BEHAVIORAL AND BRAIN SCIENCES (2010) **33**, 195–244
doi:10.1017/S0140525X09991531

# Précis of *Doing without Concepts*

**Edouard Machery**
*Department of History and Philosophy of Science, 1017CL, University of Pittsburgh, Pittsburgh, PA 15260*
**Machery@pitt.edu**
**www.pitt.edu/~machery/**

**Abstract:** Although cognitive scientists have learned a lot about concepts, their findings have yet to be organized in a coherent theoretical framework. In addition, after twenty years of controversy, there is little sign that philosophers and psychologists are converging toward an agreement about the very nature of concepts. *Doing without Concepts* (Machery 2009) attempts to remedy this state of affairs. In this article, I review the main points and arguments developed at greater length in *Doing without Concepts.*

**Keywords:** categorization; concept; concept combination; dual-process; eliminativism; exemplar; induction; meaning; natural kinds; prototype; reference; theory

The study of concepts is in an odd state of disarray. Cognitive scientists working on categorization, induction, and reasoning have discovered a dazzling amount of phenomena. New work on prototypes in the 1990s and early 2000s, innovative ideas on causal cognition in the first decade of the twenty-first century, the development of the neo-empiricist approach that assimilates the tokening of a concept to a multi-modal perceptual simulation, and the promising growth of the neuropsychology of concepts have rejuvenated the field. At the same time, this extraordinary amount of findings has yet to be organized in a coherent theoretical framework. The current theories of concepts – prototype theories, exemplar theories, theory theories, and neo-empiricist theories – fail to explain all the known phenomena, and there is very little agreement about what concepts are. *Doing without Concepts* (Machery 2009) attempts to provide such a theoretical framework. In this article, I review the main points and arguments developed at greater length in the book, and I conclude that abandoning the very notion of concept is probably required to remedy the state of disarray of the current psychology of concepts.

## 1. Regimenting the use of *concept* in cognitive science

Because cognitive scientists rarely spell out the notion of concept in detail, I begin by making explicit the notion of concept that is typically used within cognitive science. My goal in chapters 1 and 2 is threefold: To clarify this notion, to regiment the use of the term *concept*, and to show that philosophers and cognitive scientists theorize about different things when developing theories of concepts.

The cognitive processes that underwrite cognitive competences are typically assumed to access some relevant information or knowledge. Some bodies of information are only accessed by particular processes: For instance, our implicit knowledge of the syntax of the natural languages we speak (e.g., English) is only accessed by the processes involved in parsing and in producing sentences. When this is the case, I will say that the relevant body of information is "proprietary to a particular cognitive process." By contrast, some information is "non-proprietary": It is accessed by the cognitive processes that underlie several distinct cognitive competences. Cognitive scientists often assume that the cognitive processes underlying our higher cognitive competences access the same bodies of knowledge. For instance, the processes underlying categorization, induction, and speech are hypothesized to access the same body of knowledge about dogs when people classify something as a dog, when they make an induction about dogs, and when they understand sentences containing the word *dog*. This knowledge is assumed to be stored in long-term memory.

These preliminary points having been made, I propose to characterize the notion of concept as follows: Within cognitive science, a concept of *x* is a body of information about *x* that is stored in long-term memory and that is used by default in the processes underlying most, if not

EDOUARD MACHERY is Associate Professor of History and Philosophy of Science at the University of Pittsburgh, a resident fellow of the Center for Philosophy of Science (University of Pittsburgh), a member of the Center for the Neural Basis of Cognition (Carnegie Mellon University and University of Pittsburgh), and an associate editor of the *European Journal for Philosophy of Science*. His research focuses on the philosophical issues raised by psychology and cognitive neuroscience with a special interest in concepts, moral psychology, the relevance of evolutionary biology for understanding cognition, modularity, the nature, origins, and ethical significance of prejudiced cognition, and the methods of psychology and neuroscience. Machery has published more than 50 articles and chapters on these topics in venues such as *Analysis*, *The British Journal for the Philosophy of Science*, *Cognition*, *Mind & Language*, *The Monist*, *Philosophy and Phenomenological Research*, and *Philosophy of Science*. He is also involved in the development of experimental philosophy, having published several noted articles in this field.

 0140-525X/10 $40.00

all, higher cognitive competences when they result in judgments about *x*. I call this characterization "C."

It is important to highlight some significant properties of concepts, so understood. First, concepts can be about classes of objects (e.g., DOG), events (e.g., GOING TO THE DENTIST), substances (e.g., WATER), and individuals (e.g., BARACK OBAMA). Second, concepts are non-proprietary: DOG is used by the processes underlying categorization, induction, linguistic understanding, metaphor building, planning, and perhaps other competences. Third, the elements of information that are constitutive of a concept can vary over time and across individuals. Fourth, it might be unclear whether a given element of information about *x* belongs to a concept of *x*. Finally, concepts are used *by default* in the cognitive processes underlying higher cognitive competences (I call "Default" the hypothesis that some bodies of knowledge are retrieved by default when one is categorizing, reasoning, drawing analogies, making inductions, and so on). This entails that our concept(s) of, say, dogs is (are) only a subset of our whole knowledge about dogs: The knowledge that is constitutive of DOG is the knowledge about dogs that is retrieved by default from long-term memory when we reason about dogs, categorize things as dogs, and so forth. I call "background knowledge" about dogs the knowledge about dogs that is not part of the concept(s) of dogs. Our background knowledge about a category, a substance, some kind of events, and the like, can be called upon occasionally when the default body of knowledge is insufficient to solve a cognitive task.

At this point, it is useful to spell out the notion of being used by default at greater length. A default body of knowledge about *x* is the body of knowledge that is presumptively taken to be relevant when one reasons about *x*, when one categorizes things as *x*, and so on. The knowledge that is stored in a concept of *x* is preferentially available when we think, reason, and so forth, about *x*. So to speak, it spontaneously comes to mind.

The proposed characterization of the notion of concept captures much of what is implicit in the use of the term *concept* in cognitive science. However, it is also clear that C is partly at odds with some characterizations of the notion of concept found in the literature, which are discussed at length in Chapter 1 of *Doing without Concepts.* In any case, with the proposed characterization of the notion of concept, I do not merely aim at capturing the use of *concept* in cognitive science. I also want to regiment it: I contend that some bodies of knowledge are retrieved by default from long-term memory when one reasons, categorizes, and so forth (see section 3 for a defense of this claim), and I propose that the term *concept* should be used to refer to these bodies of knowledge.

## 2. Individuating concepts

It is certainly possible for a given individual to have several concepts of the same category (e.g., several concepts of chairs) or of the same substance (e.g., several concepts of gold): For instance, one might think of chairs in several distinct ways, each of which corresponds to a different concept of chairs. This possibility raises the following question, which is examined in Chapter 3 of *Doing without Concepts:* What does it mean for two bodies of knowledge about *x* (e.g., the knowledge that water is typically transparent and the knowledge that water is made of molecules of $H_2O$) to be part of the same concept of *x* (WATER) rather than of two distinct concepts ($WATER_1$ and $WATER_2$)?

I propose the following individuation criteria (respectively, *Connection* and *Coordination*). When two elements of information about *x*, A and B, fulfill either of these criteria, they belong to two distinct concepts:

1. If retrieving A (e.g., water is typically transparent) from long-term memory and using it in a cognitive process (e.g., a categorization process) does not facilitate the retrieval of B (water is made of molecules of $H_2O$) from long-term memory and its use in some cognitive process, then A and B belong to two distinct concepts ($WATER_1$ and $WATER_2$).
2. If A and B yield conflicting judgments (e.g., the judgment that some liquid is water and the judgment that this very liquid is not water) and if I do not view either judgment as defeasible in light of the other judgment (i.e., if I hold both judgments to be equally authoritative), then A and B belong to two distinct concepts ($WATER_1$ and $WATER_2$).

Let's clarify these two criteria a bit further. *Connection* unpacks the idea that the components of a given concept must be connected (or "linked" as I say in *Doing without Concepts*): If the beliefs that water is typically transparent and that water is made of molecules of $H_2O$ are part of the same concept, WATER, then using the first belief to classify some sample as water enables me to conclude that it is made of molecules of $H_2O$. Accessing one part of a concept makes the other parts accessible and poised to be used in cognitive processing (for related discussion, see Millikan 2000, Ch. 10).

*Coordination* expresses the idea that the parts of a single concept should not result in conflicting judgments that are both taken to be equally authoritative. Coordination is compatible with the parts of a given concept yielding conflicting judgments, provided that all but one are viewed as defeasible. To illustrate, although the belief that grandmothers have gray hair and the belief that grandmothers are mothers of parents would yield different judgments if they were used to categorize a young-looking grandmother (e.g., Sarah Palin), they could still both be part of the concept GRANDMOTHER if one of the two judgments (Sarah Palin is not a grandmother) is defeated by the other one (Sarah Palin is a grandmother). In effect, the first belief would be treated as a heuristic that sometimes leads us astray.

These two criteria merely explain what it is for a given individual to have one or several concepts about some class of objects, some substance, and so on (within-person individuation). It does not explain what it is for two or more individuals to have the same concept of *x* (between-person individuation). Now, one might expect a theoretician of concepts to provide individuation criteria for this situation, too. However, I doubt that these are needed for the psychology of concepts, since they seem to play no explanatory role in psychology.

## 3. Defending the proposed notion of concept

Some cognitive scientists and philosophers of psychology reject Default (viz., the assumption that some bodies of knowledge are retrieved by default when one is categorizing, reasoning, drawing analogies, and making inductions).

In Chapters 1 and 8 of the book, I rebut the criticisms mounted against this assumption.

Several findings suggest that typicality varies across contexts. Roth and Shoben (1983) have shown that depending on the linguistic context (e.g., when participants are presented with "Stacy volunteered to milk the *animal* whenever she visited the farm" or "Fran pleaded with her father to let her ride the *animal*"), participants judge that different animals are typical (cows and goats for the first linguistic context, horses and mules for the second). Similarly, Barsalou (1985) reports that judgments of typicality vary across contexts. In Study 2, participants' typicality judgments about the members of two groups differed when these groups were conceptualized differently (as physical education teachers and current event teachers, on the one hand, and as two invented types of programmers – namely, Q programmers and Z programmers – on the other hand). This study also shows that when participants are familiarized with a given category in different contexts, their judgments of typicality vary. Barsalou (1987; 1993) also reports that the typicality of objects varies when participants are asked to take different points of view on these objects. For example, people judge differently the typicality of birds when they take the point of view of someone from China and when they report their typicality evaluation from their own point of view. In addition, the correlation between typicality judgments across participants is low (circa .5) and lower than expected for a given subject on two different occasions (around .8).

Theorists have used such findings to challenge the idea that some bodies of knowledge are retrieved by default when one categorizes, draws inductions, reasons, and the like. Barsalou (1985) concludes that:

> People may not retrieve the same concept from long-term memory every time they deal with a particular category. Instead they may construct a diverse variety of concepts in working memory to represent a particular category across different situations such that the concept used to represent a category is rarely, if ever, the same. According to this view, long-term memory does not contain invariant concepts. (Barsalou 1985, p. 646)

(For similar arguments, see also Barsalou 1987; Smith & Samuelson 1997, p. 170.)

Theorists who reject Default have drawn two distinct conclusions. As discussed in Chapter 8 of *Doing without Concepts*, some theorists, such as L. Smith (Smith & Samuelson 1997), contend that Default is part of the very notion of concept, and they conclude that, since there are no such things as bodies of knowledge retrieved by default from long-term memory, there are no such things as concepts. Smith and Samuelson (1997, p. 190) conclude that "a successful theory of categories [. . .] might require that we give up timeless abstractions such as concepts."

While agreeing that Default is part of the notion of concept typically used in cognitive science, other theorists, such as Barsalou and Prinz, propose to redefine the notion of concept. Concepts should be thought of as the bodies of knowledge in working memory that are used at a given time in a given task: They are constructed on the fly to deal with the peculiarities of the task at hand; and they typically vary from time to time (Barsalou 1993, p. 29; Prinz 2002; Malt & Sloman 2007).

Though Barsalou (1993), Prinz (2002), L. Smith (see Smith & Samuelson 1997), and Malt and Sloman (2007) take the body of evidence reviewed above to establish either conclusion, I demur on three grounds. First, the hypothesis that some bodies of knowledge are retrieved by default from long-term memory and used in the processes underlying the higher cognitive competences is consistent with some variation in the bodies of knowledge that are used at any given time. This variation can have two sources. When we reason about $x$, in addition to the default body of knowledge about $x$, we sometimes retrieve some specific elements from our background knowledge about $x$. In addition, once retrieved from memory, the body of knowledge that is retrieved by default can be tailored to the peculiarities of the given situation. On this view, knowledge retrieval would be a two-step procedure: (1) Retrieve the default body of knowledge from long-term memory; (2) tailor it to the situation. (Sperber & Wilson [1998] present similar ideas.) Thus, the mere fact that performances in experimental tasks vary from time to time does not show that there are no bodies of knowledge retrieved by default from long-term memory. What would not be consistent with Default is a very large variability across contexts of the knowledge brought to bear on tasks. And, indeed, Barsalou claims that there is a "*tremendous variability* in performances . . . not only in category membership, but also in typicality, definitions, and probably most other categorization tasks" (1993, p. 34; my emphasis). However, as we shall now see, the relevant variability in performances is in fact moderate.

Second, the nature of the variation found by Barsalou, Malt, Sloman, and others is either irrelevant to evaluate Default or supportive of it. Let us consider first the pieces of evidence that turn out to be irrelevant to evaluate Default. Many findings about the context-sensitivity of typicality are misleading. Roth and Shoben's (1983) findings merely show that (unsurprisingly) people evaluate differently the typicality of target animals with respect to the category of animals that get milked and the category of animals one uses to ride. That typicality varies when evaluated from different points of view (Barsalou 1993), does not show that people's concepts vary across occasions, since, in effect, one asks participants not to use their own concept of $x$ to complete a task when one asks them to make judgments about $x$ from someone else's point of view. Similarly, the variability of the typicality judgments made by different individuals says nothing about whether a given subject retrieves a default body of knowledge across occasions. Showing, as Barsalou (1985) does, that the typicality structure of a given class of objects can substantially vary when it is conceptualized differently (viz., as physical education teachers and current events teachers, on the one hand, and as Q programmers and Z programmers, on the other hand) is interesting; but this finding does not show that the concept of a given category varies across contexts and circumstances because, properly speaking, current events teachers and Z programmers are two distinct categories, although they are composed of the same individuals. Showing that the typicality structure of a given category varies when people are familiarized with this category in different situations is also interesting; but again, it says little about whether someone who is familiarized with a given category one way will rely on a default body of knowledge about this category. In addition, some results that allegedly undermine Default in fact support it. As noted, Barsalou (1987; 1993) reports that, on average, the test-retest reliability of typicality judgments is at least .8. It is also higher when participants are re-tested one hour and a day after the first test. Furthermore, Barsalou reports that the typicality of highly typical and atypical items does not change much. These results are evidence that, across occasions, a default concept is retrieved from long-term memory.

Finally, a large body of evidence supports Default. Consider linguistic understanding (Ziff 1972, discussed in Murphy & Medin 1985). The sentence, "A cheetah can outrun a man," is meaningful, and most people would agree with it. However, as Murphy and Medin put it (1985, p. 303), it is true only if the represented cheetah is not "a 1-day old cheetah, or an aged cheetah with arthritis, or a healthy cheetah with a 100-pound weight on its back." But when we read, "A cheetah can outrun a man," these representations of cheetahs do not come to mind. This phenomenon suggests that when a speaker utters "A cheetah can outrun a man," or when a hearer or a reader understands this sentence, she retrieves from memory a default body of knowledge about cheetahs. Perhaps one will object that when one reads the sentence, "A cheetah can outrun a man," one merely constructs a context-appropriate interpretation of "cheetah" rather than retrieving a default body of knowledge about cheetahs. If this were true, then people would also construct a context-appropriate representation of cheetahs if they had to decide whether the sentence, "A man can outrun a cheetah," is true. They would, for example, imagine an old, three-legged cheetah, and the sentence, "A man can outrun a cheetah," would then be judged true too. However, I predict that, under time pressure, people would judge the sentence, "A man can outrun a cheetah," to be false. This would be evidence that in such conditions they retrieve the very default body of knowledge they retrieve when they read, "A cheetah can outrun a man." Naturally, with no time pressure, they could construct an interpretation of *cheetah* under which the sentence, "A man can outrun a cheetah," is true. But this is consistent with the existence of default bodies of knowledge, because, as I have already proposed, people can and do retrieve some additional information (viz., some information not contained in their concepts) from their background knowledge.

Behavioral studies also show that some information about a category, substance, and so on, is retrieved automatically in every context (Barsalou 1982; Whitney et al. 1985). Barsalou (1982) found that when people judge that a property (e.g., *stinks*) spontaneously "comes to mind" when they read a given noun (e.g., "skunk"), reaction times in a property-verification task are similar when the noun is presented in a relevant linguistic context ("The skunk stunk up the entire neighborhood") and when it is presented in an irrelevant linguistic context ("The skunk was under a large willow"). By contrast, reaction times are larger in the latter condition ("The roof had been renovated prior to the rainy season") than in the former condition ("The roof creaked under the weight of the repairman") when people judge that a property (e.g., *can be walked upon*) does not spontaneously come to mind when they read a given noun (e.g., "roof"). Barsalou calls the first kind of property "context-independent" and the second kind "context-dependent."

Cognitive neuroscience provides further evidence in support of Default (although the relevant studies were not developed to test this hypothesis). After having trained participants with novel tools, Weisberg et al. (2006) recorded brain activation in a perceptual task (a visual matching task). To complete this task, one needs only appeal to some structural information about the shape of the novel tools; thus, one would expect the medial portion of the fusiform gyrus to be activated (for review, see Martin 2007). Interestingly, activation was also recorded in the intraparietal sulcus, the premotor cortex, and the medial temporal gyrus, areas of the brain that are known to store information about the typical movements associated with tool use. It thus seems that the perceptual task resulted in the automatic retrieval of information that was not needed to solve the task, consistent with the idea that people have bodies of knowledge that they retrieve by default (for similar findings, see also James & Gauthier 2003; Hoenig et al. 2008).

## 4. Developing a psychological theory of concepts

It is important to keep in mind that the notion of concept proposed earlier (viz., C) does not amount to a *theory* of concepts. Rather, C does two things: It spells out what is implicit in cognitive scientists' use of the term *concept*, and it proposes to regiment this use. So, what does a psychological theory of concepts consist in?

As I explain in Chapter 1 of *Doing without Concepts*, psychological theories of concepts typically attempt to identify the properties that are typical of concepts ("the general properties of concepts"). Five kinds of properties are of interest to cognitive scientists. First, cognitive scientists are interested in the nature of the information that is constitutive of concepts. For instance, cognitive scientists want to know whether concepts consist of some statistical information about the properties that are characteristic of a class or of a substance, as prototype theorists have proposed (e.g., Hampton 1979; 1981; 2006; 2007; Smith 2002), or whether they consist of causal generalizations (e.g., Gopnik & Meltzoff 1997; Griffiths et al. 2007; Murphy & Medin 1985; Rehder 2003a; Tenenbaum et al. 2007). Second, cognitive scientists are interested in the nature of the processes that use concepts. For instance, some psychologists have argued that these processes are based on similarity (e.g., Hampton 1993), while others disagree (e.g., Rips 1989). Third, cognitive scientists develop hypotheses about the nature of the vehicles of concepts: Thus, neo-empiricists such as Barsalou and Prinz contend that the vehicle of concepts is similar to the vehicle of perceptual representations (Barsalou 1999; 2008b; 2009; Machery 2006c; Prinz 2002; 2005). Fourth, for about a decade, cognitive scientists have attempted to identify the brain areas that are involved in possessing concepts (for reviews, see, e.g., Mahon & Caramazza 2009; Martin 2007; Pulvermüller 2005). Finally, cognitive scientists have developed hypotheses about the processes of concept acquisition (e.g., Ashby & Maddox 2004; Gopnik 2003).

In addition to developing hypotheses about the general properties of concepts, cognitive scientists have shown some interest in distinguishing different types of concepts and in identifying the properties of these types of concepts. Medin et al. (2000) have rightly insisted on the importance of this task and on its relative neglect by cognitive scientists.

Why do cognitive scientists want a theory of concepts? Theories of concepts are meant to explain the properties of our cognitive competences. People categorize the way they do, they draw the inductions they do, and so on, because of the properties of the concepts they have. Thus, providing a good theory of concepts could go a long way toward explaining some important higher cognitive competences.

## 5. *Concept* in cognitive science and in philosophy

The term *concept* is used in philosophy, particularly in the philosophy of mind, as well as in cognitive science. Chapter 2 of *Doing without Concepts* examines the relation between these two uses. It is common among philosophers to assume that *concept* is used in the same sense in philosophy and in cognitive science and that psychologists' theories of concepts aim at answering the issues philosophers are interested in (Edwards 2009; Fodor 1998; 2008; Laurence & Margolis 1999; Margolis 1994; 1995; Margolis & Laurence 2006; Rey 1983; 1985; 2009b). In addition, it is common to hold that as answers to the issues of interest in philosophy, psychological theories of concepts are defective. Thus, Fodor (2003) concludes his review of Gregory Murphy's book, *The Big Book of Concepts*, as follows:

> It is part of our not knowing how the mind works that we don't know what concepts are or what it is to have one. Just about everything that current cognitive science says about either topic is wrong .. . . Gregory Murphy's book tells you most of what there is to the psychology of concepts. Read it, therefore, by all means; but don't even consider believing it. (Fodor 2003, p. 4)

It is also not uncommon to see some philosophical theories criticized for being unable to explain how we categorize, make inductions, and so on (Prinz 2002; for discussion, see Edwards 2009).

Philosophers' take on psychological theories is mistaken: Philosophical and psychological theories of concepts are not meant to answer the same questions and are thus not competing. Typically, by *concept*, philosophers refer to that which allows people to have propositional attitudes (beliefs, desires, etc.) about the objects of their attitudes. The concept of a triangle is therefore that which allows people to have propositional attitudes (beliefs, desires, etc.) about triangles. A theory of concepts in philosophy attempts to determine the conditions under which people can have propositional attitudes about the objects of their attitudes (Fodor 1998; 2008; Peacocke 1992; 2008;), but not to explain the properties of our higher cognitive competences. By contrast, psychologists attempt to explain the properties of our categorizations, inductions, and so forth, but they do not attempt to determine the conditions under which people are able to have propositional attitudes about the objects of their attitudes. Furthermore, psychologists do not need to hold, and typically do not hold, that we are able to have propositional attitudes about the objects of our attitudes by virtue of having specific bodies of knowledge about them. For instance, prototype theorists do not need to hold, and typically do not hold, that having a prototype is a condition for being able to have attitudes about the objects of our attitudes. In fact, prototype theorists are silent on this question.

The upshot of this argument should be clear. Although both philosophers and cognitive scientists use the term *concept*, they are not talking about the same things. Cognitive scientists are talking about a certain kind of bodies of knowledge, whereas philosophers are talking about that which allows people to have propositional attitudes. Many controversies between philosophers and psychologists about the nature of concepts are thus vacuous.

