mature psychology? Perhaps these are the questions we should be asking.

Concepts and theoretical unification[1]

doi:10.1017/S0140525X10000427

Eric Margolis[a] and Stephen Laurence[b]

[a]*Department of Philosophy, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada;* [b]*Department of Philosophy, University of Sheffield, Sheffield S3 7QB, United Kingdom.*

margolis@interchange.ubc.ca **s.laurence@shef.ac.uk**
http://web.mac.com/ericmargolis/primary_site/home.html
http://www.shef.ac.uk/philosophy/staff/profiles/slaurence.html

Abstract: Concepts are mental symbols that have semantic structure and processing structure. This approach (1) allows for different disciplines to converge on a common subject matter; (2) it promotes theoretical unification; and (3) it accommodates the varied processes that preoccupy Machery. It also avoids problems that go with his eliminativism, including the explanation of how fundamentally different types of concepts can be co-referential.

In *Doing without Concepts*, Machery (2009) claims that philosophers and psychologists are not talking about the same thing when they use the term *concept*, and that this is a consequence of their having differing explanatory interests. But there are reasons to reject Machery's division between philosophical and psychological subject matters regarding concepts. First, we should recognize the significant influence that philosophical and psychological theorizing have had on each other. For instance, prototype theorists have been inspired by philosophical critiques of definitions, theory-theorists have drawn upon philosophical accounts of natural kind terms, and developmental psychologists have prioritized addressing the philosophical challenge of explaining how learning enriches a conceptual system. Likewise, philosophers have been deeply influenced by psychological work on typicality effects, essentialist thinking, and conceptual change in childhood (to name just a few examples). Second, even where philosophers and psychologists do have differing explanatory agendas, the same can be said of just about any two fields in cognitive science and in science generally. Linguists and psychologists have differing explanatory aims too, as do *cognitive* psychologists and *neuro*-psychologists – not unlike biologists and chemists. This hardly shows that theorists in these fields aren't talking about the same thing (e.g., NPs, conditioning, or DNA). Third, there is a payoff to identifying a single subject matter that underlies the efforts in philosophy, psychology, and other areas of cognitive science. The result is greater theoretical unification – a prized explanatory virtue.

Now *concept* is a term of art. But we would suggest that Machery gets off on the wrong foot by characterizing concepts as "bodies of information." Instead, concepts should be taken to be mental symbols that have semantic structure (which fixes the propositional content of thoughts via a compositional semantics) and processing structure (which explains how concepts figure in various mental processes). Rather than saying that prototypes, exemplars, and theories constitute fundamentally different types of concepts, it is better to locate such bodies of information in a concept's processing structure. On this approach, the concept DOG is akin to a word in a sentence, and its structure *includes* a prototype, a theory, and so forth (Laurence & Margolis 1999). The principal advantage to viewing concepts in this way is that it makes sense of how philosophers and psychologists *can* be talking about the same thing,

while illuminating the fertile cross-disciplinary interactions that the study of concepts enjoys. And though we (two philosophers) are promoting the idea that concepts are mental symbols, this is not an exclusively philosophical viewpoint. Versions of it are endorsed by many cognitive scientists (e.g., Carey 2009; Jackendoff 2002; Pinker 1997; Pylyshyn 2007; Sperber & Wilson 1995).

Is our account of concepts a hybrid theory? Yes and no. It does bring together prototypes, exemplars, and theories by saying that they are bound to the same mental symbols. The concept DOG, for example, sometimes activates a prototype, sometimes exemplars, and sometimes a theory. But a concept need not have each type of processing structure, and the activation of one part does not require activating other parts. Machery argues that the heterogeneity hypothesis has the explanatory advantage of accounting for the diverse psychological processes that are associated with higher cognitive capacities. But a theory that unites diverse processing structure through links to a common mental symbol can handle this diversity just as well.

Machery asks why theorists who reject the heterogeneity hypothesis do not concede that the various bodies of information (the prototype, theory, etc.) amount to distinct yet co-referential concepts of fundamentally different types – his own view (2009, p. 64). But what justifies Machery's claim that, on his account, a dog prototype, a dog exemplar, and a dog theory *are* co-referential? To the extent that these representational structures have referents, the referents are hardly likely to be the same. For example, a dog-prototype would refer to things that are similar to the central tendency that the prototype describes, while a dog-theory would cover things that are at odds with the central tendency (e.g., the offspring of two dogs that doesn't look anything like typical dogs). By contrast, on our account, the issue of concept identity is easily handled in terms of the type identity of the mental symbol that unifies these various knowledge structures. This symbol's identity *is* a matter of what it refers to, plus features of the representation's vehicle for distinguishing among co-referential concepts with differing cognitive significance (Margolis & Laurence 2007).