## 6. The heterogeneity hypothesis versus the received view

Cognitive scientists of concepts naturally acknowledge differences between concepts: The concept of dogs is clearly different from the concept of cats. More interesting, they also acknowledge differences between kinds of concepts: For instance, there has been much work in experimental and developmental psychology on the differences between the concepts of animals and the concepts of artifacts (e.g., Bloom 1996; Gelman 1988; 2003; Gelman & Markman 1986; Malt & Sloman 2007). But above and beyond these differences, cognitive scientists often assume that concepts share many properties that are scientifically interesting. In Chapter 3 of *Doing without Concepts*, I call this assumption "the received view." It is well expressed by Gregory Murphy:

> The psychology of concepts cannot by itself provide a full explanation of the concepts of all the different domains that psychologists are interested in.. . . The details of each of these must be discovered by the specific disciplines that study them. . .. Nonetheless, the general processes of concept learning and representation may well be found in each of these domains. (Murphy 2002, pp. 2–3)

The received view has been instrumental in the debates that have marked the history of the psychology of concepts since the 1970s. Cognitive scientists who are committed to different theories of concepts (say, a particular prototype theory and a particular exemplar theory) have attempted to discover properties of our higher cognitive competences (e.g., the exemplar effect reported in Medin & Schaffer 1978) that were easily explained by the theory they endorsed (e.g., the exemplar theory), but were not easily explicable by the competing theory (prototype theories do not naturally explain the exemplar effect; for discussion, see Smith & Minda 2000). This research strategy makes sense only if one supposes that a single theory of concepts should be able to account for all the relevant phenomena. If, contrary to the received view, the class of concepts divides into several kinds that have little in common, the distinct theories of concepts that characterize these kinds of concepts will account for different phenomena, and the fact that theory A, but not theory B, explains some phenomenon, such as the exemplar effect, will not necessarily constitute evidence against theory B.

As I explain in Chapter 3, the received view stands in sharp contrast with a view about concepts developed in my own work (see also Machery 2005): the Heterogeneity

Hypothesis. According to this hypothesis, the class of concepts divides into several distinct kinds that have little in common – "the fundamental kinds of concepts." Because the class of concepts divides into distinct fundamental kinds, it is a mistake to assume that there are many general properties of concepts, and that a theory of concepts should attempt to describe these. Although the heterogeneity hypothesis can be developed in several ways (Machery 2005; 2006a; Piccinini & Scott 2006), I contend that a given category (e.g., dogs), a given substance (e.g., water), or a given kind of events (e.g., going to the dentist) is typically represented by several distinct concepts (e.g., $\text{DOG}_1$ and $\text{DOG}_2$). These coreferential concepts belong to the fundamental kinds of concepts. Each coreferential concept can be used to categorize, draw inductions, understand the relevant words, make analogies, and so forth (Fig. 1).

In addition, the heterogeneity hypothesis contends that these concepts are often used in distinct processes. That is, we have several categorization processes, several induction processes, and the like, each of which uses a distinct fundamental kind of concepts (Fig. 2).

If the heterogeneity hypothesis is correct, the class of concepts is not a natural kind. Natural kinds are classes whose members share many scientifically important properties in virtue of one or several causal mechanisms (Boyd 1991; 1999; Griffiths 1997; Machery 2005). Water and dogs are natural kinds in this sense; for example, samples of water have many properties in common in virtue of consisting of the same molecules of $H_2O$. In a given science, the scientific classificatory scheme is developed to identify the natural kinds in the relevant domain because identifying these kinds allows scientists to discover new generalizations. Scientific classificatory schemes are modified when they do not identify the relevant natural kinds (as happened during the chemical revolution in the eighteenth century), and scientific notions are often eliminated when it is found that they fail to pick out natural kinds (for discussion, see sect. 13). Because the hypothesized fundamental kinds of concepts have little in common, the class of concepts cannot be a natural kind if the heterogeneity hypothesis is correct.

## 7. What kind of evidence could support the heterogeneity hypothesis?

Chapter 5 of *Doing without Concepts* describes three kinds of evidence that can provide support for the heterogeneity hypothesis. I consider them in turn in this section.

Suppose that the class of concepts divides into several fundamental kinds, and suppose also that coreferential concepts are often used in distinct cognitive processes (i.e., distinct categorization processes, distinct induction processes). What properties would we then expect to observe in experimental tasks? First, if experimental conditions can be designed that trigger only one of the hypothesized categorization processes or only one of the hypothesized induction processes, we should expect some experimental findings to be best explained if the concepts used in the relevant experimental tasks are identical to a first fundamental kind of concepts, other experimental findings to be best explained if the concepts used in the relevant experimental tasks are identical to a second fundamental kind of concepts, and so on. For instance, if one hypothesizes that the fundamental kinds of concepts are exemplars and prototypes, then one might find categorization tasks where participants' categorization performances are best explained if the concepts used in these tasks are prototypes and other categorization tasks where participants' categorization performances are best explained if the concepts used in these tasks are exemplars.

Second, suppose that in some conditions, several of the hypothesized categorization (or induction) processes are triggered at the same time. Then, in some circumstances, these processes will produce congruent outputs (e.g., categorization judgments), while they will produce incongruent outputs in other circumstances. When the latter

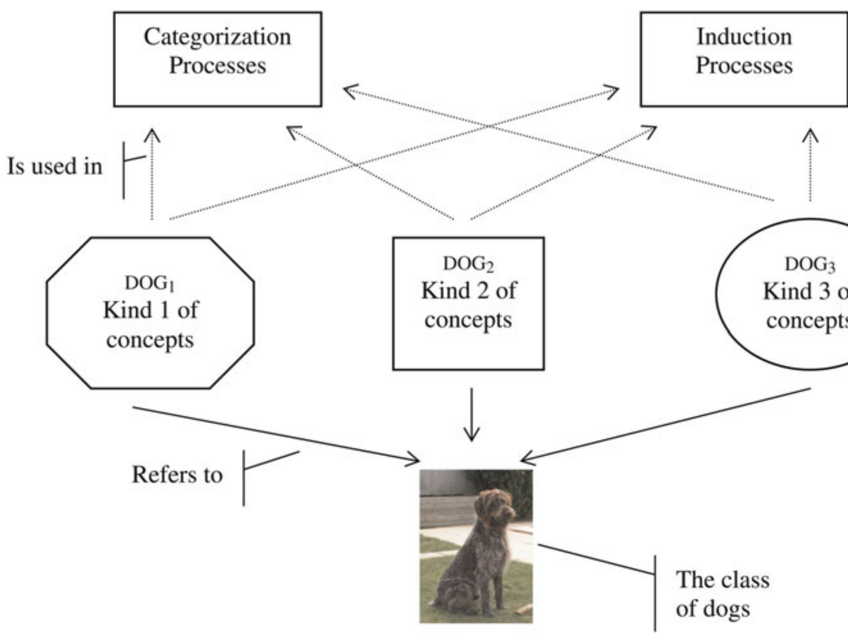


Figure 1. The heterogeneity hypothesis.

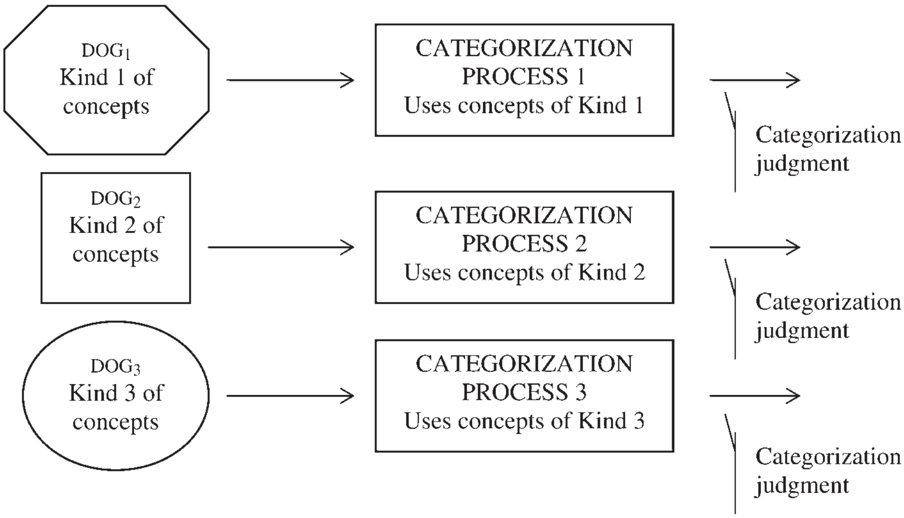


Figure 2. Several processes underlying a given cognitive competence.

happens, participants will have to decide between conflicting judgments. Participants should thus be expected to be slower when the hypothesized processes are expected to yield conflicting outputs than when they are not. Test-retest reliability should also be expected to be lower in the experimental conditions where it is hypothesized that the hypothesized categorization (induction) processes will result in incongruent outputs than when it is hypothesized that they will result in congruent outputs. Noticeably, this kind of evidence (particularly, slower reaction times) has extensively been used in cognitive science to argue that a given task involves two independent cognitive processes (e.g., Greene et al. 2001).

Finally, experimental and neuropsychological dissociations can be used to determine whether a given task involves several processes. The epistemology of dissociations is intricate (Ashby & Ell 2002; Caramazza 1986; Glymour 1994; Dunn & Kirsner 1988; 2003; Plaut 1995; Shallice 1988; Van Orden et al. 2001), but I maintain that dissociations provide evidence about the number and nature of the processes underlying a given competence.

## 8. The fundamental kinds of concepts

Now that the nature of the evidence required to support the heterogeneity hypothesis has been clarified, it is time to lay my cards on the table: What are the fundamental kinds of concepts? And what is the evidence for their existence? In what follows, I will briefly describe the kind of evidence supporting the heterogeneity hypothesis, but due to limitations of space this will not amount to a comprehensive articulation of the evidence adduced in Chapters 6 and 7 of *Doing without Concepts.*

In Chapter 4, I propose that the class of concepts divides into at least three fundamental kinds of concepts – prototypes, exemplars, and theories. These three theoretical constructs are well known in the psychology of concepts as they correspond to the entities posited by the main theories of concepts that have been developed since the 1970s (for a review, see Murphy 2002). Although there are several distinct theories about what prototypes, exemplars, and theories are, these theories agree about the distinctive features of each type of concept. In substance, prototypes are bodies of statistical knowledge about a category, a substance, a type of event, and so on. For example, a prototype of dogs could store some statistical knowledge about the properties that are typical of dogs and/or the properties that are diagnostic of the class of dogs. According to prototype theories, when I categorize, draw an induction, make an analogy, and so forth, I spontaneously bring to mind the properties that are typical, diagnostic (etc.) of the relevant category, substance, and so forth. Prototypes are typically assumed to be used in cognitive processes that compute the similarity between a prototype and other representations, such as the representations of the objects to be categorized, in a linear manner (Hampton 1979; 1993; 2006; 2007; Rosch & Mervis 1975; Smith 2002). Exemplars are bodies of knowledge about individual members of a category (e.g., Fido, Rover), particular samples of a substance, and particular instances of a kind of event (e.g., my last visit to the dentist). For instance, according to exemplar theories, a concept of dogs would consist of a set of bodies of knowledge about specific dogs (Rover, Fido). When I categorize, draw an induction, make an analogy, and so on, I spontaneously bring to mind the properties of specific members of the relevant categories, of specific samples of the relevant substances, et cetera. Exemplars are typically assumed to be used in cognitive processes that compute the similarity between a set of exemplars and other representations, such as the representations of the objects to be categorized, in a nonlinear manner (Medin & Schaffer 1978; Nosofsky 1986; 1992; Nosofsky & Stanton 2005). Theories are bodies of causal, functional, generic, and nomological knowledge about categories, substances, types of events, and the like. A theory of dogs would consist of some such knowledge about dogs. When I categorize, draw an induction, make an analogy, and so on, I spontaneously bring to mind this causal, functional, generic, and nomological knowledge. Recent work on causal knowledge suggests that theories might be used in cognitive processes that are similar to the algorithms involved in causal reasoning (Gopnik et al. 2004).

Thus, the heterogeneity hypothesis proposes that for many categories, substances, kinds of events, we typically have a prototype, a set of exemplars, and a theory about them. Thus, we might have a prototype of dogs, a set of exemplars of particular dogs, and a theory about dogs. Furthermore, prototypes, exemplars, and theories are often used in distinct processes. The heterogeneity hypothesis proposes that we have a prototype-based categorization process, an exemplar-based categorization process, and a theory-based categorization process. Note that the hypothesis is not merely that our knowledge about dogs includes some knowledge about their typical or diagnostic properties; some knowledge about some particular dogs; and some causal, functional, and generic knowledge (as Rey [2009b] erroneously believes). This would be a fairly uncontroversial claim. Rather, the claim is that, for most categories, substances, et cetera, we have several bodies of knowledge that are retrieved by default and that are often used in distinct cognitive processes (e.g., several distinct categorization processes).

The heterogeneity hypothesis also contends that the fundamental kinds of concepts have little in common. This is indeed the case if these fundamental kinds really consist of prototypes, exemplars, and theories. They consist of different types of knowledge, they are used in different kinds of processes, and they are probably acquired by distinct processes. Given what cognitive scientists working on concepts are interested in (see sect. 4), they count as very different kinds of entities.

One might perhaps object that prototypes, exemplars, and theories do have some properties in common. In particular, they are all bodies of knowledge, they are all stored in long-term memory, and they are all used in the processes underlying higher cognition. This, however, does not undermine the heterogeneity hypothesis, for the claim that prototypes, exemplars, and theories have little in common really states that the fundamental kinds of concepts have in common few properties that are *scientifically interesting* and *discovered empirically*. Prototypes, exemplars, and theories have in common numerous properties that are not of interest to cognitive scientists (e.g., they are all mental states). In addition, far from being discovered empirically, the mentioned commonalities between prototypes, exemplars, and theories (e.g., they are all bodies of knowledge, they are all stored in long-term memory, etc.) are in fact used to identify what concepts are.

So, what is the evidence for the claim that our long-term memory stores prototypes, exemplars, and theories? When one examines 30 years of research on categorization and induction, as I do in Chapters 6 and 7, one finds out that in both areas of research, some phenomena are well explained if the concepts elicited by some experimental tasks are prototypes; some phenomena are well explained if the concepts elicited by other experimental tasks are exemplars; and yet other phenomena are well explained if the concepts elicited by yet other experimental tasks are theories. As already noted, if one assumes that experimental conditions prime the reliance on one type of concepts (e.g., prototypes) instead of other types (e.g., exemplars and theories), this provides evidence for the heterogeneity hypothesis.

Let's illustrate this situation with the work on categorical induction – the capacity to conclude that the members of a category possess a property from the fact that the members of another category possess it and to evaluate the probability of this generalization (for review, see Feeney & Heit 2007; Heit 2000; Murphy 2002, Ch. 8; Sloman & Lagnado 2005). A large number of phenomena suggest that prototypes or exemplars are sometimes involved in induction (Osherson et al. 1990; Sloman 1993). Similarity-based models of induction, which assume that processes underlying induction are defined over either prototypes or exemplars, explain best two well-known findings about induction – the similarity effect and the typicality effect. Other phenomena are best explained if the concepts involved in the relevant experimental conditions are causal theories. Investigating the judgments made by tree experts (landscapers, taxonomists, and parks maintenance personnel) about the strength of inductive conclusions about trees, Proffitt et al. (2000) found that, rather than relying on typicality (as predicted, for instance, by Osherson and colleagues' similarity-coverage model), the pattern of answers and the justifications provided suggest that experts often base their judgments on theories about hypothetical causal mechanisms (see also López et al. 1997). As explained in section 7, the fact that different properties of our inductive competence are best explained by theories positing different theoretical entities (viz., prototypes, exemplars, or theories) constitutes evidence for the existence of distinct kinds of concepts used in distinct processes. Strikingly, this conclusion is consistent with the emerging consensus among psychologists working on induction that people rely on several distinct induction processes (Murphy 2002; Proffitt et al. 2000; Rehder 2006; Sloman & Lagnado 2005).

A natural question raised by these findings concerns the conditions that prime the reliance on prototypes rather than exemplars and theories or on theories rather than prototypes and exemplars (and so on) in induction (see sect. 12). Because cognitive scientists have rarely fully embraced the idea that there are several distinct kinds of concepts and several processes defined over them, there is little systematic work on this question (but see Rehder 2006).

The research on categorization and concept learning, reviewed in Chapter 6, tells an even clearer story, providing evidence for the existence of prototypes, exemplars, and theories that are used in distinct categorization processes. The research on concept combination, reviewed in Chapter 7, also shows that when people produce a complex concept, they appeal to exemplars, prototypes, and theories. However, in contrast to the research on induction and categorization, it appears that a single process uses prototypes, exemplars, and theories (instead of several distinct combination processes, each of which uses a distinct kind of concepts).

## 9. Neo-empiricism

A number of cognitive scientists have recently developed a new approach to the nature of concepts (Barsalou 1999; 2008a; 2009; Barsalou et al. 2003; Gallese & Lakoff 2005; Glenberg 1997; Martin 2007; Martin & Chao 2001; Prinz 2002; 2005), which I have called "neo-empiricism" (Machery 2006c; 2007). Although there are differences between neo-empiricist theories, they all endorse the two following theses:

1. The knowledge that is stored in a concept is encoded in several perceptual and motor representational formats.

2. Conceptual processing involves essentially reenacting some perceptual and motor states and manipulating these states.

Thesis 1 is about the format of concepts: Neo-empiricists claim that conceptual knowledge is encoded in perceptual and motor representational formats. By contrast, amodal theorists contend that our conceptual knowledge is encoded in a representational format that is distinct from the perceptual and motor representational formats (Barsalou et al. 2003, p. 85). This distinct representational format is usually thought of as being language-like, although, importantly, amodal representations need not form a language (see further on). Thesis 2 concerns the nature of the cognitive processes underlying categorization, induction, deduction, analogy-making, planning, or linguistic comprehension. The central insight is that retrieving a concept from long-term memory during reasoning or categorization consists in tokening some perceptual representations, a process called *simulation* or *reenactment.* Cognitive processing consists in manipulating these reenacted percepts (e.g., Barsalou 1999, p. 578). Following Barsalou (1999), I will use the term *perceptual symbols* to refer to concepts understood in accordance with Theses 1 and 2.

Perceptual symbols might be one of the fundamental kinds of concepts, but I argue in Chapter 4 of *Doing without Concepts* that the evidence provided so far falls short of establishing this. I have identified three main shortcomings of the research in support of neo-empiricism (see also Machery 2007; for other arguments, see Dove 2009; Machery 2006c; Mahon & Caramazza 2008).

First, what I have called "Anderson's problem" in reference to Anderson's (1978) work on the controversy between imagistic and propositional theories of thinking. Neo-empiricists typically contrast the predictions made by amodal theories of concepts and the predictions made by neo-empiricist theories of concepts (e.g., Pecher et al. 2004; Solomon & Barsalou 2001; 2004; Yaxley & Zwaan 2007), and they then attempt to show that the neo-empiricist predictions, but not the amodal predictions, are verified. The problem is that there is no such thing as *the* amodal prediction of concepts; rather, different amodal theories of concepts make different predictions, depending on what they assume about the processes that use amodal concepts. In numerous cases, some amodal theories of concepts make exactly the same predictions as the neo-empiricist theories of concepts developed by cognitive scientists such as Barsalou (for some examples, see Machery 2007; 2009; Mahon & Caramazza 2008). As a result, neo-empiricist findings do not distinguish between neo-empiricism and amodal theories of concepts in general. Rather, they provide evidence against specific amodal theories of concepts, while being naturally accommodated by other amodal theories of concepts.

The second shortcoming is what I have called "the problem from imagery." Most proponents of amodal theories of concepts (e.g., Fodor 1975; Simon 1995) acknowledge that in some situations, people rely on imagery (e.g., visual imagery). For instance, we visualize our own home when we are asked how many windows it has. What proponents of amodal theories of concepts deny is that imagery is the only type of processes people have: People also have amodal concepts that are used in non-perceptual processes. The fact that proponents of amodal theories of concepts recognize the role and importance of imagery entails that when amodal theorists expect people to rely on imagery to solve a particular task, showing that people use imagery in this task fails to provide evidence for neo-empiricism and against amodal theories of concepts (for some examples, see Machery 2007).

The third shortcoming is what I have called "the generality problem." Neo-empiricists typically assume that all concepts are perceptual symbols. However, it could be that perceptual symbols constitute only a kind of concepts – a hypothesis that would naturally be consistent with the heterogeneity hypothesis. In fact, research suggests that at least some conceptual representations – namely, the representations of the magnitudes of classes of objects or sequences of sounds – are not perceptual, but amodal (Dove 2009; Machery 2007). Although these representations do not form a language and thus are different from the hypothesized representations of Fodor's (1975; 2008) language of thought, they are not perceptual either (Dove 2009; Machery 2006c). Dove (2009) has developed the generality problem in great detail, showing that the research in support of neo-empiricism has typically focused on a single kind of concepts – namely, "concrete or highly imageable concepts" (2009, p. 431) – and that neo-empiricist findings are unlikely to be found with concepts with low imageability such as abstract concepts.

Others have identified further difficulties. Reviewing a range of neuroscientific work on concepts and various important behavioral studies, Mahon and Caramazza (2008) grant that the perceptual and motor systems are often activated during conceptual processing, but they insist that this activation falls short of supporting neo-empiricism, for it can be interpreted in two different ways. First, the interpretation preferred by neo-empiricists: The brain areas involved in perceptual and motor processing or the areas near those are activated because concepts are perceptual and motor representations and perceptual representations are realized in these areas. Second, the amodal interpretation of these findings: The activation of these brain areas results from the activation of other brain areas, not involved in perceptual processing, and from this activation spreading from the latter areas to the former (a well-known phenomenon). Both interpretations account equally well for the neo-empiricist findings.

Finally, let's say a few words about the neo-empiricist research in cognitive neuroscience. A large number of fMRI studies show that tasks meant to tap into the processes underlying higher cognition (particularly, tasks involving the understanding of words) activate either the very brain areas involved in perceptual and motor processing or brain areas near those (see, e.g., Barsalou 2008a; Kiefer et al. 2007; Martin 2007; Martin & Chao 2001; Pulvermüller 2005; Simmons et al. 2007; Thompson-Schill 2003). However, in contrast to neo-empiricists' usual interpretation of these findings, I contend that much of the neuroscientific research on concepts challenges this approach. Because neo-empiricists insist that tokening a concept means tokening some perceptual representations, they are committed to the view that concept retrieval should activate our perceptual areas (Simmons et al. 2007). However, a typical finding in neuroscience is that the brain areas activated are near, *and thus not identical to*, the brain areas involved in perceptual or motor processing (a point acknowledged by Simmons et al. 2007). Furthermore, in much neo-empiricist research on concepts in neuroscience, the brain areas that are activated in the tasks meant to tap into the processes underlying higher cognition are anterior to the brains areas activated in perceptual processing (e.g., Boulenger et al. 2009; Chao & Martin 1999; Hauk et al. 2004; Kable et al. 2005; Martin et al. 1995; Pulvermüller & Hauk 2006). A plausible interpretation is that the brain areas activated in the tasks tapping into higher cognition are amodal representations, which are distinct from the perceptual representations activated in the tasks tapping into perceptual processes, but near them. To conclude, it might be that perceptual symbols are a fundamental kind of concepts, but research still fails to establish it beyond doubt.

## 10. Hybrid theories of concepts

Several hybrid theories of concepts have been developed since the 1970s, and there is a fair amount of differences between them, but they all agree on several crucial points (Anderson & Betz 2001; Keil 1989; Keil et al. 1998). Hybrid theories of concepts grant the existence of several types of bodies of knowledge, but deny that these form distinct concepts; rather, these bodies of knowledge are the parts of concepts. Like the heterogeneity hypothesis, hybrid theories of concepts typically concur that these parts store different types of information. For instance, some hybrid theories (Gelman 2004) have proposed that one part of a concept of *x* might store some statistical information about the *x*'s, while another part stores some information about specific members of the class of *x*'s, and a third part some causal, nomological, or functional information about the *x*'s. Furthermore, they often contend that the distinct parts that compose a given concept are used in different processes (e.g., Osherson & Smith 1981). For example, the parts that compose a given hybrid concept might be used in distinct categorization processes, distinct induction processes, and so on.

Although hybrid theories of concepts and the heterogeneity hypothesis agree on several points, they are far from being identical. In section 2, I proposed two individuation criteria that specify when two bodies of knowledge about *x* form two distinct concepts rather than a single concept. Hybrid theories of concepts contend both that the different coreferential bodies of knowledge are connected and that they are coordinated. The heterogeneity hypothesis assumes that at least one of these two claims is false.

Evidence tentatively suggests that prototypes, set of exemplars, and theories are not coordinated. Malt's (1994) work on how people conceptualize water shows that people have at least two distinct concepts of water – a theoretical concept of water that identifies water with any substance composed of molecules of $H_2O$ and a prototype that identifies water with any substance that shares some typical properties (origins, use, appearances). More recently, Machery and Seppälä (forthcoming) have shown that many participants (between 20% and 80%, depending on the pair of sentences considered) are willing to endorse apparently contradictory sentences of the following form:

1. In a sense, tomatoes are a fruit.
2. In a sense, tomatoes are not a fruit.
3. In a sense, whales are fish.
4. In a sense, whales are not fish.

That is, many participants say that both (1) and (2) and both (3) and (4) are true. Although there are several possible explanations of these findings, a plausible explanation is that people retrieve different concepts of tomato when they read (1) and (2). When they retrieve a theory of tomatoes, they answer that (1) is true, whereas they answer that (2) is true when they retrieve a prototype of tomatoes. This suggests that their prototype and their theory form two distinct concepts rather than two parts of the same concept.

## 11. Multi-process theories

The heterogeneity hypothesis proposes that prototypes, exemplars, and theories are often used in distinct cognitive processes (e.g., distinct categorization processes). I call "multi-process theories" those theories that contend that a given cognitive competence (e.g., categorization, induction, or the capacity to make moral judgments) is underwritten by several distinct processes. Chapter 5 of *Doing without Concepts* is dedicated to examining this kind of cognitive theory. Dual-process theories, which have been embraced in social psychology, are a type of multi-process theory, characterized by a distinction between two types of processes (slow, analytic, intentional processes and fast, automatic processes; for discussion, see, e.g., Evans 2007; Evans and Frankish 2009; Gigerenzer & Regier 1996; Sloman 1996). Gigerenzer and Todd's fast-and-frugal-heuristics research program is another kind of multi-process theory (Gigerenzer et al. 1999).

The default hypothesis in cognitive science is that a cognitive competence is underwritten by a single cognitive process, and the burden of proof typically is on those who hold a multi-process theory for some cognitive competence. In light of the recent work on a range of cognitive competences, this state of affairs should be revised. Evidence suggests that cognitive competences are commonly underwritten by several distinct processes.

Postulating that a given cognitive competence is underwritten by distinct processes raises a host of questions that have rarely been explicitly confronted by proponents of multi-process theories (but should be). The two most important issues are the following ones:

A. In what conditions are the cognitive processes underlying a given cognitive competence triggered? Are they all always triggered? Are they rather triggered in distinct circumstances? Or, perhaps, in overlapping circumstances? If they are not all always triggered, what cues or processes determine their triggering? Is their triggering under intentional control?

B. If the cognitive processes that underlie a given cognitive competence are triggered in the same conditions, how does the mind choose between their outputs or integrate them?

It is fair to say that current multi-process theories, such as the dual-process theories, have typically failed to give clear answers to these questions. This limits their capacity to genuinely predict experimental outcomes.

What about the prototype-based, exemplar-based, and theory-based cognitive processes? In what conditions are they triggered? And if they are triggered simultaneously, how does the mind choose between their outputs? There is no systematic work on these issues; in fact, I hope that this book will invite cognitive scientists to systematically investigate them.

What is known can be presented briefly. It appears that the categorization processes can be triggered simultaneously (e.g., Allen & Brooks 1991; Smith et al. 1998), but that some circumstances prime reliance on one of the categorization processes. Reasoning out loud seems to prime people to rely on a theory-based process of categorization (Smith & Sloman 1994). Categorizing objects into a class with which one has little acquaintance seems to prime people to rely on exemplars (Smith & Minda 1998). The same is true of these classes whose members appear to share few properties in common (Minda & Smith 2001; Murphy 2002; Smith & Minda 2000). Very little is known about the induction processes except for the fact that expertise seems to prime people to rely on theoretical knowledge about the classes involved (López et al. 1997; Proffitt et al. 2000).

## 12. Open questions

One of the virtues of the heterogeneity hypothesis is to bring to the fore a range of questions that have not been systematically examined by cognitive scientists. I now summarize some of these issues.

First, psychologists should investigate the factors that determine whether an element of knowledge about *x* is part of the concept of *x* rather than being part of the background knowledge about *x*. Frequency of use is the only factor that has been systematically investigated (Barsalou 1982). Other factors should be considered – including attention and explicit teaching.

Second, there are several prototype theories, several exemplar theories, and several theory theories. Although evidence indicates that we have prototypes, exemplars, and theories, it remains unclear, however, which prototype theory, exemplar theory, or theory theories is correct. That is, the exact nature of prototypes, exemplars, and theories remains to be investigated. Cognitive scientists have typically attempted to show that concepts are prototypes, or that concepts are exemplars, or that concepts are theories, but they have paid little attention to investigating the nature of prototypes, exemplars, and theories in great detail. Similarly, it is important to determine which prototype-based model of categorization (induction, etc.), which exemplar-based model of categorization (induction, etc.), and which theory-based model of categorization (induction, etc.) is correct (instead of comparing, say, a specific exemplar-based and a specific prototype-based model of categorization, as has usually been done). Recently, some psychologists have taken up the important task of comparing the models of categorization and of induction developed by prototype theories (Sloman & Lagnado 2005), as well as the models of categorization and of induction developed by theory theorists (Rehder & Kim 2006). Such efforts should be systematically pursued.

Third, multi-process theories are also another important research area that requires systematic attention. I have sketched a framework for developing multi-process theories of the higher cognitive competences, identifying several key questions that need to be answered by proponents of these theories. Multi-process theories need to deal with some important issues that have not been fully solved. Of particular importance is the kind of evidence that can support multi-process theories. Among the three kinds of evidence I have distinguished, the legitimacy of dissociations remains controversial and should be investigated further. It is also plausible that other kinds of evidence can support multi-process theories. While contemporary psychologists often endorse dual theories of cognition that distinguish System 1 and System 2 processes, there are numerous other types of multi-process theory. Furthermore, existing multi-process theories, such as dual-process theories, do not specify in which conditions the hypothesized processes are triggered and how their outputs are integrated, and, as a result, they are unable to yield clear predictions instead of mere post hoc accommodations.

Psychologists should also develop detailed multi-process theories of those cognitive competences that are the best candidates for being realized by several distinct processes – namely, categorization and induction (see sect. 11). So far, we know very little about how the distinct cognitive processes that realize competences such as categorization and induction are organized. We do not really know whether outside the lab the categorization (or induction) processes are triggered simultaneously or in distinct conditions. We do not really know what determines their triggering. And we do not know what happens to the outputs of the categorization (or induction) processes when these processes are simultaneously triggered.

## 13. Concept eliminativism

Let us take stock. We have seen that the class of concepts divides into several distinct kinds of concepts, namely, prototypes, exemplars, and theories, which have little in common. Categories, substances, events are often represented by several coreferential concepts (a prototype, a set of exemplars, and a theory). These are not parts of concepts, but are rather bona fide concepts. Prototypes, exemplars, and theories are also typically used in distinct cognitive processes, for example, in distinct categorization processes, although little is known about the organization of these processes. It is rarely the case that a given cognitive competence is underwritten by a single cognitive process; rather, the mind usually includes several distinct processes that do the same thing. Finally, the heterogeneity hypothesis focuses attention on a range of empirical questions, for which systematic empirical information is missing, partly because cognitive scientists have not considered the heterogeneity hypothesis seriously.

To conclude this article, I want to discuss a radical proposal inspired by the views about concepts presented in *Doing without Concepts* and in this précis article: Cognitive scientists might be better off renouncing the very notion of concept. Rather, they should use theoretical terms introduced to refer to the fundamental kinds of concepts – namely, *prototype, exemplar*, and *theory*. The heterogeneity hypothesis contends that the class of concepts is not a natural kind: It does not support many causally grounded generalizations because the class of concepts divides into several fundamental kinds that have little in common. Furthermore, theoretical terms are often rejected when it is found that they fail to pick out natural kinds. To illustrate, some philosophers (Murphy & Stich [1999] building on Griffiths [1997]) have proposed to eliminate the term *emotion* from the theoretical vocabulary of psychology on precisely these grounds. The proposal here is that *concept* should be eliminated from the vocabulary of cognitive science for the same reason.

Chapter 8 of *Doing without Concepts* examines the intricate and controversial logic of eliminativist arguments (see also Mallon et al. 2009; Murphy & Stich 1999; Stich 1996). Many eliminativist arguments attempt to conclude that there are no *x*'s (e.g., no beliefs: see Churchland [1981] and Stich [1983]; or no races: for discussion, see, e.g., Mallon [2006]) from the fact that the definition of *x* is not satisfied. For instance, if nothing satisfies the definition of *belief* or *concept*, it is concluded that there are no beliefs of concepts. However, such eliminativist arguments are bound to be unsuccessful because they are enmeshed with controversial issues concerning how words such as *belief* or *concept* refer (Mallon et al. 2009). What we need is another kind of eliminativist argument, which clarifies when it is legitimate to eliminate a scientific term from a scientific classificatory scheme.

In a nutshell, I propose that scientific terms should be eliminated on pragmatic grounds (this is what I have called "scientific eliminativism"): To determine whether *x* has a legitimate place in the vocabulary of a given science or whether it should be eliminated, one should examine whether using *x* helps to fulfill the goals of this science – particularly, whether it helps its classificatory purposes. Picking out natural kinds is the primary function of theoretical terms in many sciences (Quine 1969; but perhaps not in all sciences: Russell 1948). Thus, when it is found out that a scientific term fails to pick out a natural kind, there is a presumption that it should be eliminated from the relevant science. However, one still needs to consider and weigh the costs and benefits of eliminating this term. Perhaps keeping this term might simplify communication between scientists. On the other hand, keeping this term might prevent the development of a more appropriate classificatory scheme (a common situation, I suspect). If the relevant term does not pick out a natural kind and if the benefits of keeping it do not clearly overweigh the costs, then one should eliminate it.

Because *concept* does not pick out a natural kind if the heterogeneity hypothesis is correct, there is a presumption that it should be eliminated from the theoretical vocabulary of psychology. Furthermore, the continued use of *concept* in cognitive science might invite cognitive scientists to look for commonalities shared by all concepts or to develop another theory that would encompass all the phenomena known about the processes underlying higher cognition. If the heterogeneity hypothesis is correct, these efforts would be wasted. By contrast, replacing *concept* with *prototype*, *exemplar*, and *theory* would bring to the fore the urgent open questions discussed in section 12. For instance, speaking of a prototype-based categorization process, an exemplar-based categorization process, and a theory-based categorization process makes it clear that there are several categorization processes and brings to the fore the questions of the organization of these categorization processes.

Now, one might worry that eliminating the word *concept* would make communication among cognitive scientists cumbersome. To some extent, this is likely to be true, as is suggested by the frequent use of this term in this article. But I doubt that the elimination of *concept* would make communication too cumbersome; after all, when required, cognitive scientists can always appeal to the description "bodies of knowledge used in higher cognition." It seems likely that using such a description will not invite cognitive scientists (or at least not to the same extent) to look for commonalities among all bodies of knowledge used in higher cognition. If this is correct, then the costs resulting from the elimination of *concept* are limited, and cognitive scientists should eliminate the notion of concept from their theoretical apparatus.

ACKNOWLEDGMENTS
I would like to thank Guy Dove, Kevan Edwards, and Paul Bloom for helpful comments on previous drafts.

# Open Peer Commentary

## Default knowledge, time pressure, and the theory-theory of concepts

doi:10.1017/S0140525X10000324

Thomas Blanchard
*Department of Philosophy, Rutgers University, New Brunswick, NJ 08901-1107.*
**tblanchard@philosophy.rutgers.edu**

**Abstract:** I raise two issues for Machery's discussion and interpretation of the theory-theory. First, I raise an objection against Machery's claim that theory-theorists take theories to be default bodies of knowledge. Second, I argue that theory-theorists' experimental results do not support Machery's contention that default bodies of knowledge include theories used in their own proprietary kind of categorization process.

Edouard Machery (in *Doing without Concepts*, Machery 2009, p. 12) claims that psychologists (including theory-theorists) take concepts to be "bodies of knowledge that are used *by default* in the processes underlying the higher cognitive competences" (p. 11). According to Machery's gloss, default knowledge about, for example, dogs is the knowledge that is "preferentially available" (p. 11) and "spontaneously comes to mind" (p. 12) in most contexts in which we make judgments about dogs. People's knowledge about a category is not exhausted by their default knowledge, but their non-default (background) knowledge is less easily retrievable and is used only when default knowledge is insufficient for the task at hand. Machery suggests that whether or not a subject's judgment about *x* provides evidence about the content of her default body of knowledge relative to *x* depends (at least partly) on whether or not she made the judgment under time pressure. Thus, he notes in his Précis (target article, sect. 3, para. 8), the fact that under no time pressure people may retrieve a representation of a cheetah that makes the sentence "A man can outrun a cheetah" true is no evidence that this representation belongs to their default knowledge rather than to their background knowledge about cheetahs. In situations with no time pressure, people can and do retrieve information from their background knowledge.

This turns out to be crucial when we consider the nature of the theory-theorists' experiments. There is a stark contrast between the kinds of experimental tasks relied upon by, say, prototype-theorists on the one hand and theory-theorists on the other hand. Prototype theorists have usually relied upon experimental tasks in which subjects were instructed, for example, to list properties associated with a category in a short time period or make categorization judgments under explicit time pressure (see, e.g., Hampton 1979; Rosch & Mervis 1975). By contrast, many of the experimental tasks designed by theory-theorists do not involve any element of time pressure. Consider, for instance, Rips's (1989) famous pizza experiment. As Smith and Sloman (1994) have rightly noted, "there was no mention of speed in Rips' instructions" (p. 380). Or consider Keil's (1989) transformation experiments, in which subjects were asked to make judgments about the biological membership of an animal that has undergone unusual transformations. Nothing in Keil's text indicates that subjects had to make their categorization judgments under any particular time pressure.

This raises two issues for Machery's discussion of the theory-theory. First, because Rips's and Keil's experimental tasks did not involve any element of time pressure, they do not show that the knowledge retrieved by subjects to solve these tasks belongs to their "default knowledge" in Machery's sense. Subjects might have retrieved information from their background knowledge to make their categorization judgments. (Given the unusual nature of the tasks they were asked to solve, this would not be surprising). However, theory-theorists take Rips's and Keil's experiments to provide crucial insights into the nature and structure of concepts. If theory-theorists assume concepts to be default bodies of knowledge, then they have misinterpreted Rips's and Keil's results. A more charitable hypothesis is that, *pace* Machery, theory-theorists do not consider that being used by default in higher-level cognitive competences is a necessary condition for a piece of information to count as conceptual knowledge. This hypothesis is supported by some of the theory-theorists' own claims. For example, both Murphy and Medin (1985, p. 296) and Gelman (2003, p. 244) take theories to be concepts, but allow that in many contexts, subjects rely by default on prototypical features to make routine and quick categorization judgments, and retrieve theoretical information only when prototypical information is insufficient for the task at hand.

Here is the second issue. Machery (2009, pp. 52, 119) contends that for most categories, our default knowledge includes a theory of that category, and that theory-like default bodies of knowledge are associated with a distinct kind of categorization process. He mentions Rips's and Keil's results in support of this claim (pp. 183–87). However, these experiments do not support Machery's contention, since they are consistent with the claim that theories belong to background rather than to default knowledge. Moreover, the literature generated by Rips's results partly disconfirms Machery's claim. Thus, Smith and Sloman (1994, pp. 379–80) failed to replicate Rips's results when they instructed subjects to make their categorization judgments quickly. This suggests that the theoretical information on which subjects relied in Rips's original experiment did not belong to their default knowledge.

Now, some experimental results mentioned by Machery do support the idea that default bodies of knowledge include theoretical information. Thus, Luhmann et al. (2006; cited in Machery 2009, p. 186) have shown that subjects can use causal knowledge to make categorization judgments even under time pressure. In the learning phase of the experiment, subjects learned about the causal relations between various features of fictional animals. In a subsequent categorization phase, subjects relied upon this causal knowledge to make categorization judgments, even when they were instructed to respond as quickly as possible. This is evidence that their default knowledge about those fictional animals included causal (hence theoretical) knowledge. However, as Luhmann et al. argue, their results are consistent with (and according to them, even support) the idea that their subjects' theory-driven categorization process was not "qualitatively different from traditional similarity-based processes based on weighted feature matching" (2006, p. 1104): Subjects may simply have assigned strong weights to causal features when they first learned the category, and later used those features weights in a simple similarity computation when they made categorization judgments. Thus, these results do not support Machery's claim (2009, p. 119) that theory-like default bodies of knowledge are used in a categorization process similar to scientific reasoning and different in kind from the categorization processes associated with, for example, prototypes.

ACKNOWLEDGMENT
I thank Alvin Goldman for helpful advice and comments.

## Carving nature at its joints using a knife called concepts

doi:10.1017/S0140525X10000336

Justin J. Couchman, Joseph Boomer,
Mariana V. C. Coutinho, and J. David Smith
*Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY 14260.*
**jjc38@buffalo.edu jboomer@buffalo.edu**
**mvc5@buffalo.edu psysmith@buffalo.edu**

**Abstract:** That humans can categorize in different ways does not imply that there are qualitatively distinct underlying natural kinds or that the field of concepts splinters. Rather, it implies that the unitary goal of forming concepts is important enough that it receives redundant expression in cognition. Categorization science focuses on commonalities involved in concept learning. Eliminating "concept" makes this more difficult.

Categorization researchers hypothesize that minds group objects to reflect natural groupings in the world. For example, theorists often point to the family-resemblance relationships embodied by biological taxonomies. Poodles resemble each other more than they do Labradors; dogs resemble each other more than they do bears, and so forth. These similarities presumably hold in the world and in our mental lives. Of course, there is a fundamental separation between categories in the world and concepts in the mind. Fortunately, though, categorization science finds its grounding and coherence at both the ecological (world) and psychological (mind) levels of analysis. Unfortunately, this often leads to the misconception that "category" and "concept" are equivalent or interchangeable terms – as Machery correctly points out in *Doing without Concepts* (Machery 2009, pp. 8–14).

The world contains coherent assemblages of objects that are sometimes called natural kinds, family-resemblance categories, or consequential stimulus regions. All organisms share the profoundly important goal of learning the concepts that summarize these assemblages. These concepts are behavioral and psychological equivalence classes – they let creatures behave similarly toward similar things and understand similar things similarly. Concepts are the mind's tool for representing and responding to natural kinds.

The conceptual goal of living organisms is unitary – to develop psychological representational schemes that preserve these assemblages in acting and understanding. The fitness consequence of not doing so is unitary. The assemblages of things in the world – the natural kinds – may have a unitary structure, too. One sees that categorization science is motivated by this unitariness. That concepts are a staff of life for living organisms is a profound truth that would be lost by breaking the field into process splinters. This is a principal reason for preserving the coherence of the field.

Now it is true that minds lack direct access to the world's things in themselves. Therefore, minds must create mental concepts that estimate natural-kind categories. Given this necessary distinction between natural-kinds in the world and our mental estimates of them, it is not clear why concepts, prototypes, exemplars, theories, or any other mental representation would be considered a natural kind. Indeed, if there are any groups that are *not* natural kinds, mental representations are likely among them.

Machery (2009, p. 241) contends that because it does not pick out a unified natural kind, "concept" is a useless, or perhaps even harmful, term. We disagree. It does not shatter the field of concepts if it turns out – as it does – that concept estimation produces a marvelous variety of concepts and learning processes – including prototypes, exemplar memories, category rules, decision bounds, and theories. This only means that the phenomenon of concepts is rich and diverse. It also means that the unitary goal of estimating concepts is central enough to survival that it deserves and receives redundant expression within cognitive systems. To us, this centrality emphasizes the need for a coherent field of concepts and categorization.

Of course, the field sometimes seems less than coherent because researchers tailor category tasks to elicit different processes. Prototype theorists use large stimulus sets that reduce stimulus repetition and defeat exemplar memorization. These categories also have family-resemblance organization that encourages prototype formation. Exemplar theorists often use categories that share little or no family resemblance. Perhaps the most iconic of these are "5–4" categories. As Machery (2009, pp. 175–77; see also Smith & Minda 2000) correctly notes, these categories share few features, preventing prototype formation, and the stimulus sets are small. Participants see many repetitions of each stimulus and naturally memorize them. Rule-based tasks are clearly simply unidimensional in character, encouraging hypothesis testing, reasoning, and solution by rule formation.

However, the progress in this area shows the method in the madness. Researchers have found that the dominant categorization process is strongly affected by the size of categories, their perceptual coherence, the dimensionality of the task's sufficient solution, the stage of category learning, the default tendencies of the categorizing organism, and so forth (Ashby & Maddox 2005; Blair & Homa 2003; Couchman et al., in press; Homa et al. 1981; Murphy 2002; Smith & Minda 1998; 2000). Moreover, it is clear that these generalizations are not just isolated findings. It is insufficient to claim that the effect of stimuli on categorization is captured completely by a description of prototypes, exemplars, or theories (as Machery [2009, sect. 8.3.2] requires). Rather, organisms tune sensitively to the affordances of their category experience in order to choose the most cognitively economical and adaptive learning solution they can. Viewed from a broader perspective, these different strategies are instances of organisms navigating upon a multi-dimensional fitness surface. But the goal, the navigation, and the surface all embody a unitary fitness potential that must not be overlooked and that really is not divisible.

This broader perspective also raises many important theoretical questions. What species have which categorization potentialities and which constraints? What were the antecedents of humans' categorization system in the vertebrates or primates? What are the affordances of language and symbolic reasoning, and how do they change the nature of concept formation? What are the developmental stages by which humans acquire their mature concepts and adaptive action patterns in the world? What are the implications of cognitive development, cognitive aging, and neurological impairments for category learning, both regarding constraints and regarding preserved or spared capacities on which remediation could capitalize to maximize education and training? Replacing "concept" with "prototype," "exemplar," or "theory" as Machery (2009, p. 242) suggests would unwarrantedly deemphasize these important questions. In our view, answering all these questions is also best served by understanding the unitary nature of the task at hand.

That task is to carve nature at its joints using the psychological knife called concepts. It is true, it is profoundly important to know, and it is all right for the progress of science that the knife is Swiss-Army issue with multiple blades.

## Not different kinds, just special cases

doi:10.1017/S0140525X1000052X

David Danks
*Carnegie Mellon University and Institute for Human & Machine Cognition, Department of Philosophy, Pittsburgh, PA 15213.*
**ddanks@cmu.edu**
**http://www.hss.cmu.edu/philosophy/faculty-danks.php**

**Abstract:** Machery's Heterogeneity Hypothesis depends on his argument that no theory of concepts can account for all the extant reliable categorization data. I argue that a single theoretical framework based on graphical models can explain all of the behavioral data to which this argument refers. These different theories of concepts thus (arguably) correspond to different special cases, not different kinds.

One of Machery's central arguments for his Heterogeneity Hypothesis in *Doing without Concepts* (Machery 2009) is that, for each different theory of concepts, there are reliable datasets – behavioral, neuroscientific, and dissociation – that are best (or only) explained by that theory. That is, prototype-based, exemplar-based, and theory-based kinds of concepts are *all* required in order to explain all of the data. Thus, Machery concludes, we have heterogeneity: for many of the cognitive items thought to be single concepts, people have cognitively distinct prototype-like, exemplar-like, and theory-like kinds of concepts.