As realists about concepts, we also do not find Machery's case for eliminativism compelling. For one thing, we would argue that concepts as we understand them do constitute a natural kind by Machery's criteria. But also, Machery's standard for the reality of psychological kinds is too high. If his standard were enforced – if a kind has to play an important role in many scientific generalization beyond those that characterize it – we'd have to give up on many core psychological constructs, such as *module*, *computation*, and *representation.* But while these high-level kinds may not satisfy Machery's criteria for realism, they play a key role in describing the mind's operations and helping scientists to empirically investigate its overall organization. Moreover, Machery's standard probably cannot even be maintained for his fundamentally different types of concepts; for example, numerous distinct types of structures tend to get lumped together under the heading of a *theory.* And though we lack the space to press the point here, Machery's approach to elimination would also have dire consequences outside of psychology. Arguably, we would have to give up most high-level kinds, including *cell*, *vertebrate*, and *chemical element.*

In sum, a realist account of concepts as mental symbols with both semantic and processing structure addresses the explanatory concerns that Machery raises while avoiding the problems connected to his eliminativism. Taking psychological and philosophical theories of concepts to be about a single subject matter allows for far greater theoretical unification, placing concepts at the center of a broad investigation into the nature of cognitive processes, cognitive development, meaning, justification, and the mind's relation to the world.

ACKNOWLEDGMENT

Eric Margolis would like to thank Canada's Social Sciences and Humanities Research Council for supporting this research.

NOTE

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

Where are nature's joints? Finding the mechanisms underlying categorization

doi:10.1017/S0140525X10000439

Arthur B. Markman

Department of Psychology, University of Texas, Austin, TX 78712.

markman@psy.utexas.edu

http://www.psy.utexas.edu/psy/FACULTY/Markman/index.html

Abstract: Machery argues that concepts are too heterogeneous to be a natural kind. I argue that the book does not go far enough. Theories of concepts assume that the task of categorizing warrants a unique set of cognitive constructs. Instead, cognitive science must look across tasks to find a fundamental set of cognitive mechanisms.

There is a persistent worry that cognitive scientists may not be carving nature at its joints. This fear underlies debates over whether computational representations or dynamical systems best explain cognitive processing (e.g., Markman & Dietrich 2000; Spivey 2007). It lies at the heart of critiques of the use of brain imaging to understand cognitive function (Uttal 2001). This issue is also central to the target book, *Doing without Concepts* (Machery 2009).

This question is important, because cognitive scientists typically organize theories around tasks. Memory is explored by having people study items and then probing their memory for those items at some later time. Decision-making research involves presenting people with a set of options and having them select one. As Machery points out, categorization research often involves specific tasks such as classification and category-based induction. Theories then aim to explain performance in these tasks.

Machery takes the structure of the cognitive science literature on concepts as a given and then suggests that the notion of a concept is misleading. On his view, there are (at least) three distinct types of concepts: prototypes, exemplars, and theories. Using a single term – concepts – to refer to all of these is dangerous, because it fails to carve nature at its joints.

I suggest that the problem is even worse than Machery makes it out to be. Fundamentally, the set of tasks that we study involves a series of cross-cutting cognitive mechanisms. At present, cognitive science assumes that tasks like classification and category-based induction require an explanation that involves some set of representations and processes that are shared (to some degree) across different kinds of categorization tasks, but are relatively distinct from the kinds of representations and processes that are involved in decision-making, memory, or attention.

Ultimately, we need to reorient our theories to find the commonalities across tasks that are typically thought of as different. In the study of concepts, there are already some hints in the literature that this reorientation is starting to take place.

The most prominent shift in research on categorization comes from work relating categorization to memory. For example, the research by Ashby, Maddox, and colleagues draws parallels between behavioral and neurobiological research on categorization and memory (Ashby et al. 1998; Maddox & Ashby 2004). This work incorporates research from neural systems involved in memory to make predictions for performance in category learning experiments. Research on the kinds of categories that amnesics can learn is also inspired by the desire to create parallels between memory and categorization (Knowlton et al. 1994).

Research on the influences of learning tasks on category learning also forms parallels between categorization and memory