This argument requires that these three theories/kinds of concepts be qualitatively different, not special cases of a more general framework. If there were a unifying account that encompassed all three theories of concepts, then that theory would (by hypothesis) be able to explain all the same data as the heterogeneity hypothesis that Machery proposes in his book. Hybrid theories of concepts are the most prominent instances of such a unifying account, and Machery provides numerous arguments against them precisely because they have the potential to undermine his data-based argument for the heterogeneity hypothesis. Hybrid theories are not, however, the only way to unify the three theories of concepts.

The behavioral data used in Machery's argument can all, I contend, be explained by a single theory of concepts based on graphical models. At a (very) high level, a graphical model encodes a set of relationships, whether informational/probabilistic, causal, communication, taxonomic, or other. More formally, a graphical model has two components: (1) a graph composed of nodes/vertices and edges (directed and/or undirected) that encode the qualitative relationships; and (2) a representation of the quantitative relationships. The precise quantitative component depends on model-specific features; joint probability distributions and structural/linear equations are typical. Bayesian networks (causal or probabilistic), structural equation models, and Markov random fields/networks are probably the most common graphical models, though there are many other types. (Lauritzen [1996] provides a comprehensive formal overview of graphical models.)

One instance of the theory-theory (causal model theory) is already explicitly formulated in the language of graphical models (specifically, causal Bayesian networks). A deeper connection is suggested by the fact that all three theories of concepts ultimately understand concepts as structured relations among features, components, causes and effects, and so on, though there are between-theory differences in the nature of the objects and relations.

The suggested deep connection exists: All three theories of concepts can provably be represented in the graphical models framework (Danks 2004; 2007). More precisely, for each particular prototype-based, exemplar-based, or theory-based[1] theory of concepts, there is a corresponding class of graphical models such that (A) there is a one-to-one mapping between (i) a concept in the "traditional" psychological theory, and (ii) a particular graphical model in that class; and (B) inference/reasoning using that graphical model is behaviorally indistinguishable from using the corresponding prototype-based, exemplar-based, or theory-based concept. As a concrete example, multiplicative prototype concepts with second-order features are isomorphic to Markov random fields with restricted clique potential functions. (Danks [2004; 2007] proved this result for categorization; the results have since been extended to all of the activities that Machery discusses.) These three "kinds" of concepts are (at least, formally) each representable as a different set of restrictions within the shared graphical models framework, specifically the framework of so-called chain graphs.

These formal results open the door for *homogeneity* to reemerge as a live possibility. Of course, this homogeneity occurs at the level of graphical models, rather than exemplars versus prototypes versus theories. An individual can have what appear to be both prototype-based concepts and exemplar-based concepts simply by having different specific graphical models. The force of the argument for heterogeneity is thus blunted: There is a single theory that can account for all of the (behavioral) data. Importantly, this proposed theory is *not* a hybrid theory: It does not hold that a single concept is composed of different graphical models corresponding to the different types of concepts. Rather, the proposal is that any particular concept corresponds to only one graphical model, and diversity in our graphical models leads to *apparent* diversity in types of concepts.

This account focuses on the formal/computational structure of the different theories of concepts, and one might object that it ignores other, more "metaphysical" claims made by their proponents. This concern goes to the hard question of how to define or characterize a theory. At one extreme, we could define a theory (of concepts) by its behavioral predictions, and ignore all other assertions made about the theory. At the other extreme, we could say that a theory is given by a maximal set of consistent claims made by proponents of that theory, regardless of the nature of those claims. The graphical models account takes the intermediate view that a theory should be interpreted relatively minimally as the set of claims necessary to explain the phenomena that purportedly fall within its reach (similar in spirit to the approach in Strevens 2000). I contend that (though I do not have room to argue for) the proper "minimal" understanding of theories of concepts is in terms of the formal/computational structure that they attribute to particular concepts. Although there is a clear rhetorical difference between saying a concept is based on a "prototype" versus an "exemplar," I contend that the difference matters only when it leads to a difference in the concept's formal/computational structure.

Machery provides an admirable survey of the vast behavioral/psychological literature on concepts, and his conclusion – there are at least three distinct *kinds* of concepts – is reasonable if there is no unifying framework for those "kinds." But all of the behavioral data that Machery discusses can in fact be explained as bodies of knowledge encoded as graphical models: The differences in behavior can be explained as different graphical structures – that is, different special cases – rather than truly distinct kinds.

ACKNOWLEDGMENT
The author is partially supported by a James S. McDonnell Foundation Scholar Award.

NOTE

1. Statements about "the" theory-theory of concepts apply only to the few, computationally precise instances of that idea (e.g., causal model theory).

## An additional heterogeneity hypothesis

doi:10.1017/S0140525X10000348

Guy Dove
*Department of Philosophy, University of Louisville, Louisville, KY 40297.*
**guy.dove@louisville.edu**

**Abstract:** In this commentary, I make three points concerning Machery's response to neo-empiricism. First, his methodological critique fails to remove the threat that neo-empiricism poses to his conceptual eliminativism. Second, evidence suggests that there are multiple semantic codes, some of which are not perceptually based. Third, this representational heterogeneity thwarts neo-empiricism but also raises questions with respect to how we should "do without concepts."

Machery (2009) provides a substantial and convincing argument that prototypes, exemplars, and theories form distinct mental kinds. The real question is whether or not this heterogeneity should cause us to eliminate concepts as a mental kind. Neo-empiricism (e.g., Barsalou 1999; Glenberg 1997; Prinz 2002) poses a direct threat to this proposal because it provides a unified account of concepts. As Machery notes, neo-empiricism involves two core claims: that conceptual knowledge is encoded in sensorimotor representations and that conceptual processing involves some kind of perceptual simulation. If these claims apply to prototypes, exemplars, and theories, then Machery's case for eliminativism falls apart. This is not simply a hypothetical argument; at least one prominent neo-empiricist (Prinz 2002) endorses both the independence of prototypes, exemplars, and theories and the theoretical unity of concepts.

In *Doing without Concepts*, Machery claims that neo-empiricism faces two main problems. The first is that certain amodal models (i.e., ones containing non-perceptual representations) are compatible with the empirical findings cited in support of neo-empiricism. The second is that many amodal theorists acknowledge that perceptual imagery is important to some cognitive processes. Machery (2009, p. 116) concludes that "there is no strong evidence that concepts (or some concepts) are in fact similar to perceptual representations." This argument is inconclusive, though, because it fails to provide strong evidence *against* perceptually based conceptual representations or, indeed, *for* amodal ones.

In his Précis, Machery identifies an additional problem for neo-empiricism, which he refers to as the generality problem. This problem arises because there are robust bodies of evidence suggesting that some specialized conceptual representations are amodal. Although the generality problem clearly undermines strong forms of neo-empiricism, it is compatible with weaker forms (Machery 2007; Weiskopf 2007).

I suggest that neo-empiricism faces a more serious problem. This problem is not merely a lack of generality but rather the presence of heterogeneity. Evidence from a number of sources suggests that conceptual knowledge is encoded in multiple representational formats, some of which are not perceptual (Dove 2009). In other words, it supports what might be called the *representational heterogeneity hypothesis* (RHH).

The case for the RHH can be seen most clearly in recent neuropsychological and neuroimaging research on imageability. Traditionally, cognitive scientists examined imageability in terms of processing advantages for high-imageable concepts over low-imageable ones in several cognitive tasks (Paivio 1987; Wattenmaker & Shoben 1987). Two major theories emerged in response to the behavioral evidence: the dual-code theory (Paivio 1987) and the context-availability theory (Schwanenflugel & Shoben 1983). Recent evidence from cognitive neuroscience supports both to some degree (Dove 2009).

Because my concern is the RHH, I will focus on the evidence for distinct representational formats. First, consider neuropsychological case studies. Several research teams describe aphasic patients with significant left hemisphere damage who exhibit a selective semantic impairment for high-imageable words (Berndt et al. 2002; Bird et al. 2003; Crepaldi et al. 2006). Patients with a selective semantic impairment for low-imageable words have also been found (Marshall et al. 1996).

Second, a number of event-related potential (ERP) experiments support a neuroanatomical distinction between concepts of high and low imageability. For example, Holcomb et al. (1999) created a task that involved manipulations of both context and concreteness. ERP recordings were time-locked to sentence final words in a word-by-word reading task in which participants made semantic congruency judgments (e.g., Armed robbery implies that the thief used a *weapon* vs. Armed robbery implies that the thief used a *rose*). They found that sentence-final concrete words generated a larger and more anterior N400 than sentence-final abstract words in both contexts. Further studies have found context-independent topographic effects associated with imageability in single-word presentations (Kellenbach et al. 2002; Swaab et al. 2002). Thus, ERP studies employing diverse tasks support the notion that different cognitive systems are associated with the semantic processing of high- and low-imageable words.

Third, the idea that neural activity is modulated by imageability is generally supported by neuroimaging data. A number of studies find that high-imageable words elicit greater activation than low-imageable words in superior regions of the left temporal lobe (Binder et al. 2005; Giesbrecht et al. 2004; Mellet et al. 1998; Noppeney & Price 2004; Sabsevitz et al. 2005) and inferior regions of the left prefrontal cortex (Binder et al. 2005; Giesbrecht et al. 2004; Goldberg et al. 2006; Noppeney & Price 2004; Sabsevitz et al. 2005). Giesbrecht et al. (2004), for example, manipulated both imageability and semantic priming (a measure of the influence of context) in an event-related fMRI study. Participants were presented with prime word followed by a target word. The words were either semantically related (*bread* and *butter*) or unrelated (*wheat* and *slipper*). Half of the pairs consisted of two high-imageable words, and half of the pairs consisted of two low-imageable words. In keeping with the general finding that context effects are distinct from imageability effects, each of these manipulations modulated activity in anatomically distinct areas of the left hemisphere.

The RHH seems to be grist for Machery's mill. It certainly undermines the threat posed by neo-empiricism to unify concepts. In keeping with this, Piccinini and Scott (2006) have argued that the divide between cognitive processes that require language and those that do not provides a better case for conceptual eliminativism than the heterogeneity identified by Machery.

However, the RHH also raises an interesting problem. Because there is every reason to suppose that prototypes, exemplars, and theories have high- and low-imageable variants, these kinds are likely to be representationally heterogeneous. Following Machery's reasoning, this should threaten their status as theoretical kinds. It seems reasonable to resist this inference, but then the challenge is to provide an argument for resisting eliminativism in the context of prototypes, exemplars, and theories that does not apply to the larger context of concepts. Whether this can be accomplished or not remains to be seen.

## Unity amidst heterogeneity in theories of concepts

doi:10.1017/S0140525X10000543

Kevan Edwards
*Philosophy Department, Syracuse University, Syracuse, NY 13210.*
**kedwar02@syr.edu**
**http://philosophy.syr.edu/FacEdwards.htm**

**Abstract:** This commentary raises two concerns with Machery's approach in Doing without Concepts. The first concern is that it may be possible to preserve a unified theory of concepts by distinguishing facts about concept individuation from facts about cognitive structures and processes. The second concern questions the sharpness of the distinction Machery draws between psychological and philosophical conceptions of concepts.

I think Machery is entirely right, in *Doing without Concepts* (2009), to draw attention to the phenomenon that he refers to as the *heterogeneity* of concepts (as the term "concept" is used in cognitive psychology). I also agree that this stands in the way of any unified account of concepts as exemplars, prototypes, or constituents of theories. Nevertheless, I take issue with Machery's *eliminativist* conclusion. The logical heart of my worry is that heterogeneity across exemplars, prototypes, theories, and so on, precludes a unified theory of concepts only insofar as one assumes that such a theory needs to be built out of the kinds of structures and processes described by advocates of these approaches. I think this assumption is worth questioning, and I want to suggest an alternative approach that rejects it.

The alternative begins with the idea that concepts are individuated by considerations more abstract than their role in cognitive structures and processes. The specific version of this idea that I favor individuates concept types in terms of the individuals, kinds, or properties they *represent* or *refer to* (see Edwards 2009; 2010). Before saying more about this approach, I want to mention another issue that might partially explain why Machery is relatively blind to – or perhaps it is more charitable to say, disinterested in – this kind of alternative.

When it comes down to how to construe the starting point for a theory of concepts, Machery (2009, pp. 32–51) is quick to side with psychologists as opposed to philosophers. Machery chooses sides in part to undermine various philosophical concerns on the grounds that philosophers are engaged in a fundamentally different research project. I think Machery may be overlooking some potentially relevant issues here. Just to be clear, I agree that philosophers typically begin from a different starting point than psychologists and tend to emphasize different considerations. However, this is always the case with topics located at the intersection of academic disciplines. Admittedly, this realization does not show that philosophers and psychologists are converging on the same theoretical entity in this particular case. Nevertheless, I think one ought to be very cautious in concluding that different apparent starting points amount to a fundamental difference in subject matter. Given space constraints, I can't properly address Machery's claims on this subject; but I'll briefly mention several issues regarding which philosophers may be able to contribute to how psychologists think about concepts.

One topic to highlight – not surprisingly – is the notion of *representation*. It seems to me this is a place where philosophers and psychologists tend to take recognizably different approaches. Psychologists, even those who clearly are working within the framework of a Representational Theory of Mind, often appear to presume something in the vicinity of representational content without making this explicit. Consider, for example, how natural it is to talk about exemplars or prototypes being *of* or *about* a category. In contrast, philosophers, in particular philosophers of mind, have been something close to obsessed with either explaining such talk in naturalistically respectable terms or explaining how to do without it. Examples of the former include Dretske (1981), Fodor (1987; 1992), Millikan (1984; 1993); examples of the latter include Field (2001), Horwich (1998a; 1998b). I think Machery's neglect of these discussions partially explains why a representation-based view of concepts isn't on his radar. I should note that I am here echoing concerns raised by Rey (2009b).

Another topic much discussed in recent philosophy can be used to shed light on the potential relationship between representation and more blatantly psychological considerations: I have in mind the debate between so-called reductionists and non-reductionists, especially insofar as this has involved questions about *multiple-realization* (see, in particular, the debate between Fodor (1974; 1997) and Kim (1992; 1998). Speaking very roughly, the upshot of this debate has been that both anti-reductionists and their opponents have confronted the fact that reality exists (or is properly describable/conceptualized) at different levels of abstraction. At a minimum, this forces one to be careful about the *level* at which a particular theoretical entity is construed. The cost of picking too low a level of analysis is that one wrongly identifies the target entity with a disjunction of lower-level entities each of which is better described as a *realization* (or instance) of the target. To repeat a now hackneyed example, it is a mistake to identify doorstops with the arrangements of physical stuff (slices of wood, hinged pieces of metal, bricks, old printers, etc.) out of which particular doorstops are made. The now standard functionalist line is that something is a doorstop by virtue of playing a characteristic role in a larger system; for example, a system involving doors, people, and so on. Someone who holds a representational view of concepts can make a similar suggestion. Type-individuating concepts in terms of their representational properties frees up the possibility for different instances or realizations of a concept type to show up in substantially different cognitive structures and processes.

The alternative approach to which I have gestured brings many questions in its wake. I have not, for example, supplied arguments for the conclusion that a representationalist view of concepts should *supplant* rather than *supplement* accounts that appeal to mental structures and processes. Similarly, I have not said anything about whether embracing a representationalist alternative forces a substantive (rather than merely terminological) shift in work done by advocates of exemplar, prototype, and theory-based approaches. Moreover, there are familiar reasons to worry about a representation-based approach. The goal of the present commentary is to make two relatively modest points: (1) There is an admittedly radical approach to concepts that needs to be undermined before one opts for eliminativism. (2) Various issues that have featured prominently in recent philosophy have the potential to impinge upon discussions of concepts in psychology, and to do so in potentially important ways.

## Two uneliminated uses for "concepts": Hybrids and guides for inquiry

doi:10.1017/S0140525X1000035X

Chad Gonnerman and Jonathan M. Weinberg
*Department of Philosophy, Indiana University, Bloomington, Indiana 47405.*
**cgonnerm@indiana.edu jmweinbe@indiana.edu**
**http://www.indiana.edu/~eel**

**Abstract:** Machery's case against hybrids rests on a principle that is too strong, even by his own lights. And there are likely important generalizations to be made about hybrids, if they do exist. Moreover, even if there were no important generalizations about concepts themselves, the term picks out an important class of entities and should be retained to help guide inquiry.

We concur with Machery's broad assessment in *Doing without Concepts* (Machery 2009) that the science of concepts has revealed matters to be rather a mess. But we strongly disagree with what Machery takes as the philosophical upshot of that messiness, as expressed in his book's title: that psychologists should scrub "concept" from their lexicon. First, the hybrid option is still live, and if concepts are hybrids, they may well be natural kinds. Second, even if "concept" fails to pick out a natural kind, it may yet be scientifically useful, even on Machery's own terms.

Machery correctly observes that hybrid theorists owe an account of when two bodies of knowledge are parts of the same concept. He proposes that psychologists are committed to the principle that they are genuine parts only if they are *coordinated* (Machery 2009, p. 64), that is, if they generate consistent evaluations of cases. But he adduces some evidence that coordination often fails across parts of putative hybrids: for example, one representation says tomatoes are vegetables; a bit later, the other says they are not (p. 72).

But this coordination condition is too strong, as it would break up bodies of knowledge that psychologists treat as conceptual wholes, namely, sets of exemplars. Contextual shifts, such as those that adjust the perspective of the subject (e.g., from artist to biologist), can, within selective-retrieval models, shift which exemplar is retrieved, even when both exemplars are stored as being about the same thing – and thereby can shift, for particular cases, whether they will be counted as in the target category (see Braisby 2005). Sets of similarly stored exemplars would thus fail Machery's coordination condition – forcing their dissolution, likely down to individual exemplars. But that's not how psychologists treat such exemplars. His argument against hybrids proves too much.

That the coordination condition is too strong does not remove the need for hybrid theorists to provide an account of conceptual wholes. Here is one proposal, building on machinery used much earlier in Machery's book, when he attempts to distinguish conceptual from background knowledge (pp. 11–12). There, he suggests that a body of knowledge about *x* is in the concept of *x* just in case it is "preferentially available when we think, reason, and so on, about *x*" (p. 11). There is nothing unique to a hybrid that keeps its fans from saying such things about its sub-conceptual parts. A hybrid theorist could therefore propose that two coreferential representations are parts of the same concept of *x* just in case each is preferentially available in such a way when we think, reason, and so on, about *x*. (We note that Machery's account of conceptualization may require some technical tweak for Frege cases; we expect that whatever would do that for his account, would also work for this proposal.) But a failure of coordination is irrelevant to this weaker condition.

Whether this proposal succeeds is an open question, and that is sufficient for our purposes here in contending that hybrids are still a live option. For, if "concepts" picks out hybrids, psychology should probably keep the term, as there will likely be informative generalizations to be made. As Machery argues, empirical research suggests that for most categories we have a prototype, a body of exemplars, and a theory, but not a definition. So, a hybrid theorist would be able to theorize that most concepts partially consist of a prototype, and so on, but not a definition. But we cannot read such generalizations off of Machery's reference-fixing description (p. 12); they are discoveries. So, if concepts are hybrids, "concept" probably picks out a class that supports many informative, useful generalizations.

Even if the above is incorrect about hybrids, "concept" would still have important work to do. As Machery suggests, scientists should keep a term if it plays a useful role (p. 239). "Concept" does that, by efficiently marking out a class that scientists want to make claims or ask questions about, even if it should turn that there are vanishingly few generalizations to be made about something merely in virtue of its being a concept. There are many terms in good scientific standing across various disciplines that play such roles, including, we would suggest, "algorithm," "sub-atomic particle," or "nutrient" (and even some specific nutrients, such as "vitamin B"; see Elder 1994, p. 259.) None of these categories seems to support many generalizations about their members *qua* members, yet each is very useful in organizing established knowledge and continuing inquiry in their respective domains. And the same holds for "concept." For example, the neo-empiricist, friendly to Machery's general take on our conceptual systems, might want to defend the substantive claim that perceptual symbols are a kind of concept. Such a claim would be substantive, to the effect that the delineated class contains some additional sorts of entities. Indeed, Machery himself wonders whether there are other kinds of concepts (p. 249). For example, he writes, "Evidence shows that people have some knowledge about ideals. What is now needed is to determine whether these bodies of knowledge qualify as concepts" (p. 249). This strikes us as a meaningful and important question, and one for which the term "concept" is obviously useful in asking.

So, independent of the question of hybrids, psychologists should keep the term. Even if Machery is right, and concepts are not a natural kind, the potential dangers here would be better addressed through reformation instead of elimination. The practical advice to take away from Machery's arguments may be, not that scientists should get rid of "concept," but that they should be more careful in understanding that this term likely fails to pick out a very tidy sort of natural kind. Doing so should allow them to steer around the sorts of dangers that Machery (2009) hypothesizes about (e.g., pp. 242–243), without sinking an otherwise fruitful vehicle of inquiry. The psychology of categorization, inference, and so on may be much messier than philosophers and psychologists have hoped. But "concept" is still likely to be a vitally important word for theorizing about that mess.

## Concept talk cannot be avoided

doi:10.1017/S0140525X10000361

James A. Hampton
*Department of Psychology, City University London, London, EC1V OHB, United Kingdom.*
**hampton@city.ac.uk**
**www.staff.city.ac.uk/hampton**

**Abstract:** Distinct systems for representing concepts as prototypes, exemplars, and theories are closely integrated in the mind, and the notion of concept is required as a framework for exploring this integration. Eliminating the term "concept" from our theories will hinder rather than promote scientific progress.

While most people interested in concepts will find much to agree with in this book (*Doing without Concepts*, Machery 2009), it is the eliminativist thesis that will find most resistance. Machery provides analogical cases in psychology such as "emotion" and "memory." Emotion and memory, it is argued, may prove to be terms referring to a varied set of phenomena, without any identifiable single associated brain system. Similar cases can be found in other sciences – for example, "species" and "planet." The concept of species is problematic because there is not always a clear criterion for differentiating one species from another; instead biological laws describe the distribution of genes over populations of individuals (Mayr 1982). While problems of definition mean that "species" is not a well-defined term in biology, it would, however, be hard to imagine biological discourse without it. There are just too many general truths that need to be expressed. Similarly, astronomers ran into trouble with the designation of Pluto as a planet, given the discovery of other large orbiting bodies that had been labeled as asteroids. But the term still has a referential meaning. Science needs more loosely defined general referring expressions in addition to the carefully defined terms that figure in theories. I argue that cognitive science still needs the notion of "concept," even if it proves multifaceted and hard to define satisfactorily.

Machery's argument rests on there being three distinct forms of knowledge that are recruited by default by cognitive processes: namely, prototypes (P), exemplars (E), and theories (T). The danger of eliminating the notion of concept is that the importance of the relations between these forms of knowledge risks being underplayed. First, there is the obvious point that the P, E, and T representations (let's call them PET) of dog all refer to the same class – they are broadly co-referential (give or take some differences in categorization resulting from exceptional contexts). What makes them co-referential is the fact that they represent the same concept. Without a notion of concept, it is hard to explain why they co-refer.

More importantly, the term "concept" is needed as part of an account of the many situations in which the PET systems interact. How does one discuss concept combination, including the formation of composite prototypes, the importing of exemplar knowledge, and the coherence checking of the result through background theory, if one cannot have the integrative term "concept" to specify just what it is that is being combined. The combination occurs at the concept level, and the description of the processes involved then requires elaboration in terms of the PET systems. Similarly, in concept learning, we need an overarching notion of concept in order to describe how PET systems interact. Experiential concepts like DOG or CUP may first be learned by a child through interacting with individuals encountered in everyday life. When a variety of individuals are known, and it is necessary to learn to use the words "dog" and "cup" correctly, then prototypes may be formed, enabling generalization to other individuals, discrimination of other classes, and the accumulation of generic knowledge. As the child then develops wider knowledge, the prototype notion of DOG may be supplemented by theoretically driven concepts like mammal or species, and by essentialist ideas about the causal properties of biological kinds, or the need to defer to expert opinion about correct classification.

Far from aiding scientific advance, treating the PET systems as largely independent of each other may impede investigation of the important ways in which information is transferred between them. It can also be argued that the three systems are not as easily distinguished as Machery would require. Consider prototypes and exemplars. Machery agrees that much of the research and debate concerning prototypes and exemplars has been directed at a very restricted form of behavior, namely, learning to classify simple geometrical shapes in a laboratory setting where the categories to be learned are not easily distinguished without extensive training. Even in this arcane area of psychology, there is considerable evidence that under different conditions people will either learn individual exemplars or will abstract prototypes (Smith & Minda 1998). If we move to the more "conceptual" domain of natural language terms, then the question of prototype versus exemplar models hardly arises. For example, Storms et al. (2000) have investigated whether typicality in superordinate categories like FISH, FRUIT, or FURNITURE is best predicted by similarity to the category prototype or by similarity to "exemplars." But in this case the exemplars are simply prototypes defined at a more specific level (e.g., CHAIR and TABLE). So the question is not which of two distinct systems is driving the behavior, but rather which level of abstraction is involved within a single representational system. Some concepts do have genuine exemplars – the concept of "Beethoven Symphony" to a musician will be heavily dependent on knowledge of the nine exemplars. But there will be a close link between knowledge of the exemplars and generalized knowledge about the typical structure and expressive vocabulary found in the works.

Likewise, there has been a rapprochement between prototype and theory-based elements of concepts. In discussing the notion of prototype (Hampton 1998), I have proposed that the distinguishing feature of prototype representations is that they represent the center of a class and not its boundary. It is this fact that gives rise to category vagueness, typicality gradients, the lack of explicit definitions, and the preponderance of generic (rather than necessary) features in people's accounts of the content of the concept. The notion of prototype as a form of schema is therefore free to be supplemented by causal connections within the representation resulting in a structured frame representation (Barsalou & Hale 1993). Mutability and centrality of properties, modal judgments of necessity, and dissociations between similarity-based typicality and theory-based categorization can all be accommodated within this single representational system.

In short, it is too soon to be counseling despair about integrating prototype, exemplar, and theory-based representations into a coherent account of the concept of concept.

## Eliminating the "concept" concept

doi:10.1017/S0140525X10000567

Stevan Harnad

*Institut des Sciences Cognitives, Université du Québec à Montréal, Montreal, Canada, H3C 3P8; School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, United Kingdom.*

**harnad@ugam.ca**

**http://users.ecs.soton.ac.uk/harnad**

**Abstract:** Machery suggests that the concept of "concept" is too heterogeneous to serve as a "natural kind" for scientific explanation, so cognitive science should do without concepts. I second the suggestion and propose substituting, in place of concepts, inborn and acquired sensorimotor category-detectors and category-names combined into propositions that define and describe further categories.

Whatever a "concept" is, we have at least one for every thing we can recognize, act on, name, or describe, including not only the things denoted by all the dictionary words we understand, but also everything we know what to *do* with (Harnad 2007), even if we don't know its name or it has none – perhaps because, like "things that are bigger than a breadbox," no one has ever bothered to name it.

"Things" can be individual objects (nonliving or living), kinds, events, actions, properties, or states. We have "concepts" of countless such things, and having the concept simply means being able to *do* something with respect to those things, an action that has a right and wrong about it – anything from approaching/avoiding the thing, to interacting with or manipulating it in some way, identifying it (correctly) by name, saying true things about it, imagining it, and thinking and reasoning about it.

In *Doing without Concepts*, Machery (2009) suggests that although there is no "natural kind" corresponding to the intersection of prototypes, examples, theories, and sensorimotor representations, each may still turn out to be a legitimate natural kind of its own. I will sketch an alternative that scraps both the use and the mention of "concept" altogether.

Consider concept's twin, "percept." If a concept is, roughly, an "idea," then a "percept" is an "image." Should we ban talk of percepts, too? Pylyshyn (1973) suggested banning talk of "images" – as unobservable, unmeasurable, homuncular, and, most important, nonexplanatory – to be replaced by propositions, and, eventually, computations, which are genuinely explanatory, in that they can *generate* the capacity that the images or "percepts" had been meant to explain (Harnad 2006).

With findings on mental rotation (Shepard & Cooper 1982), however, "percept" has made a comeback, in the form of internal analog structures and processes that have some of the properties of images but can do the internal generative work, with no homunculus, sometimes more efficiently than computation. (Digital computation can always approximate analog dynamics as closely as we like: A picture is always worth more than 1,000 words, but 10,000 words come closer. It cannot, however, be words all the way down; Harnad 1990.)

Apart from their sensory shapes, objects have sensorimotor "affordances": things that objects are amenable to having done with them (by our bodies, and their shapes). A chair (but not a pyramid or a pincushion) affords sittability-upon; a doornail, but not a doormat, affords grasping and turning. But is an affordance-detector a "representation"?

We need to be able to recognize birds, for example, before we can start doing anything with them, including talking and thinking about them. No machine vision program could perform anywhere near human level using prototype-matching to recognize birds; raw example-storage would do even worse. And without those, verbal theories could not even get off the ground (because it can't be words all the way down).

So what we need first is not bird representations, but *bird-detectors.* For most of us, visuomotor contact is our first introduction to birds, but it is not "we" who pick up the affordances; we are no more aware of the tuning of our internal category detectors than subjects in mental-rotation experiments are aware of rotating their inner images. Internal mechanisms do this "neo-empirical" work for us (Barsalou 1999; Glenberg & Robertson 2000). The work of cognitive science is to discover those mechanisms. That done, it no longer matters whether we call them concepts, ideas, notions, representations, beliefs, or meanings.

Cognitive science has not yet done this job, though Turing (1950) set the agenda long ago: Scale up to a model capable of *doing* everything we can do (Harnad 2008). The first hurdle is sensorimotor category detection: the mechanism for learning categories from sensorimotor interactions with the world, guided by error-correcting feedback. We share this capability with most other species: learning to detect and act upon sensorimotor affordances. *To categorize is to do the right thing with the right kind of thing* (Harnad 2005).

Some categories are innate: We recognize and know what to do with them because natural selection already did the "learning" by genetically pretuning our ancestors' brains. But most categories we have to learn within our lifetimes, including everything named and described in our dictionaries plus many things, actions, events, properties, and states we never bother to name: We learn to do the right thing with them, and perhaps describe them, on the fly. How did we *get* those names and descriptions? Our species is the only one that has them.

According to our account so far, we only have the categories for which we have learned through direct experience what to do with their members. One of the most adaptive things our species alone does with many of our categories is to *name* them. For, with language evolved our capacity to produce and understand strings of category names that encode truth-valued propositions, *predicating* something about something. This allowed us to acquire new categories not only by sensorimotor *induction*, but also by verbal *instruction.* For once we have a set of categories "grounded" directly in our sensorimotor capacity to detect their members and nonmembers, we can also assign each category an agreed, arbitrary name (Harnad 1990), and then we can define and describe new categories, conveying them to those who do not yet have them, by combining and recombining the names of our already grounded categories (Cangelosi & Harnad 2001) in propositions. Then and only then does the "theory-theory" come in, for verbal definitions and descriptions are higher-order category-detectors, too, as long as all their component terms are grounded (Blondin-Massé et al. 2008). Here we are right to call them "representations," for they are descriptions of categories, and can be given to and received from others without every individual's having to learn the categories directly from experience – as long as the category-names used in those descriptions are ultimately grounded in direct sensorimotor categories.

There is much ongoing research on the mechanisms of sensorimotor category learning in computers, neural nets, robots, and the brain, as well as on the origins and mechanisms of natural language processing. It is nowhere near Turing-scale, but this sketch rearranges the cognitive landscape a bit, to preview how we can, as Machery suggests, *do* "without concepts": What takes their place is innate and mostly learned sensorimotor category-detectors (for which the learning mechanisms are still not known, but neither prototypes nor exemplars are likely to play much of a role in them), progressively supplemented by verbal category representations composed of grounded category names describing further categories through propositions. The real challenge is getting this to work, Turing-scale. Alongside that momentous and substantive task, which of the landmarks we elect to dub "concepts" or "ideas" seems pretty much a matter of taste. [A fuller version of this commentary, entitled "Concepts: The Very Idea" is available at http://eprints.ecs.soton.ac.uk/18029.]

## Defending the concept of "concepts"

doi:10.1017/S0140525X10000373

Brett K. Hayes and Lauren Kearney

*School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia.*

**b.hayes@unsw.edu.au**

**lkearney@psy.unsw.edu.au**

**Abstract:** We critically review key lines of evidence and theoretical argument relevant to Machery's "heterogeneity hypothesis." These include interactions between different kinds of concept representations, unified approaches to explaining contextual effects on concept retrieval, and a critique of empirical dissociations as evidence for concept heterogeneity. We suggest there are good grounds for retaining the concept construct in human cognition.

The past decade has seen prolific growth in research on human concepts, both in terms of the phenomena studied and the generation of explanatory models. With research proceeding on so many fronts it is tempting to see the field as becoming fragmented with little prospect of a unified theory. However, it is hard to see how the field would be advanced by assuming that individual objects and events are represented by "several concepts" (*Doing without Concepts*, Machery 2009, p. 52), especially when the properties of these alternate concepts and the demarcation between them have not been specified. Moreover, we argue that there are good reasons for rejecting the "heterogeneity hypothesis." Here we identify three problems with this view.

***1. Are objects really represented in multiple ways?*** It is true that a variety of theoretical models (prototype, exemplar, theory-based) have been proposed to explain how concepts are represented and used However, Machery's implication that each has equal explanatory power and therefore warrants status as a separate kind of concept is incorrect. While there are certainly limits to the things that exemplar models can explain (see Murphy 2002), there is also little doubt that exemplar models do a better job of explaining laboratory data on category learning (see Kruschke [2005] for a review) and other key phenomena (e.g., the effects of category-irrelevant features on classification, as in Allen & Brooks 1991) than models which only assume storage of prototypical features.

The "theory" approach was originally proposed to explain aspects of conceptual experience that lie outside the purview of "data-driven" approaches, including sensitivity to conceptual coherence and the causal basis of categories (Murphy & Medin 1985). However, there have been significant advances towards integration of data-driven and theory-based approaches. Heit's (1994a) integration model incorporates prior knowledge into an exemplar model to successfully predict patterns of category learning in knowledge-rich domains (Carmichael & Hayes 2001). In Rehder and Murphy's (2003) KRES model, background knowledge is represented in a connectionist network to explain how data-driven category learning is accelerated in the presence of knowledge and how ambiguous features are reinterpreted in the light of feedback. Similar efforts to integrate data-driven and knowledge-based approaches have met with considerable success in explaining category-based induction (e.g., Kemp & Tenenbaum 2009). Note that these are not "hybrid models" in the sense used by Machery. They do not assume separate "exemplar" and "knowledge" modules. Instead, prior knowledge is represented in a format similar to that used for learning new exemplars (e.g., the integration model instantiates prior knowledge as retrieval of relevant exemplars from memory; KRES does it via feature to feature associations and similarity between known and novel concepts). These models show that the conceptual knowledge associated with "theory" and "exemplar" approaches need not be thought of as independent and that their interaction can be accounted for within a unified theoretical framework.

***2. Different representations for different tasks?*** One of the key sources of evidence cited in support of the heterogeneity hypothesis is that different aspects of object knowledge are activated in the service of different goals or tasks. Classification learning, for example, focuses attention on the differences between categories, whereas inductive predictions rely on knowledge of within-category structure (Markman & Ross 2003). The first thing to note is that this kind of flexibility is not limited to concepts. It is a ubiquitous feature of human cognition. For example, changes in retrieval conditions lead to important changes in the kinds of memories that are activated (Roediger 2008), but this does not mean that we should abandon the notion of object "memory." Second, it is important to recognise that empirical findings pointing to the "context dependency" of concepts can be oversold. A consistent empirical finding is that although context is important, a core conceptual representation of a given object is often retrieved when the object is used in a variety of conceptual tasks (Markman & Ross 2003). For example, taxonomic features appear to be activated by default in a variety of tasks involving biological concepts, whereas the activation of causal relations is dependent on task goals and domain experience (Shafto et al. 2007).

Finally, and most critically, it is possible to develop models that have the flexibility to explain context dependent changes in concept retrieval – SUSTAIN (Love et al. 2004) is a good example. Such models achieve flexibility by incorporating well-established psychological processes, such as selective attention and discrimination learning, and the use of a range of similarity metrics. However, they do so while retaining an assumption of a conceptual "core" in object representation. As well as explaining data that Machery claims are problematic for standard models, these approaches may be extended to examine how conceptual knowledge influences performance in domains such as problem-solving (Markman & Ross 2003) and recognition memory (Heit & Hayes 2008), which have traditionally been divorced from the study of concepts.

***3. What kind of evidence is needed to test the heterogeneity hypothesis?*** Machery suggests that the heterogeneity hypothesis is supported to the extent that we can identify experimental or neuropsychological dissociations across different conceptual tasks. Although similar views are frequently espoused in the psychological literature (e.g., Ashby & Maddox 2005), they should be treated with considerable caution. Deciding whether a given data pattern supports models that posit a single causal process (e.g., exemplar memory) or multiple, independent processes is a complex and tricky business. Patterns of single or double dissociations in categorization performance across tasks or patient populations can be produced by single-process models (e.g., Newell & Dunn 2008). Rather than simply assuming that patterns of dissociation point to a particular kind of cognitive architecture, future progress in mapping conceptual representations will need to apply rigorous analytical techniques such as state-trace analysis (Dunn 2008) and a careful comparison of well-articulated single- and multiple-process models.

The empirical work summarised by Machery suggests that our concept representations are complex and multifaceted. However, we argue that the balance of empirical and modelling work shows that the notion of "a concept" remains a useful heuristic in psychological explanation.

## The faux, fake, forged, false, fabricated, and phony: Problems for the independence of similarity-based theories of concepts

doi:10.1017/S0140525X10000385

Anne J. Jacobson
*Department of Philosophy, University of Houston, Houston, TX 77204-3004.*
**ajjacobson@uh.edu**
**www.class.uh.edu/phil/jacobson**

**Abstract:** Some things in our environment are not what they seem, and they provide a challenge to theories of concepts that emphasize similarity. Section 1 of my commentary explores a dilemma this situation creates for Machery. Section 2 describes a more general problem for prototype and exemplar theories. Section 3 locates a place for similarity-based concepts, and indicates an alternative to Machery's thesis.

***1. A dilemma for Machery.*** The large number of words we have for potentially deceptive things attests to the fact that they label a phenomenon that is important to us. Nonetheless, the knowledge often essential to distinguishing between the real thing and something else – causal knowledge – is left out of the prototype and exemplar accounts. Thus, Machery (2009) says, "neither prototypes nor exemplars store causal knowledge," and so tasks that require subjects "bring some causal knowledge to bear" (*Doing without Concepts*, p. 187) are evidence for a third theory of concepts, the theory or knowledge account. As a consequence, very good mimics and imitations, whose similarity to the real thing is often undetectable to the casual observer, answer to the criterion given by prototype and exemplar theories. Many red pandas look a great deal like raccoons, but they are not raccoons (Flynn et al. 2000; Sato et al. 2009). Seeming gold rings may be made of copper, cubic zirconias fool many people who take them to be diamonds, and gopher snakes are close indeed in appearance to rattlers, though they are not poisonous.

Machery argues that typically we have at least three independent but co-referential concepts for each kind of thing: a prototypical concept, an exemplar concept, and a knowledge/theory concept (see Murphy [2002] for the "knowledge" label). But there is a problem: Cases of good fakes (e.g., cubic zirconias vs. diamonds) either fit a correct use of the prototypical or exemplar concept or they do not. Suppose they do; if so, then the prototypical or exemplar concept of a diamond will apply to things that the knowledge/theory account concept does not, and the two will not be co-referential.

Alternatively, suppose that cubic zirconias that fit the prototypical concept are not correctly called diamonds. In this case, their independence is threatened. The criteria given by the knowledge/theory account are able to overrule those of the other two. Machery resists the idea that the knowledge/theory concept can so dominate (Ch. 3, sect. 3.3). However, given his claim that the sentence, "Tina Turner is a grandmother" (Machery 2009, p. 72) is true under one interpretation and false under another, the example shows that if one type does not dominate, different uses of "grandmother" will vary in reference.

The argumentative context in which Machery considers fakes leads him to restrict his attention to sentences. Further, his intuition does not clearly say that "Fake dollars are dollars" is false (2009, p. 72). But, as Machery insists, the range of the use of concepts covers far more, and intuitions about sentences do not settle whether criteria from the knowledge/theory account will overrule in practice.

***2. A more general problem.*** The failure to encode causal knowledge creates another problem for prototype and exemplar concepts and, by implication, Machery's thesis. Experimental work in Machery's discussion of prototype and exemplar theories is largely represented by descriptions of testing subjects on lists, sentences, pictures, and drawings, including patterns of dots. In contrast, Murphy (2002, p. 60) remarks that his knowledge/theory account rejects the idea that we "learn concepts in isolation from everything else (as is the case in many psychology experiments)." But our classifications of objects need to work to identify and track them in a dynamical environment. That sort of task carries quite different demands; among other things, classification needs to yield some clues about how our environment will unfold.

Infants have a capacity relevant to identifying kinds in a changing environment, one that Carey and Xu (2001) claim controversially does not appear until 12 months. That is, at some point infants expect kinds of objects to persist and not change into another kind. Before then, mere spatial-temporal continuity dominates; a duck moving behind a screen and an emerging rabbit do not have to be two distinct objects. Knowledge that kinds persist and do not turn into one another looks like causal knowledge of how things work; and Carey and Xu maintain it is preceded by other causal knowledge. Further, the presumption that kinds persist facilitates the acquisition of knowledge about how objects of some kind interact with their environment. Such knowledge is important because it enables one to anticipate some of what will happen.

From this perspective, there is at least a tension between saying that concepts are important in classifying and saying that they do not encode causal knowledge. It is unclear why using non-causal concepts would have any survival advantages, except in situations where the tasks to be performed have restricted success conditions. We look at such a task in the next section.

***3. Conclusion.*** Prototypes and exemplar theories may still characterize what is needed to fulfill some of the tasks vision needs to undertake. One is that of perceptual organization. Our vision at one stage is a succession of saccades; the successive-saccades stage leads to our experience of a scene of stable objects. This important result appears to be driven by gestalt groupings, but it can be facilitated by top-down categorizing that may in many cases be triggered by similarity to a prototype or exemplar (Öğmen 2007; Öğmen et al. 2006). Success in this sort of task is simply getting some organization. It can be achieved even if one is being taken in by fakes and the faux.

The preceding material suggests that there are layers to our uses of concepts, ones distinguishable in terms of the complexity of the knowledge needed. In creatures made for action, perceptual organization is required for almost any vision-guided action. Its success conditions may be much less demanding than those of another task, anticipating how one's environment is going to unfold. Finally, human cognitive life and public communication, among other things, have a considerable interest in getting it right, and not being taken in by potential deception. These different stages can be seen as posing less and more demanding tasks for our uses of concepts. Machery has instead posited multiple independent processes, but we have seen reasons for questioning the independence. That said, one should nonetheless expect that bold hypotheses from a subtle mind like Machery's will prove more resilient than one might first expect.

ACKNOWLEDGMENTS
I am indebted to Josh Brown and Josh Weisberg for discussions of this material.

## Hybrid vigor and conceptual structure

doi:10.1017/S0140525X10000555

Frank Keil
*Department of Psychology, Yale University, New Haven, CT 06520.*
**frank.keil@Yale.edu**

**Abstract:** Machery rightly points out a diverse set of phenomena associated with concepts that create challenges for many traditional views of their nature. It may be premature, however, to give up such views completely. Here I defend the possibility of hybrid models of concept structure.

In *Doing without Concepts*, Machery (2009) provides a service to us all by reminding us of the challenges of specifying what concepts are and how they are mentally represented. Moreover, by moving to the radical position that we should do away with the concepts altogether, he forces all of us to think more deeply about why we might want to preserve such notions. Finally, he is correct in pointing out the ways in which philosophical and psychological approaches to concepts often seem to be asking and answering different questions. Despite all this, it seems too extreme to assume that more traditional notions of concepts are bankrupt or that philosophers and psychologists are always talking past each other.

Here I focus on one alternative to Machery's proposal that he seems to dismiss too lightly – the idea that concepts might have a hybrid structure. I am not yet sure about the extent to which the potentially hybrid facets to concepts are actually parts of the concepts per se, or whether they should instead be considered as linked to concepts that are themselves much simpler atoms in the manner described by Fodor (1998). Cognitive science still has not fully answered Fodor's reasons for doubting that many cognitive phenomena associated with concepts reflect internal structural features of concepts as opposed to aspects of how we use and work with concepts; but to accept Fodor's arguments is to discount Machery's view as well, so let us assume here that we do want to assume internal structures to representations of concepts and that those structures help explain many psychological phenomena associated with concepts such as induction, categorization, and conceptual change.

Fodor (1998) has characterized concepts as "the smallest units of thought" and, in this respect, many psychologists and philosophers agree. Does such a characterization compel us to Machery's heterogeneity hypothesis, namely, that we must have several distinct concepts of water because, depending on context, we seem to use the concept in different ways? It is difficult to see why. Machery discounts the hybrid alternative by arguing that people will endorse conflicting statements about kinds such as tomatoes, whales, and the like, and those conflicts can only be explained by assuming that they are drawing on different concepts. These different uses are supposedly not "coordinated," and therefore people cannot be referring to the same concepts. But this coordination problem does not seem to be so lethal for hybrid views. If a given concept has a hybrid structure consisting, for example, of typicality-based information, causal schema, functional relations, and logical entailments, it might well be the case that different contexts cause people to weight those properties very differently and respond in different manners across tasks. Ever since Lakoff's (1972) demonstrations that different "hedges" such as "technically speaking" and "loosely speaking" can cause us to categorize kinds such as whales differently, it has been known that some hallmark ways we use concepts, such as categorization, can show strong variations as a function of situational and sentential concepts. But, if hybrid models are right, they seem more than adequate for dealing with such phenomena. Machery would need to provide a detailed model of internal hybrid representations of concepts that showed how they were intrinsically unable to computationally model such effects, and he has not yet done so.

Machery suggests that the parts of hybrid concepts must be "coordinated" such that this cannot create inconsistencies, such as categorization judgments that whales both are and are not fish. This coordination property is seen as an essential part of hybrid models, and hybrid models are described as incoherent or empty without it. This was not an obvious conclusion. Consider, for example, contexts in which we might describe a person as "short" and then "tall." If we see a 2-meter person practicing with Olympic gymnasts, we might well call him tall; but when observing him practice with an Olympic basketball team, we would call him short. We may well know his true height and the true heights of the other players, but the contexts call for different ways of assigning thresholds on the vertical dimension that we would then use to consider someone tall or short. Is this to be taken as evidence for multiple concepts of tall and short? Are there then an indefinitely large number of such concepts that are depending on all the micro-contexts that could shift the thresholds to tiny degrees? There is a strong tendency to resist such a route, and it seems that, for similar reasons, we should resist claims that hybrid structures are undermined by conflicting categorization judgments in different contexts. Categorization inconsistencies do not pose a problem if there are still systematic ways that categorization judgments can be shown to vary across contexts as a function of a description of their internal hybrid structure. Machery would be correct in pointing out that such systematic accounts are not yet fully worked out, but there are no obvious reasons why they might not be in the long run.

Hybrid approaches also have other appeals. They can, for example provide continuity and coherence to models of conceptual change over time, as, for example, when the causal or rule-based aspects of a concept become more differentiated as a child grows older and come to be weighted more and more relative to the associative components (Keil 1989; Keil & Newman 2010). Machery's heterogeneity alternative sees the child as progressing through a series of unconnected concepts that somehow magically tend to unfold in the same way across children. Hybrid models can also help explain how concepts differ across broad categories such as natural kinds and artifacts, where different components of the hybrid might be present to different degrees and accordingly assigned different weights, as well as being processed in different ways (Hampton et al. 2009).

Machery has done us all a great service. He raises a host of interesting troubles for many accounts of concepts, and he is to be commended for trying to build a larger common ground of inquiry between philosophers and psychologists. His book is a refreshing new perspective that prods all of us to further develop our own theories of concepts.

## The function and representation of concepts

doi:10.1017/S0140525X10000397

Sangeet S. Khemlani[a] and Geoffrey Goodwin[b]

[a]*Department of Psychology, Princeton University, Princeton, NJ 08540;*
[b]*Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104.*

**khemlani@princeton.edu**
**ggoodwin@psych.upenn.edu**
**http://www.princeton.edu/~khemlani**
**http://www.psych.upenn.edu/people/ggoodwin**

**Abstract:** Machery has usefully organized the vast heterogeneity in conceptual representation. However, we believe his argument is too narrow in tacitly assuming that concepts are comprised of only prototypes, exemplars, and theories, and also that its eliminative aspect is too strong. We examine two exceptions to Machery's representational taxonomy before considering whether doing without concepts is a good idea.

In *Doing without Concepts* (Machery 2009; hereafter *DwC*), Machery proposes that heterogeneity in the mental representation of "concepts" is sufficient to render that term useless. As he argues, the term can refer to exemplars, prototypes, and theories. However, it can also refer to defaults (Connolly et al., Fodor et al. 2007), aspects (Prasada & Dillingham 2009), Boolean concepts (Feldman 2000; Goodwin & Johnson-Laird 2010; submitted; Shepard et al. 1961), and connections yet to be discovered. Thus, in our view, Machery's taxonomy is too narrow, and it underestimates the degree of heterogeneity that exists in the representation of concepts. It excludes a variety of conceptual phenomena that do not fall within the purview of prototypes, exemplars, and theories. We turn next to describe two examples of such conceptual phenomena – Boolean concepts, and connections expressed by generics.

Boolean concepts – those that are composed out of negation (*not*), disjunction (*or*), and conjunction (*and*) – are an important kind of everyday concept. They occur frequently in the form of laws, rules, or procedures. Indeed, Machery's criteria for individuating concepts are themselves Boolean concepts (see sect. 2 of Machery's Précis of *DwC* in this issue). A concept does not need to be entirely Boolean in order for it to contain relevant Boolean structure, however. How individuals learn Boolean concepts is still not resolved, but many of the current leading contenders are not based on exemplars, prototypes, or theories (Feldman 2006; Vigo 2009). We have recently proposed an alternative theory, based on mental models, which analyzes the complexity of concepts in terms of the number of distinct possibilities that a concept can be compressed into (Goodwin & Johnson-Laird, submitted). This theory predicts the acquisition of Boolean concepts as well as, if not better than, the other leading contenders, and it too is not based on the representational mechanisms that Machery assumes to be exhaustive in explaining conceptual knowledge.

Recently, we published a paper documenting the occurrence of "conceptual illusions," in which people think that particular instances of a Boolean concept are possible when in fact they are not, and vice versa (Goodwin & Johnson-Laird 2010). A typical example is the concept: *green and large, or else green*, in a context in which all possible objects are either green or not, and either large or small (the "or else" here represents exclusive disjunction). A large percentage of individuals think that an object that is both green and large is possible given this description, which is in fact erroneous. The exclusive disjunction between the two clauses means that the only possible object is one that is both green and small. This error, as well as others like it, is predicted by the mental model theory's principle of conceptual truth. And as far as we can tell, accounts based on prototypes, exemplars, or theories have no way to explain these errors.

Other examples of conceptual phenomena unaccounted for by the taxonomy described in *DwC* include the connections and relations that link concepts together. Such connections can be concepts unto themselves, and are revealed by generic assertions such as "tigers are striped," "barns are red," or "ticks carry Lyme disease," which express generalizations about kinds of things (Carlson & Pelletier 1995; Gelman 2003; Lawler 1973). All three assertions are true for different reasons, and as such, generics provide a means for studying the types of connections we represent between concepts of kinds and properties. We have found that for statements such as "tigers are striped," the relation between the kind ("tigers") and the predicate ("are striped") can be distinguished from logical, statistical, and causal connections (Khemlani et al., submitted; Prasada et al., submitted). These distinctions could account for phenomena in concept learning and conceptual development without importing any assumptions made by other theories of concepts. Thus, by studying generics, it is possible to discover the conceptual structure of generalizations without assuming the representational structure of the concepts to which they refer. Machery proposes that future research should examine differences between generics, and particularly how they differ from quantified assertions (*DwC*, p. 200), and we agree wholeheartedly. Unfortunately, the proposals in *DwC* do not leave room to explore such advances in conceptual organization, as they encourage researchers to couch their work as falling within the domain of three fundamental classes of conceptual representation.

We do not think these phenomena, which point to even greater heterogeneity in the mental representation of concepts than Machery suggests, strengthen Machery's eliminativist argument to do away with concepts. The elimination of the term "concept" in favor of greater specificity tacitly endorses the assumption that prototypes, exemplars, and theories (and whatever else), are all that comprise concepts. It thus presupposes that contingent facts about the mental representation of concepts are the sole criterion for deciding whether "concept" ought to be preserved. But this presupposition ignores the common *function* that diverse sorts of concept play in representing knowledge and in communication. Concepts represent and convey systematic bodies of information, and they would retain this function regardless of how they are mentally represented. In other words, we think that the question of what counts as a concept needs answering at the computational level, not at the algorithmic one (cf. Marr 1982).

Thus, Machery's eliminativist argument is too powerful. It gives no grounds for thinking that the term "concept" is in an especially precarious position. In much the same way that the term organizes a wide array of representational processes, so too do terms like "thinking," "attention," and "memory." Would Machery have us do away with these terms as well, given heterogeneity in the cognitive processes to which they refer? Perhaps, but we think this is going too far.

In sum, the key functions of concepts are to represent and communicate knowledge, and this general functional property is what argues in favor of preserving the term "concept." We believe that heterogeneity at the level of mental representation is no obstacle to the further empirical investigation of concepts.

## Concepts are a functional kind

doi:10.1017/S0140525X10000403

Elisabetta Lalumera

*Dipartimento di Psicologia, Università di Milano-Bicocca, 20126 Milano, Italy, and Cogito Research Centre in Philosophy, Università di Bologna, 40100 Bologna, Italy.*

**elisabetta.lalumera@unimib.it**

**http://sites.google.com/site/elisabettalalumera/**

**Abstract:** This commentary focuses on Machery's eliminativist claim, that "concept" ought to be eliminated from the theoretical vocabulary of psychology because it fails to denote a natural kind. I argue for the more traditional view that concepts are a functional kind, which provides the simplest account of the empirical evidence discussed by Machery.

The novelty of Machery's proposal in *Doing without Concepts* (Machery 2009) is the claim that the term "concept" ought to be eliminated from the theoretical vocabulary of psychology, because it fails to denote a natural kind. I will not dispute the claim that concepts are not *a* natural kind. There is a growing consensus among psychologists that the structure of concepts may vary along many dimensions, depending on expertise, domain of objects categorized, and conceptual task involved. Much of this evidence is reviewed by Machery himself, as well as in other recent studies in the philosophy of psychology (Piccinini & Scott 2006; Weiskopf 2009b). My point here is that, on philosophical grounds, this evidence is perfectly compatible with the much less revisionary claim that concepts are a functional kind. Something is a concept by virtue of the function it performs within a cognitive system, and something is the concept of a certain category *C* (at least partially) by virtue of the further specific function of representing it. It is a further question whether or not the functional kind "concept" is realized by natural kinds (Weiskopf 2009b). Functional kinds can be individuated and described independently of their realizers. This, however, does not deprive them of a central role in experimental psychology.

***1. Concepts are a functional kind.*** It is disputable that a full characterization of the psychological usage of "concepts" is:

"those bodies of knowledge that are used by default in the processes underlying the higher cognitive competences" (Machery 2009, p. 7). Concepts are also normatively characterized as those bodies of knowledge that *can* perform a double function – namely, abstraction and projection of knowledge – and that *are able to* be recombined almost freely in order to form more complex concepts and thoughts. Let us focus on the first two functions, as the third one is arguably not specific to concepts only. Abstraction is the "bottom-up" process of extracting information from a single encounter with an object or property-instance, and generalizing such information to all encounters with that object or property. The experience of tasting rhubarb once and finding it bitter would be of no use if I could not store it as information about rhubarb independently of the specific episode of tasting it, by means of a general representation – a concept (Bloom 2000). Category induction is the complementary "top-down" process of projecting such knowledge to new encounters. When you tell me you like rhubarb pie, I form the expectation that you will also like other bitter-tasting foods. This is an application of my concept of rhubarb. Thus, the function of a concept is that of a "mental glue," which connects one's past experience with the present (Millikan 1998; Murphy 2002). It is because they perform this complex function that concepts are used by default in our higher cognitive capacities, and not vice versa. Psychologists are interested in discovering how such function is performed – by what structures and mechanisms. The functional properties of concepts provide constraints on what may count as an adequate concept-realizer. This is the first sense in which concepts as functional kinds are not dispensable in psychology.

***2. Concepts as a functional kind are multiply realizable in a broad sense.*** Vehicles and food are familiar non-mental functional kinds. Traditionally, the essence of a *mental* functional kind is to bring about some outcome, or to exercise some capacity (Kim 1992; Putnam 1967). With "broad sense" here, I mean that they are realizable by items characterized by different structural properties. Assuming that abstraction and category induction are the fundamental capacity characterizing the functional kind "concept," there are plausibly many ways in which a human mind can organize itself in order to exercise them. Capacities may well be individuated by their ends, not just by their means (Millikan 1998). So, for example, a chemist's capacity to abstract and project information about water may involve means that I do not possess. But so long as we are both able to abstract, accumulate, and project information about water, we share the same capacity, that is, we both possess a concept of water. The variability of structure of the functional kind "concept" is just as compatible with the empirical evidence discussed by Machery, as his own proposal is. It has the advantage of leaving it open whether prototypes, exemplars, and theories exhaust the possible concept-realizers or, more plausibly, not.

***3. Concepts as kinds are multiply realizable in a narrow sense.*** The narrow sense is the traditional sense associated with the phrase "multiple realizability" in philosophy of mind. It is the idea that a token (mental) functional kind can be realized by different physical substances. The classical example is a state of pain, multiply realized by silicon chips or C-fibers. To deny that concepts are multiply realizable in this narrow sense is very demanding. It requires a commitment to a strong form of physicalism. Contemporary neurophysiological research on concepts aims at individuating which areas of the human brain are involved in specifically conceptual tasks and explaining how this is done. It is not committed, however, to the further metaphysical claim that no other kind of matter, appropriately organized, could bring about the same capacities. If neurophysiology is not, surely psychology need not be committed to a robust reductionist agenda. That is, there is no reason why psychology should dispense with the functional kind "concept" *qua* multiply realizable. The neurophysiological evidence quoted by Machery on the variability of concepts is compatible with the anti-eliminativist functional view.

***4. The functional kind "concept" has a role in psychology.*** Machery claims that there are pragmatic reasons for the elimination of "concept," if intended as a natural kind term. This is not so if it is taken as a functional kind term. Consider the task of accounting for the fact that one's general representation of a category – say, *dogs* – changes over time. Initially it is constituted by a bunch of exemplars, then it develops into a summary prototypical representation, and later it becomes a theory. In order to describe the three structures as stages of a diachronic process of change, one needs to make reference to what is common to them, namely, realizing the capacity of representing dogs generally, or being instances of the functional kind "concept of *dogs*." This intrapersonal explanation properly belongs to the psychologist's agenda, and so do interpersonal accounts of variability.

## From conceptual representations to explanatory relations

doi:10.1017/S0140525X10000415

Tania Lombrozo

*Department of Psychology, University of California, Berkeley, CA 94720.*
**lombrozo@berkeley.edu**
**http://cognition.berkeley.edu/**

**Abstract:** Machery emphasizes the centrality of explanation for theory-based approaches to concepts. I endorse Machery's emphasis on explanation and consider recent advances in psychology that point to the "heterogeneity" of explanation, with consequences for Machery's heterogeneity hypothesis about concepts.

The "theory" approach to concepts, one of three that Machery endorses in *Doing without Concepts* (2009), reflects a widespread view in psychology that is typically taken to be promising but in need of further development (e.g., Carey 1985; Gopnik & Meltzoff 1997). According to this approach, concepts constitute or are embedded within intuitive theories about the world. However, both philosophers and psychologists have been quick to point out that the notion of an intuitive theory is underspecified, and that appeals to scientific theories run the risk of substituting one mystery for another (Laurence & Margolis 1999). Machery is aware of these concerns, but he identifies the locus of the theory view's commitments not in the notion of an intuitive theory or in appeals to scientific theory, but in *explanation.* For an advocate of the theory approach, concepts store "knowledge that can explain the properties of the category members," so "much hangs on the notion of explanation" (Machery 2009, p. 101).

Machery is right to focus on the central role of explanation for theory-based approaches to concepts. In a set of influential papers that kick-started theory theorizing, theories were defined as "any of a host of mental explanations" (Murphy & Medin 1985, p. 290) and characterized in terms of "laws and other explanatory mechanisms" (Carey 1985, p. 201). Categorization and category learning were described as "special cases of inference to the best explanation" (Rips 1989, p. 53), with a concept invoked "when it has a sufficient explanatory relation to an object" (Murphy & Medin 1985, p. 295). But such appeals to explanation were not proposed against a backdrop of psychological theories of explanation. Rather, they predominantly appealed to what Machery calls a "folk understanding of explanation" (2009, p. 102), contributing to the concern that theory-based approaches to concepts are underspecified.

What to do? One option would be to develop psychological theories of explanation, with our folk understanding as a guide. A second option would be to ground psychologists' appeals to explanation in philosophical theories of scientific explanation. Given the analogy between science and cognition that motivates 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

(Markman & Ross 2003). A growing body of research suggests that the tasks people perform while learning categories influences what people learn about those categories. For example, Yamauchi and Markman (Yamauchi & Markman 1998; Yamauchi et al. 2002) found that people tend to learn about features that distinguish one category from another when learning by classifying examplars. In contrast, people tend to acquire category prototypes when learning by making predictive inferences about new category members that are missing some features (see also Chin-Parker & Ross 2002). Markman and Ross (2003) drew on the memory literature and argued that this type of category acquisition involves transfer appropriate processing, which is also observed in memory (e.g., Morris et al. 1977).

The relationship between categorization and memory is also being driven forward by work on the influences of communication on category acquisition. Garrod and colleagues (Garrod & Anderson 1987; Garrod & Doherty 1994) found that people tend to arrive at a common method for referring to information in the environment over the course of conversing with others about those objects. Markman and Makin (1998) observed that the similarity of categories across people is made more uniform by communicating with others. The act of establishing a common reference influences memory by ensuring that category labels refer to a common set of properties.

These communicative factors have a significant influence on the ability to extract information about categories from memory. For example, Malt et al. (2003) found that it is hard to predict the labels people give to a common set of objects (such as jars, bottles, boxes, and containers) based solely on similarities among the items. Furthermore, there is no clear relationship between the labels given to objects in different languages. Instead, the labels given to objects reflect communicative history of those objects in the language. The labels people learn for objects, then, are determined in part by the utility of those labels for communication. We use a particular word for an object, because we know that others will know what we are talking about when we use that label.

I am not suggesting that we abandon theoretical devices like prototypes, exemplars, and theories as constructs for helping us to understand categorization behavior. Instead, the theoretical basis of research on categorization needs to be modified.

We must recognize that the tasks that we use to study psychology in the lab do not cut nature at its joints. Instead, we must look for the cross-cutting psychological mechanisms that are involved in a variety of different tasks. That means that theoretical constructs from memory, communication, attention, and motivation are all relevant to the study of category acquisition.

In addition, the valuable constructs from research on categorization must be exported to the study of other cognitive processes where they may be valuable. Undoubtedly, prototypes, exemplars, and theories have value in understanding a variety of aspects of cognitive processing beyond categorization. For example, exemplars play an increasingly important role in theories of automaticity (Logan 1988; 2002).

Ultimately, we must transition from a cognitive science in which psychological tasks organize our understanding of psychological mechanisms to one based on an understanding of the way a core set of psychological mechanisms permit us to perform those tasks.

## Concepts versus conceptions (again)[1]

doi:10.1017/S0140525X10000440

Georges Rey

*Department of Philosophy, University of Maryland, College Park, MD 20742.*
**georey@earthlink.net**
**http://sites.google.com/site/georgesrey**

**Abstract:** Machery neglects the crucial role of concepts in psychological explanation, as well as the efforts of numerous "externalists" of the last 40 years to provide an account of that role. He rightly calls attention to the wide variation in people's epistemic relations to concepts – people's *conceptions* of things – but fails to appreciate how externalist and kindred proposals offer the needed stability in concepts themselves that underlies that variation.

In proposing to dispense with concepts in *Doing without Concepts*, Machery (2009) neglects to notice how psychological explanation presupposes them. How could we begin even to describe, for example, the Müller-Lyer illusion unless people share a concept of *longer than*; or the gambler's fallacy, without them sharing *more likely*? Concepts seem to be natural kinds at least to the extent that they are the entities over which psychology generalizes.

But what is a concept that can play this role? Many recent philosophers (other than the atypical Peacocke [1992]) have offered proposals that Machery amazingly ignores, such as the various forms of "externalist" views that have been proposed by, for example, Kripke (1972/1980), Putnam (1975), Dretske (1981), Devitt (1981; 1996), Millikan (1984), Burge (1986), and Fodor (1990; 1998), according to which the identity conditions for a concept are provided in part by historical and counterfactual relations the thinker bears to phenomena in the external world. At most, Machery considers some of the "intuitions" that motivate externalism, only to dismiss them as too cross-culturally variable and unreliable.[2]

Machery correctly notes that a problem with many *traditional* philosophical accounts like Peacocke's is that they fall afoul of Quine's attack on the analytic/synthetic distinction (Machery 2009, p. 39). But he misses how that attack threatens any purely *epistemic* proposal, including his own appeal to "bodies of knowledge" (p. 12). This is puzzling, given Fodor's extensive discussion of many of these issues in at least *five* of his recent books, a discussion quite often directed explicitly at the psychological theories Machery considers. The problem is that "bodies of knowledge" vary between people and stages in a person's life. Unless one restricts the *relevant* knowledge in some principled way, no two people (or stages) will share a concept, since, short of coincidence, no two people (or stages) will bring exactly the same knowledge or procedures to bear in making many of their judgments.[3] Call this the problem of epistemic variability. Rather than inviting us to abandon the notion of concept, perhaps this variability is simply a reason to abandon an epistemic conception of it.

Machery would probably reply that I'm merely pressing here the philosophical notion of concept, which is concerned with the individuation of propositional attitudes (2009, pp. 32–33), a topic that, surprisingly, he claims is not the concern of psychological theory (p. 34). Indeed, he claims that when the two tasks are properly distinguished, "most philosophical attacks against the psychological theories of concepts are decisively undermined" (p. 51).

"Decisively"? There are a number of important reasons to think not: In the first place, as Machery himself notes (2009, pp. 35–37), many psychologists themselves are hardly clear about the difference between the two concerns, often presenting their work as refuting the Classical View of traditional philosophy and presupposing some philosophical, usually verificationist alternative (see Rey 1983).

Secondly, concept identification would seem to be an issue not about how people do think under pressure, but how they could think if they were to reflect – what they could understand – and, *pace* Machery (p. 34), this seems as apt a topic as any for psychological research. It may well be that peoples' prototype of a doctor is of a man in a white coat; but if they found the thought of young woman doctor in a dark one as unintelligible as a round square, that would be a reason to think they didn't have the concept doctor. It's because people have a concept doctor that transcends their prototypes that it's worth reasoning with them, that is, modifying their epistemic position by citing evidence or argument, but keeping the concept the same. As I emphasized in Rey (1985), we need to distinguish the concept of something from merely the (epistemic) conceptions of it that have been too much the focus of the psychological research Machery reviews. Concepts are what remain stable across variability in conceptions, and so give argument a point, framing the questions of what people could learn and what might be the limits of reason and thought.

Lastly, as Machery notes, there is a need of a "coherent framework" (2009, p. 247) for bringing together the psychologists' different proposals. Something like the externalist strategies may be just the sort of thing for the purpose. They at least address the problem of epistemic variability in a promising way, proposing that concepts are constituted by what is "explanatorily basic."[4] A crucial feature of this strategy is that it makes no commitment to the character of the representations people use in ordinary circumstances requiring rapid reasoning, or even in "acquiring" the concept (thereby also allowing for many concepts to be innate).

Nothing I have said here is meant to suggest that any externalist (or other) proposals are yet satisfactory (see Segal [2000] and Rey [2009a] for serious qualms). It's not that Machery should have endorsed externalist or related strategies; my point is simply that he should have discussed them, particularly before giving up on the concept of concept entirely.

NOTES

**1.** This commentary summarizes my longer review of Machery's book available at: http://ndpr.nd.edu, http://ndpr.nd.edu/review.cfm?id=16608, which readers should consult for more detail. It and other relevant material of mine are also available at: http://sites.google.com/site/georgesrey.

**2.** See Devitt (forthcoming) for reasons this dismissal is rash. Note that Machery fails to notice that, for example, Burge's (1986) arguments for externalism don't rest on ordinary intuitions alone, but on explanatory features of a Marrian theory of vision.

**3.** Of course, there may be *similarities* and *overlap* in many people's knowledge and procedures (at least relative to the stability of their *other* concepts); but mere similarity and overlap aren't identity, and it is identity in concepts that is needed to sustain serious explanations, such as ones about cognitive development, vision, or language; cf. Fodor (1998, pp. 30ff).

**4.** Devitt (1996) and Horwich (1998a, p. 41) propose that treating the meaning constitutive conditions as the ones on which all other uses of a symbol explanatorily depend, can be seen as a mixed internalist/externalist variant of Fodor's (1990) "asymmetric dependency" that achieves some of its same effects, but without its strong externalist commitments; see Rey (2009a) for discussion.

## Why *don't* concepts constitute a natural kind?

doi:10.1017/S0140525X10000452

Richard Samuels and Michael Ferreira

*Department of Philosophy, The Ohio State University, Columbus, OH 43210-1365.*
**samuels.58@osu.edu**
**ferreira.23@.osu.edu**

**Abstract:** Machery argues that concepts do not constitute a natural kind. We argue that this is a mistake. When appropriately construed, his discussion in fact *bolsters* the claim that concepts are a natural kind.

***Introduction.*** A central claim of Machery's Précis (target article) – and of his book, *Doing without Concepts* (Machery 2009) – is that concepts do not constitute a natural kind. Until reading his work on the topic, we would have been inclined to agree. But he has changed our minds. Machery's discussion, when appropriately construed, provides grounds to suppose that concepts *do* constitute a natural kind.

***What is a natural kind? .*** Though the notion of a natural kind has been characterized many times over, philosophers of science have, in recent years, reached a consensus – or as close to consensus as philosophers ever get – according to which natural kinds are *Homeostatic Property Clusters* (HPC; Boyd 1991; Machery 2005). According to the HPC account, a kind *K* is natural if:

H1. It is associated with a contingent property cluster: a range of characteristics or symptoms which tend to be co-instantiated by instances of the kind, but need not be genuine necessary conditions for membership.
H2. There is some set of empirically discoverable causal mechanisms or structures – a *causal essence* – that explains the covariation of these symptoms.
H3. To the extent that there is any real definition of what it is for something to be a member of the kind, it is not the symptoms but the presence of the causal essence producing the symptoms that are definitive of kind membership (Boyd 1989).

This conception of natural kinds is intended to capture the central features of those kinds paradigmatically studied by science. Most importantly, the fact that the existence of property-clusters depends on the operation of an underlying causal essence helps explain why such kinds are typically subsumed by many non-accidental, empirical generalizations (Griffiths 1997; Machery 2005).

***Why suppose that concepts constitute a natural kind?*** Machery endorses the HPC account, yet he denies that concepts constitute a natural kind We maintain, however, that if his claims about the cognitive scientist's notion of a concept are correct, then concepts are plausibly a natural kind in the HPC sense. First, consider some of the characteristics that, according to Machery, concepts possess:

1. Concepts consist in *bodies of information.*
2. Concepts are *stored in long-term memory*, and *persist* even when not actively deployed.
3. Concepts are *non-proprietary.*
4. Concepts are *default* representations automatically retrieved from long-term memory.
5. Concepts are *subject to temporal and inter-subjective variation.*
6. Concepts are *internally connected.*
7. Concepts are *internally coherent.*

No doubt other generalizations hold of concepts as such. But even the characteristics 1–7 just listed provide grounds to suppose that concepts are a natural kind. If true, they suggest that concepts satisfy the main conditions imposed by the HPC account. Specifically:

C1: Concepts exhibit a reliably *covarying property cluster*, including internal coherence, connectedness, persistence, and so on.
C2: The existence of the property cluster depends largely on the operation of an abstractly, functionally characterizable kind of process. Roughly: Most of the properties depend on facts about long-term memory, the operations that it performs, and its relations to other higher cognitive processes.
C3: To the extent that inclusion in the kind, *concept*, can be defined, it should be the condition of being subject (in the right kind of way) to the relevant causal mechanism – i.e., long-term memory – that defines kind membership. The condition that Machery labels "Default" partially captures this idea.

In short, Machery's comments, appropriately construed, provide reason to suppose that concepts are a natural kind in the HPC sense.

***Why would one think otherwise?*** Clearly, Machery is not inclined to accept this conclusion Why not? Here are some brief comments on the most plausible arguments that we could extract from his discussion.

*Argument 1:* Different kinds – for example, prototypes and theories – satisfy the conditions on concepthood. So, while prototypes and theories might be natural kinds, concepts as such are not.

*Response:* There is nothing wrong with superordinate natural kinds. For example, *metal* is plausibly a natural kind, even though it contains subordinate kinds such as gold and lead. For all the present argument shows, the same is true of *concepts.*

*Argument 2:* Characteristics 1–7 are intended to explicate the cognitive scientist's *notion* of a concept, not to characterize the kind as such. If so, treating them as parts of the kind-syndrome is tantamount to confusing issues about the semantics of the word "concept" with issues about the nature of the kind.

*Response:* Even if 1–7 are parts of the cognitive scientist's notion of a concept, there is no inconsistency between this claim and the claim that they are also parts of the property cluster exhibited by members of the kind.

*Argument 3:* Members of a natural kind are supposed to share *many* covarying properties and, hence, be subsumed by many empirical generalizations. But while concepts share *some* properties, and are subsumed by *some* empirical generalizations, there are too few to merit natural kind status.

*Response:* How many properties must members of a kind share in order for the kind to be natural? This is a silly question because there is no reason to suppose any precise cut-off point. Presumably some kinds manifest more common properties than concepts do, and, in that regard, are better examples of natural kindhood than concepts. But this is a *very* unsurprising conclusion. (Who would have supposed otherwise?) Moreover, it surely does not follow from this that concepts are not a natural kind *simpliciter.* On the contrary, there are two obvious considerations that explain the relative lack of properties and generalizations associated with concepts without impugning their natural kind status:

1. *Concept* is a *psychological* kind; and psychological kinds quite generally appear to underwrite fewer empirically rich generalizations than the kinds cited by many other disciplines. (Exercise: Compare chemistry or molecular biology with psychology.)
2. *Concept* is a *superordinate* kind. So, in point of logic, the kinds it subordinates –for example, prototypes and exemplars – will have the characteristics of concepts as such, plus their own specific characteristics. Still, it would be wrong to suppose that only subordinate kinds are natural, while kinds like *concept* (or metal, or alkali, etc.) are not.

***Conclusion.*** Concepts plausibly constitute a natural kind, in the HPC sense. So, *pace* Machery, we do not advocate their elimination from cognitive science. Instead, we think that Machery's discussion supports a rather more banal conclusion, namely: Concepts are, in some regards, a less good example of natural kindhood than many other kinds studied by science. But this is very unsurprising, and is largely explained by two facts: Psychology appears to generate fewer robust empirical generalizations than many other sciences; and in point of logic, superordinate kinds manifest fewer regularities than the kinds they subordinate.

## Evidence of coordination as a cure for concept eliminativism

doi:10.1017/S0140525X10000464

Andrea Scarantino

*Department of Philosophy, Georgia State University, Atlanta, GA 30302.*
**ascarantino@gsu.edu**

**Abstract:** I argue that Machery stacks the deck against hybrid theories of concepts by relying on an unduly restrictive understanding of coordination between concept parts. Once a less restrictive notion of coordination is introduced, the empirical case for hybrid theories of concepts becomes stronger, and the appeal of concept eliminativism weaker.

In *Doing without Concepts*, Machery (2009) makes a persuasive case that there is no unique body of information that plays the concept-role. An important methodological consequence follows: Psychologists of concepts should stop asking whether the realizer of the concept-role is a prototype, an exemplar, or a theory. None of these bodies of information alone can explain all phenomena pertaining to higher cognitive competences. What I reject is the additional thesis that the term "concept" should be eliminated from the vocabulary of psychology because it does not designate a natural kind, roughly a maximal set of entities about which many scientifically interesting generalizations can be formulated.

Machery is inspired by Griffiths' (1997) proposal that we split emotion theory into the study of several heterogeneous kinds of emotions, such as affect programs (e.g., fear of a suddenly looming object), higher cognitive emotions (e.g., guilt about having missed a friend's birthday) and socially sustained pretenses (e.g., going postal after having been fired). Mere evidence of heterogeneity, however, isn't a sufficient reason for eliminating a kind (Piccinini & Scott 2006). If a set of heterogeneous entities is part of a higher-level structure about which scientifically interesting generalizations can be issued, the heterogeneity of the parts is compatible with the existence of a higher-level natural kind to which they jointly belong. We do not think that "elephant" does not designate a natural kind just because elephants have heterogeneous parts: these parts are integrated in a way that allows biologists to formulate many scientifically interesting generalizations about elephants.

Now, in the case of emotions, it is quite clear not only that affect programs, higher cognitive emotions, and socially sustained pretenses are heterogeneous, but also that there is no overarching higher-level entity of which they are parts. Several distinct causal mechanisms are responsible for the occurrence and unfolding of emotion episodes of different kinds, and this prevents the emergence of a unified scientific psychology of emotions.

In the case of concepts, the situation is considerably murkier, because we do have at least preliminary evidence that co-referential prototypes, exemplars, and theories are integrated parts of a larger whole (the Hybrid Hypothesis). To get clear on this topic, we must determine when bodies of information are parts of a larger whole. Machery presents two individually necessary and jointly sufficient conditions: *Connection* and *Coordination* (Machery 2009, p. 64).

*Connection* between bodies of information requires that the retrieval and use of one body of information in a given cognitive process facilitates the retrieval and use of the remaining bodies of information in other cognitive processes. *Coordination* between bodies of information requires that they never produce inconsistent outcomes, for example, inconsistent categorization judgments.

Armed with this understanding of the part–whole relation, Machery proceeds to present evidence against Coordination. Language users, he argues, can judge that some liquid is water because it fits the water prototype (water is transparent, drinkable, found in lakes, etc.), but at the same time that it is not water because it does not fit the water definition (water is $H_2O$). Since neither judgment is allegedly taken by the folk to be authoritative over the other, Machery concludes that the Coordination condition is violated, and that this counts as a strike against the Hybrid Hypothesis.

My main problem with this line of reasoning is that the Coordination condition is inadequate. In general terms, the parts of a given concept are coordinated just in case they work together, in ways to be empirically discovered, in at least some of the processes underlying higher cognitive competences.

Machery's coordination condition offers a very narrow interpretation of how such bodies of information are supposed to work together, namely, by avoiding conflict in all circumstances in which they are jointly activated. But the fact that this very specific principle of organization is not empirically supported does not constitute evidence that bodies of information are not coordinated in some other, theoretically interesting way.

First, there are forms of coordination that have nothing to do with joint activation. For instance, bodies of information can be acquired, rather than deployed, in a coordinated fashion. Finding out how they are acquired demands unveiling what we may call *generalizations of coordinated acquisition.* For example, some have argued that the statistical information contained in prototypes and the causal, functional, and nomological information contained in theories are acquired in part through a process of abstraction from exemplars (Heit 1994b). This would represent an important aspect of integration between bodies of information.

A further aspect of integration is that the specific features that are abstracted in a prototype appear to be heavily influenced by the subject's background theory (Wisniewski & Medin 1994). Some consider the integration between statistical and theoretical bodies of information to be so tight that they have proposed representing prototypes not as simple feature lists, but rather as *schemata* which make explicit the theory-based relations between statistically common features (Cohen & Murphy 1984).

Second, there are forms of integrated activation that do not aim to avoid conflict under all circumstances. Finding out about such alternative forms of coordination demands unveiling what we may call *generalizations of coordinated deployment.* Some have argued, for instance, that whether theories are activated or prototypes are activated in a categorization task depends on the perceptual richness of the input: perceptually rich inputs activate prototypes, and perceptually poor inputs activate theories (Murphy 2002, p. 168). This would be a way in which different bodies of information can work together towards a cognitive end by being differentially, as opposed to jointly, activated.

There is also evidence that in some circumstances in which jointly activated bodies of information lead to conflict, the conflict is resolved according to a general trumping principle. For instance, Keil (1989) has argued that conflicts between prototype-based and theory-based categorizations of biological categories tend to be systematically resolved in favor of the theory-based categorization (e.g., a raccoon that is made to look exactly like a skunk while preserving its internal properties is judged to be a raccoon even it if fits perfectly the skunk prototype). This particular form of coordination is lost, on the other hand, when biological categories are substituted with artifact categories.

My central point is that if enough *empirical generalizations of coordination* can be unveiled, both of the *acquisition* and of the *deployment* variety, a case can be made that "concept," *pace* Machery, designates a higher-level natural kind for the purposes of scientific psychology. Notice that this strategy for preserving the scientific integrity of the notion of concept differs from the strategy of unveiling empirical generalizations that are insensitive to the differences between prototypes, exemplars, and theories (e.g., Weiskopf 2009b). Generalizations of coordination are eminently sensitive to the differences between prototypes, exemplars, and theories, but they can provide evidence that they are components of an integrated higher-level entity.

ACKNOWLEDGMENT
Thanks to Gualtiero Piccinini for helpful comments on a previous draft.

## Conceptual atomism rethought

doi:10.1017/S0140525X10000579

Susan Schneider
*Department of Philosophy, Center for Cognitive Neuroscience, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA 19104-6304.*
**susansdr@gmail.com**

**Abstract:** Focusing on Machery's claim that concepts play entirely different roles in philosophy and psychology, I explain how one well-known philosophical theory of concepts, Conceptual Atomism (CA), when properly understood, takes into account both kinds of roles.

In *Doing without Concepts*, Machery (2009) argues that concepts should be eliminated from psychology. He further claims that psychological and philosophical theories of concepts talk past each other: "when philosophers and psychologists develop theories of concepts, they are really theorizing about different things" (p. 4). I am not convinced that concepts should be eliminated from psychological theorizing, but today, I focus on Machery's claim that concepts play entirely different roles in philosophy and psychology: I shall aim for a partial reconciliation in the context of one well-known philosophical theory of concepts, Conceptual Atomism (CA), a view pioneered by Jerry Fodor (see Fodor 1998; Laurence & Margolis 2002).

According to Machery (2009), philosophical theories are mainly interested in concepts as they figure as constituents in propositional attitudes and hence find the matter of reference determination to be of key import. Psychological theories, in contrast, focus on topics such as categorization, analogical reasoning, and induction (p. 34). As an example of how divorced philosophical and psychological interests are, Machery raises CA:

> There is little point in blaming some philosophical theories of concepts, such as Fodor's theory, for being unable to explain how we reason, how we categorize, how we draw analogies, or how we induce (as does, e.g., Prinz 2002). For, simply, a philosophical theory of concepts is not in the business of providing such explanations. (p. 37)

CA claims that the nature of a concept is determined (or, as philosophers say, "individuated"), at least in part, by the information that the symbol carries about the world. It further holds that lexical concepts are primitive, being semantically unstructured: that is, they are not comprised of further concepts (Fodor 1998; Margolis & Laurence 1999, p. 62; Prinz 2002, p. 89).

Many would agree with Machery's claim that CA ignores the role that concepts play in thought (i.e., categorization, induction, etc.). Still, Machery's conclusion is premature. Surprisingly, CA has a neglected resource for capturing the role the concept plays in thought; this is because, as we'll see, according to CA, a concept is defined by its symbol type. For, consider that CA defines primitive concepts in the following manner:

> **Existence condition**: A primitive concept exists if and only if a primitive symbol in the language of thought (LOT) has a broad content
> **Identity condition**: Primitive concepts are identical if and only if they are of the same symbol type and have the same broad content (Fodor 1998, p. 37).

(Where the broad content of a symbol is, roughly, what the symbol refers to.)

The reader may be surprised that I construe CA as saying that symbols individuate concepts. But observe that Fodor himself acknowledges that broad content alone is inadequate for the purpose of individuating primitive concepts because it fails to distinguish co-referring concepts (e.g., groundhog/woodchuck; Cicero/Tully) (Fodor 1998; 2008). He explains that it is for this reason that he distinguishes concepts in terms of their mode of presentation (MOP) types, as well as their broad contents (Fodor 1998, Chs. 1 and 2; 2008, Ch. 3, especially p. 70). And, as philosophers know, Fodor regards MOPs as being symbols. Hence, even working within Fodor's original framework, CA has the resources to individuate concepts along two dimensions: a symbolic dimension and a referential one.

Now let us see how CA's symbolic element captures the role the concept plays in thought. I have argued that symbols are defined by the role they play in computation (Schneider 2009). Although Fodor, ironically, challenges aspects of my view, notice that even Fodor himself writes that MOPs (i.e., symbols) are individuated by their role in mental processing: "If MOPs are both in the head and functionally individuated, then a MOP's identity can be constituted by what happens when you entertain it" (Fodor 1998, p. 20; see also Fodor 2008, p. 92).

Piecing these observations together, we are now ready for the payoff. When CA is properly understood, both psychological and philosophical interests are brought together into a singular package: A lexical concept is a semantically unstructured "atom" that is individuated by both its meaning (in particular, its broad content) and its symbol type, where the symbol, in turn, is individuated by the computational role that it plays in one's cognitive economy (including, importantly, its role in mental processes such as categorization, induction, and analogical reasoning) (Schneider 2009; forthcoming). So CA can say that the features of concepts that psychologists are traditionally interested in are built into concepts' very natures. For example, consider the prototype view. In the eyes of CA the experimental results in the literature on prototypes are indications of features of certain symbols' underlying computational roles, and these roles determine the relevant concept's natures.

CA is ecumenical. For now consider the competing theory-theory, which suggests that concepts are mini-theories of the categories that they represent, encompassing our beliefs about hidden features, underlying causal mechanisms, and ontological divisions. Advocates of the theory-theory suggest that it captures explanatory relations between features while the prototype theory does not. For instance, in a well-known criticism of the prototype view, children appear to use beliefs about a creature or thing's underlying essence to override categorization judgments based on superficial, sensory features (Keil 1989). In the eyes of the conceptual atomist, this debate provides insights regarding concepts' underlying computational roles. But no matter how the debate plays out, concepts' natures are nevertheless determined by their broad contents and symbol types. Indeed, perhaps some concepts have computational roles that are explained by the prototype view, while others have roles that are illuminated by the theories view.

So the conceptual atomist who pays attention to the neglected, symbolic element of concepts can offer a more comprehensive theory of concepts than is normally supposed. CA is ecumenical enough to incorporate different sorts of concepts (prototypes, theories, etc.), and it also speaks to philosophers' traditional interest in reference determination. Finally, although I have focused on the LOT approach, my remarks can apply to philosophical approaches to concepts more generally, insofar as they individuate concepts by both meaning and computational (or functional) role.

## Banishing the thought

doi:10.1017/S0140525X10000476

Nina Strohminger and Bradley W. Moore

*Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1043.*

**humean@umich.edu**

**bradmoor@umich.edu**

**Abstract:** The first seven chapters of *Doing without Concepts* offer a perfectly reasonable view of current research on concepts. The last chapter, on which the central thesis of the book rests, provides little actual evidence that using the term "concept" impedes scientific progress. It thus fails to demonstrate that this term should be eliminated from the scientific vernacular.

The newly minted cognitive psychology student is likely to have been taught that there are three major models of categorization – prototype, exemplar, and theory – and that they are vying with one another for the title of one true theory. The better part of *Doing without Concepts* (Machery 2009) is spent revising this doctrine: rather than being mutually exclusive, the overwhelming evidence suggests that each of these strategies is used for determining category membership, depending on context.

From these modest conclusions Machery derives a fairly radical idea: What we have been referring to as "conceptual processing" is actually several distinct processes, so distinct that no scientifically meaningful relationship exists between them. Because these processes do not form a natural kind – they do not overlap at the cognitive or neural level – any further discussion of "concepts" as a unit is both incoherent and detrimental to the study of these processes. Thus Machery advances, in the final chapter, a proposal for *concept eliminativism*, where the notion of a concept must be banished from science entirely.

One would think that such a proposal requires some kind of evidence that these heuristic groupings generally prevent scientific advancement. Curiously, Machery provides virtually no support to this effect. He does mention, in passing, a few instances where scientific taxonomies have undergone restructuring based on deference to natural kinds (as when chemical classifications shifted to the periodic table), but this is quite separate from demonstrating that theories reflecting natural kinds are requisite for scientific progress. In any case, examples of such sea changes in psychology are rare. Machery cites "memory" and "emotion" as two terms which psychologists have successfully done away with, a claim we find puzzling. The idea that emotion contains "few scientifically relevant properties that are common to all emotions" (2009, p. 238) is far from the standard view – if anything, it is the division of emotions into ever-smaller categories that meets with resistance (Nesse & Ellsworth 2009). Memory is an even worse example. Machery asserts that the umbrella term "memory" has been tidily replaced with a suite of finer-grained labels like "working memory" and "explicit memory." In fact, psychologists debate today not only where to carve memory into its constituent parts, but also whether to disambiguate memory at all. In one recent review, Jonides and colleagues concluded that the longstanding division between long-term and short-term memory is artificial, owing to the fact that the same mechanisms are involved in encoding, maintenance, and retrieval in both systems (Jonides et al. 2008). Here is a case where overzealous compartmentalizing has led scientists to overlook alternate possibilities that were evident in the available data.

Of course, the memory case cannot directly inform whether "concept" should be discarded in favor of exemplar, prototype, and theory theories. But if a decades-long distinction between short-term memory and long-term memory has been overstated, then even the extensive evidence that Machery gathers in support of dividing concept into subcategories may erode after further study and review.

What of those psychological terms which were never divvied up into natural kinds – how have they fared? It may be instructive to look at research on concepts, which, by Machery's account, has been hobbled with an imperfect taxonomy for years. Fortunately, a survey of this research is conveniently located in the earlier chapters of his book, which is rich with experiments probing the nature of conceptual cognition. The arms race between camps of what were once considered competing theories appears to have led to a flourishing of work on this topic. If the unnatural kind-ness of "concept" has hindered research on conceptual processing, it certainly doesn't show in Machery's book.

There is another hurdle facing would-be converts to the eliminativist project, which is the delicate matter of what to do once the word "concept" has been cast from the psychological vernacular. Machery is happy to supply us with a more technically accurate replacement: "bodies of knowledge used by default in the processes underlying most higher cognitive competencies" (p. 239). It is not clear why Machery endorses this ungainly locution, which would seem to retain all the scientific baggage of "concept." And it is probably not a good sign that Machery himself can hardly go more than a page without resorting to the forsaken term.

Although Machery presents his project as a fundamentally pragmatic one, concept eliminativism does not seem particularly practical. It requires that we not only convolute our vocabulary, but also renounce our natural inclination to think of these processes as related on a functional level. A truly pragmatic revolution can hardly fail to take into account the minds that must work within the confines of the new paradigm. In this sense, concept eliminativism and old-fashioned eliminativism (which Machery takes pains to distance himself from) bear quite a bit in common: The proposals to eradicate inaccurate terminology, even if technically more precise, are so unwieldy that they would be unrealistic to adopt.

If science were performed by robots, eliminating concepts might well be a reasonable prescription. Of course, the degree to which these rough intuitions hamper scientific progress would still need to be demonstrated. In considering proposals such as Machery's, we must weigh the inconvenience of the newer, less fluent way of thinking against the advantages of doing so. In this case, the inelegance of the project is evident, and the benefits of adopting it are, as yet, untested and unknown. We therefore see no reason to jettison "concept" from the scientific discourse.

## Are prototypes and exemplars used in distinct cognitive processes?

doi:10.1017/S0140525X10000488

James Virtel and Gualtiero Piccinini

*Department of Philosophy, University of Missouri–St. Louis, St. Louis, MO 63121-4400.*

**jlv4z3@umsl.edu**

**piccininig@umsl.edu**

**www.umsl.edu/~piccininig/**

**Abstract:** We argue that Machery provides no convincing evidence that prototypes and exemplars are typically used in distinct cognitive processes. This partially undermines the fourth tenet of the Heterogeneity Hypothesis and thus casts doubts on Machery's way of splitting concepts into different kinds. Although Machery may be right that concepts split into different kinds, such kinds may be different from those countenanced by the Heterogeneity Hypothesis.

In *Doing without Concepts* (2009), Machery's argument that concepts split into different kinds is bold and inspiring but not fully persuasive. We will focus on the lack of evidence for the fourth tenet of Machery's Heterogeneity Hypothesis (HH), according to which, "prototypes, exemplars, and theories are typically used in distinct cognitive processes" (Machery 2009, p. 4).

Machery proposes three types of evidence that may support his fourth tenet (p. 124). If any of the following is shown for two kinds of concept, then it is likely that the two kinds of concept are used in distinct cognitive processes:

1. The neural systems implementing the cognitive processes that use the two kinds of concept are doubly dissociable.
2. The cognitive processes that use the two kinds of concept exhibit a difference in their input-output functions.
3. The cognitive processes that use the two kinds of concept do so by means of different algorithms.

We accept these three criteria with one exception pertaining to Criterion 2: While we agree that a difference in outputs is evidence of distinct processes, we deny that a difference in the inputs alone is good evidence of distinct processes.

Machery maintains that an input difference in categorization, for example, categorizing some items by means of prototypes and other items by means of exemplars, is evidence enough for his fourth tenet (Machery 2009, p. 124). Assuming for the sake of the argument that some items are indeed categorized by means of prototypes and others by means of exemplars, this shows only that we possess both prototypes and exemplars and use both in categorizing. Given this evidence, it may be that prototypes and exemplars are used in *distinct* processes, but it may also be that both prototypes and exemplars are used in the *same* process. Neither possibility is favored by an input difference. With this caveat in place, we argue that Machery does not fulfill any of his three criteria with respect to prototypes and exemplars. Some of our considerations go even further and raise doubts about Machery's splitting of concepts into prototypes, exemplars, and theories.

As to Criterion 1, Machery presents no evidence of doubly dissociable neural systems involving prototypes and exemplars. The only evidence he describes is of a *single* dissociation in amnesic patient E.P. (Machery 2009, p. 214). Since E.P. could not recognize previously seen items, E.P. was unable to add new exemplars to his long-term memory. But E.P. could still correctly categorize simple dot patterns in a way that suggests he used prototypes (Machery 2009, p. 215).

After citing E.P.'s case, however, Machery cites evidence that previous exposure to category members is not necessary to perform well in the dot pattern task used to test E.P. A similar performance may be obtained by relying solely on short-term memory (Palmeri & Flanery 1999). Thus, as Machery points out (2009, p. 217), E.P.'s performance does not show that E.P. categorized by means of prototypes in the absence of exemplars (a single dissociation). And even if that were shown, a double dissociation would also require an additional case in which exemplars are used without prototypes.

As to Criterion 2, Machery's best evidence comes from experiments in which subjects learn some new categories A and B and then categorize some new stimuli as either A or B (Malt 1989). Under some conditions, subjects appeared to categorize a new stimulus by comparing it to the old stimulus most closely resembling it. Malt interpreted this as exemplar-based categorization. Under other conditions, subjects appeared to categorize a new stimulus by determining which features it possessed among those that were typical of a category. Malt interpreted this as prototype-based categorization. Based on Malt's experiments, Machery concludes that people categorize some items using prototypes and others using exemplars, and he implies that the processes involved are distinct (Machery 2009, pp. 180–82). We reject Machery's conclusion for two reasons.

First, as we pointed out earlier, input differences alone are not good evidence of distinct processes. Hence, experiments such as Malt's do not support the fourth tenet of HH unless there are also output differences. Malt (1989) reports no output differences. She does report a priming effect that occurs only under the allegedly exemplar-based strategy. But the priming effect changes a subject's response time, not the output.

Second, Malt's experiments do not even show that, at least in the wild, subjects store both prototypes and exemplars properly so called (i.e., representations of particular objects) and use both in categorization. The stimuli in Experiments 1–3 were drawings of real animals from different species (see Malt 1989, Fig. 1). They depict what a typical member of a species looks like, without any features that would distinguish one particular animal from another. Such stimuli are too generic to provide convincing evidence that subjects store exemplars properly so called. By contrast, the stimuli in Experiments 4–6 were artificial categories (Malt 1989, Figs. 2 and 3) whose structure is too different from natural categories to warrant any firm conclusion about the ordinary process of categorization. Machery himself worries that "these experiments tap into ad hoc strategies only used by subjects to deal with abnormal learning and categorization conditions" (Machery 2009, p. 183).

As to Criterion 3, Machery points out that while both prototype-based models and exemplar-based models postulate that categorization involves a computation of similarity, the two classes of models are different in one respect. Prototype-based models usually employ a *linear* similarity measure, while exemplar-based models usually employ a *nonlinear* similarity measure. This may suggest that prototypes are processed following an algorithm that uses a *linear* similarity measure, while exemplars are processed following an algorithm that uses a *nonlinear* similarity measure.

But as Machery also points out, linear measures of similarity are not required for prototype-based models (2009, p. 90), and nonlinear measures of similarity are not required for exemplar-based models (2009, p. 98). Thus, the use of linear versus nonlinear measures does not determine whether an algorithm is prototype-based or exemplar-based. Therefore, there is no clear evidence that prototypes and exemplars are used in processes that follow different algorithms.

In conclusion, Machery has provided no convincing evidence that prototypes and exemplars are typically used in distinct cognitive processes. This lack of evidence is enough to at least partly undermine the fourth tenet of HH. The considerations we have presented are part of a larger set of doubts on Machery's way of splitting concepts into prototypes, exemplars, and theories. Machery may yet be right that there are different kinds of concept, but there might be a more fruitful way to split concepts into kinds than that postulated by HH.

Specifically, one of us (Piccinini) has argued that the two main kinds of concept are implicit concepts and explicit concepts. Implicit concepts encode information about a category in an implicit form that cannot be accessed directly by the language faculty, whereas explicit concepts encode information in an explicit form that can be manipulated by the language faculty (Piccinini, forthcoming; Piccinini & Scott 2006). This proposal fits with and may contribute to explain a wide range of evidence about implicit versus explicit cognition (Evans & Frankish 2009).

ACKNOWLEDGMENTS
Thanks to Edouard Machery and Andrea Scarantino for helpful comments.

## Doing with development: Moving toward a complete theory of concepts

doi:10.1017/S0140525X1000049X

Haley A. Vlach, Lauren Krogh, Emily E. Thom, and Catherine M. Sandhofer
*Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563.*
**haleyvlach@ucla.edu**
**laurenkrogh@gmail.com**
**emilyt0623@ucla.edu**
**sandhof@psych.ucla.edu**

**Abstract:** Machery proposes that the construct of "concept" detracts from research progress. However, ignoring development also detracts from research progress. Developmental research has advanced our understanding of how concepts are acquired and thus is essential to a complete theory. We propose a framework that both accounts for development and holds great promise as a new direction for thinking about concepts.

In *Doing without Concepts*, Machery (2009) provides a solid argument for how the current construct of "concept" has led to useless controversies. While agreeing that research on concepts needs to be refocused, we contend that Machery's proposal is only a small step towards a new framework for thinking about concepts. We suggest that a promising direction for concepts exists in research that is too often ignored – the acquisition, formation, and development of concepts.

Unfortunately, like many theories in cognitive science and philosophy, Machery's proposal largely ignores development. *Doing without Concepts* avoids issues that are central to a theory of concepts: how concepts are acquired, how conceptual structures change across development, and how concepts that are coordinated early in development become uncoordinated over time. In fact, Machery's proposal acknowledges the sentiment that development is less important than other areas of research on concepts (2009, p. 18). This blanket rejection of development is erroneous and dangerous – identifying the mechanisms by which concepts are acquired, how knowledge changes over time, and how cognitive processes give rise to such changes is essential to our understanding of how concepts operate and are organized. Because of the focus on such issues, many developmental researchers have taken a step back from the assumption that concepts originate from existing mental structures or representations.

One way that developmental research has advanced our understanding of concepts is by introducing the idea that concepts are formed in the moment (e.g., Gibson 1969; Samuelson et al. 2009; Smith et al. 1999; Spencer & Schöner 2003; Thelen & Smith 1994). By this account, performance on tasks is a reflection of the dynamic interaction between the learner and the learning environment – prior experiences recalled from long-term memory act together with task dynamics, perception, and action, to generate behavior. For example, one study (Samuelson et al. 2009) had children generalize novel nouns for rigid and deformable objects in two different tasks: a forced-choice task, in which children had to choose one of three objects that shared the same name with an exemplar, and a yes/no task, in which children had to respond whether each of the three objects shared the same name with an exemplar. Children's performance in the two tasks suggested conflicting conclusions about how children organized categories. Performance on the forced-choice task suggested that children treat rigid and deformable things differently when assigning labels: Rigid things are named by similarity in shape, whereas deformable things are named by similarity in material substance. However, performance on the yes/no task suggested that children did not distinguish between rigid and deformable things in naming and generalized names for both kinds by shape similarity. Given that children were all at the same developmental level and were presented with the exact same stimuli, differences in performance between the two tasks were likely not attributable to possessing different object concepts. Instead, children's object concepts were formed in the moment given the demands of the different tasks.

Machery notes that the feasibility of concepts being formed in the moment fails because there is not a significant body of research suggesting a "variability across contexts of the knowledge brought to bear on tasks" (2009, p. 22). However, Machery ignores countless examples of developmental research demonstrating great variability in the knowledge that is brought to bear in a particular task (e.g., Plumert 2008; Samuelson et al. 2009; Sandhofer & Doumas 2008; Siegler 1994; Smith et al. 1999; Thelen & Smith 1994; van Geert & van Dijk 2002; Vlach et al. 2008). This literature has demonstrated that performance on tasks is flexible (e.g., Sloutsky & Fisher 2008), context dependent (e.g., Samuelson & Smith 1998), and altered by seemingly minor changes in the conditions of the task (e.g., Sandhofer & Doumas 2008; Sandhofer & Thom 2006; Vlach et al. 2008). For example, altering the timing of exemplar presentation by a matter of seconds can lead to marked differences in children's performance on a generalization task (e.g., Vlach et al. 2008).

Machery also rejects the idea that concepts are formed in the moment, too quickly dismissing the proposal that there are not enduring mental structures for prototypes, exemplars, and theories. The book refers to this perspective as the "anti-representationalist argument" (2009, p. 222). However, developmental research has provided substantial evidence suggesting that variability and behavior can be explained without the presence of a mental structure or representation (e.g., Samuelson et al. 2009; Smith et al. 1999; Spencer & Schöner 2003; Thelen & Smith 1994). For example, Piaget (1954) originally attributed errors in the A-not-B task to infants' lack of an "object concept" (i.e., an enduring mental structure about the spatial and temporal consistencies of objects). Over the last few decades, research has demonstrated that errors on this task result from factors other than a lack of an object concept (e.g., Smith et al. 1999), such as visual input (e.g., Butterworth et al. 1982), direction of gaze (e.g., Horobin & Acredolo 1986), and memory (e.g., Diamond 1985). Moreover, performance on this task can be explained by the interaction of these processes alone (e.g., Smith et al. 1999). Altogether, this work has provided a compelling argument for how concepts and performance are shaped in the moment from a confluence of factors, rather than being accessed from an enduring mental structure.

To conclude, thinking about concepts needs a better direction. However, any theory of concepts would be remiss if it did not account for development. We propose that thinking about concepts should be situated within a dynamic framework that includes the learner and learning environment. Future research should explore the mechanisms by which concepts emerge in the moment – perception, action, and memory, interacting with properties of the environment, will reveal how this process works. This framework accounts for how concepts arise and change over time and thus holds great promise as the new direction for thinking about concepts.

## The theoretical indispensability of concepts

doi:10.1017/S0140525X10000506

Daniel A. Weiskopf

*Department of Philosophy, Georgia State University, Atlanta, GA 30302.*

**dweiskopf@gsu.edu**

**http://www2.gsu.edu/~phldaw/**

**Abstract:** Machery denies the traditional view that concepts are constituents of thoughts, and he more provocatively argues that concepts should be eliminated from our best psychological taxonomy. I argue that the constituency view has much to recommend it (and is presupposed by much of his own theory), and that the evidence gives us grounds for pluralism, rather than eliminativism, about concepts.

What are concepts? A long tradition in philosophy and psychology takes them to be the constituents of thoughts. In *Doing without Concepts*, Machery rejects this, defining them instead as bodies of knowledge stored in long-term memory and used by default in a range of higher cognitive processes (Machery 2009, p. 12).

Machery's arguments against the constituency view, however, are not compelling. He suggests that the notion of a constituent is ill-understood (2009, p. 26). But he also notes that the language of thought (LOT) hypothesis (Fodor 1975; 2008) gives us one fairly clear sense of what this might mean (Machery 2009, p. 27). So why *shouldn't* we adopt precisely this sense? Moreover, it is hard to understand many of his own claims about conceptual processing without appealing to constituency. Prototypes, exemplars, and theories are all complex representations that bear structural relations to their parts, over which inferences, similarity computations, and the like, might operate. He may wish to remain neutral on issues of what the vehicles of thought are, but connectionist and dynamical systems models of these phenomena are notably unpromising. The best candidates for bodies of knowledge that can fill the role he posits are ones organized, inter alia, by relations of constituency.

Constituency also plays a role in psychological explanation. Not everything about a category in long-term memory is or can be accessed in a single task; only some packets are extracted and used at once. Tokening a complex representation makes its constituents available to working memory for processing. And complex representations may make greater processing demands than simpler ones. The notion of a representational constituent is needed in describing what packets are retrieved from the vast reserves of long-term memory, and how these copies in working memory affect task performance.

So we can safely embrace the notion that concepts are constituents of thoughts. But I agree with Machery that these constituents are heterogeneous (*non-uniform*, in my terms). My own provisional list of types of concepts includes prototypes, exemplars, causal models, ideals and norms, and some perceptual and linguistic representations. This is central to the pluralist view of concepts I defend (Weiskopf 2009a; 2009b). While Machery and I agree on much of the empirical data, we disagree on its import. Where I see evidence for pluralism, he favors eliminativism. I suggest that we should be optimistic about the study of concepts as such.

Concepts are a functional kind, like most in psychology and neuroscience (Weiskopf, forthcoming). Consider how functional explanation proceeds. We decompose cognitive systems into a host of nested and interconnected subsystems, and populate them with representations, processes, and resources such as memory stores. This is obviously true in explaining competencies such as visual perception and numerical cognition, and it is no less true for concepts. Inductive and deductive reasoning, decision making, long-term planning, theory construction and testing, language use, and a host of "higher" capacities require explanation, and concepts are the representations, whatever they may be, that are proprietary to the system that underlies these capacities.

Machery argues that concepts have nothing in common beyond this functional description, and hence are not a "natural kind" in his sense (2009, pp. 243–44). But functional kinds are empirically discovered, and are posited in order to explain a (possibly open-ended) range of capacities that creatures possess. And concept possessors are strikingly different from creatures lacking concepts. They have a cognitive repertoire that is flexible – that is, sensitive to, but substantially independent of, ongoing perceptual input in terms of both content and processing – and that displays integration of information freely across domains (Weiskopf 2010). Concepts also explain the productive character of human thought in virtue of being able to combine open-endedly; Machery does not discuss productivity, but it is widely taken to be a central property of conceptual thought, and one that separates concepts from other types of representation.

The fact that a separate cognitive system is needed to explain these capacities is a discovery, not an a priori deliverance. Otherwise we could have predicted from the armchair that Skinnerian behaviorism and its modern descendents (e.g., Brooksian robotics) were doomed to fail. Instead, the limits of these models are demonstrated by their failure to capture the relevant phenomena. Concepts constitute a kind because positing them gives us the needed explanatory leverage over a wide range of creatures and their capacities. If we posit them, we simultaneously gain the ability to account for phenomena that would otherwise have been inexplicable, and to capture similarities among otherwise dissimilar creatures. This is how the functional kinds posited in models of cognitive systems typically earn their distinctive status. If a model containing a functional category $F$ has greater explanatory and unifying power than ones that lack it, then $F$ is prima facie a kind. The failure of models of human cognition that lack anything corresponding to concepts shows that they satisfy this condition.

But suppose we followed Machery's lead and eliminated the term "concept," talking only of prototypes, exemplars, and so on. These representations may occur in *many* cognitive subsystems. Visual perception may involve generating and storing such representations; hence we often find talk of "perceptual prototypes" in the psychological literature. Without distinguishing concepts as such, we would be unable to state the ways in which perceptual prototypes differ from their conceptual kin. Conceptual prototypes, as opposed to perceptual ones, are capable of free recombination with other conceptual representations, are capable of being generated for non-perceptual categories, and so forth. Talk of prototypes alone will not do this crucial taxonomic work for us. For this we need the theoretical notion of a concept.

Indeed, this explanatory need can be seen even in Machery's own definition of concepts as being involved in "higher" cognitive capacities. For what makes one capacity "higher" than another? A tempting answer is that the "higher" ones are just the concept-involving ones. The fact that we need to appeal to concepts even to isolate these various types of representations in the first place suggests that concepts will be an essential part of our taxonomy of psychological kinds. Happy news for pluralists, but not for eliminativists.

## Developing without concepts

doi:10.1017/S0140525X10000518

Yevdokiya Yermolayeva and David H. Rakison

*Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213.*

**yyermola@andrew.cmu.edu**

**rakison@andrew.cmu.edu**

**http://www.psy.cmu.edu/people/yermolayeva.html**

**http://www.psy.cmu.edu/people/rakison.html**

**Abstract:** We evaluate the heterogeneity hypothesis by considering the developmental time course and the mechanism of acquisition of exemplars, prototypes, and theories. We argue that behavioral and modeling data point to a sequential emergence of these three types of concepts within a single system. This suggests that similar or identical underlying cognitive processes – rather than separate ones – underpin representation acquisition.

*Doing without Concepts* (Machery 2009) proposes an interesting solution to the problem of applying the term "concept" to prototypes, exemplars, and theories, which, according to the author, are unrelated. Each type of concept engages a distinct cognitive process – such as similarity comparison or causal inference – so that a unified label is inappropriate. The book synthesizes an impressive amount of literature in psychology and philosophy to provide evidence for this heterogeneity hypothesis. From the point of view of developmental psychology, however, two key questions remain unanswered. First, what is the time course for the emergence of prototypes, exemplars, and theories? Second, and more importantly, what is the mechanism behind their formation? Specifically, does each require a dedicated mechanism, or is a single system sufficient? In our view, an answer to the second question is particularly important for our ability to evaluate the proposal that distinct cognitive processes underlie the use of prototypes, exemplars, and theories.

Answering the first question is an important component to answering the second question. If exemplars, prototypes, and theories emerge in succession and not simultaneously, then it is possible that they build upon each other. This could suggest the development of a single mechanism, or, at the very least, the development of three related mechanisms. While no single study provides definitive evidence, a pattern of successive emergence can be observed across studies. As an example, we can examine infants' knowledge about individuals. Three-month-old infants can discriminate an image of their mother's face from that of a stranger (Barrera & Maurer 1981), which suggests that they have stored an exemplar of their mother's appearance. By 6 months of age, infants can extract a prototype from a series of faces and display a preference for a novel face when it is presented with either a familiar face or the previously unseen prototype (Rubenstein et al. 1999). By 10 months, infants display more theory-like knowledge about individuals in that they do not generalize goal-directed actions, such as reaching for an object, from one individual to another (Buresh & Woodward 2007). This task requires not only theoretical knowledge about the properties of goals, but also the ability to store exemplars of the individuals so that goals may be matched correctly. Taken together, these studies provide some support for the sequential emergence of exemplars, prototypes, and theories.

In addition to determining the time course for these processes, the most important developmental question with respect to the heterogeneity hypothesis relates to the mechanism of acquisition. Do dedicated mechanisms exist for of prototypes, exemplars, and theories, or are all three acquired within the same system? In our view, the literature points to the latter. We suggest that if the underlying mechanism of acquisition is shared, then entirely distinct cognitive processes do not underlie different types of concepts.

According to Machery, prototypes and theories have little in common: the former involves extraction of the statistics of a category, while the latter involves causal inference. However, as has been suggested by Sobel and Kirkham (2007), statistical learning is involved in the emergence of causal reasoning. In the backwards blocking paradigm, children's and adults' judgments about an object's causal effectiveness are influenced by prior knowledge about frequencies with which causal and noncausal objects are present in the environment. Sobel and Kirkham argue that causal knowledge involves probability distributions and likelihoods of particular hypotheses – to reason causally, children must have the ability to extract statistical regularities from the environment. The formation of prototypes and theories, then, relies on processing of statistical information.

Work in computational modeling provides additional evidence for a shared mechanism by demonstrating that reasoning based on similarities (using exemplars or prototypes) and reasoning based on theories does not require separate architectures. Chaput and Cohen (2001) used hierarchical self-organizing maps to model changes in infants' perception of simple collision events in which one ball causes another to move. Studies have shown that younger infants respond to such events based on temporal or spatial similarity. Older infants respond based on causal features of events: Noncausal events with a temporal gap in the sequence are perceived to be the same as those with a spatial gap, and different from continuous causal events. Chaput and Cohen (2001) produced a model in which the intermediate layers responded based on temporal and spatial components early in training, much like younger infants. As training progressed, the top layer integrated these components and began to respond based on causal information.

Similarly, Verguts and Fias (2009) used modeling to demonstrate that similarity and rule-based responding can be thought of as lying on the same continuum. Similarity judgments are made based on many readily perceivable features; rule judgments are made based on fewer internally generated features. The model replicated human performance on a prediction task in which participants who received little training used similarity to observed cases to make predictions, and those who received more training used rules. With an increased number of training trials, the model progressed from making similarity judgments using the components of the input to making rule judgments by extracting regularities among components. Taken together with Chaput and Cohen's (2001) work, these findings suggest that separate mechanisms are not necessary for the emergence of prototypes, exemplars, and theories, and that theories can emerge through the reorganization of similarity information within the same system.

The proposal that exemplars, prototypes, and theories are underwritten by distinct processes is a convenient way to account for the conflicting psychological data on concepts. However, an examination of the developmental literature is necessary for the evaluation of this proposal. Behavioral and modeling studies suggest that exemplars, prototypes, and theories develop sequentially, and can do so within the same system without the need for three dedicated mechanisms. In our view, if the mechanism of acquisition is shared, then the cognitive processes underlying prototypes, exemplars, and theories must be partially, if not completely, overlapping, casting doubt on the heterogeneity hypothesis. From the developmental perspective, the three are not so distinct, and "doing without concepts" may be unnecessary.

## Parsimony and the triple-system model of concepts

doi:10.1017/S0140525X10000531

Safa Zaki[a] and Joe Cruz[b]

[a]*Department of Psychology and Program in Cognitive Science, Williams College, Williamstown, MA 01267;* [b]*Department of Philosophy and Program in Cognitive Science, Williamstown, MA 01267.*

**szaki@williams.edu**

**http://www.williams.edu/Psychology/Faculty/Zaki/zaki.html**

**jcruz@williams.edu**

**http://www.williams.edu/philosophy/fourth_layer/faculty_pages/jcruz/jcruz.html**

**Abstract:** Machery's dismissive position on parsimony requires that we examine especially carefully the data he provides as evidence for his complex triple-system account. We use the prototype-exemplar debate as an example of empirical findings which may not, in fact, support a multiple-systems account. We discuss the importance of considering complexity in scientific theory.

In *Doing without Concepts*, Machery (2009) postulates a novel multiple-system theory of concepts: a triple-system theory. That is, he proposes a model that has a prototype process, a theory process, and an exemplar process. He cites experimental evidence in favor of each of these component systems and argues that none of them alone is sufficient to account for extant experimental data. In defense of the manifest complexity of the model, Machery argues that parsimonious theories are only to be preferred if they have historically "won out" in a particular domain:

> [O]ne can take parsimonious theories as more likely to be empirically supported than less parsimonious theories if and only if in the past, parsimonious theories have been better supported than less parsimonious theories. (Machery 2009, p. 126)

In his view, the unified, parsimonious view of cognition does not possess this historical advantage, so a lack of parsimony is no challenge to his multi-system theory.

We find Machery's treatment of parsimony odd. If a single-process theory does a credible job in matching a multiple-systems theory, it ought to be favored on the basis of parsimony. Machery, however, has closed off this appeal. His historical presumption makes it impossible for scientifically meritorious simple theories to triumph if they are late-comers, because by then, presumably, there would exist data sets that at least equally favor complex models. After all, by their nature, complex theories will be able to accommodate more of these data points than less complex models. Therefore, all that the complex models have to do is get there first.

Of course, the serious downside to the ease with which a complex system can accommodate data is the loss of predictive ability. In our view, this is the point of privileging parsimony. Evidently Machery maintains that a multiple-systems theory is to be assessed solely in terms of data in its favor, regardless of complexity. To his credit, he canvases an impressive number of studies in the field. But because there is no presumption in favor of parsimony and therefore in favor of single-system accounts, we are inclined to look closely at the data that he uses as evidence of the several component processes in category learning. In at least one area, Machery erroneously concludes that the data favor his triple-system theory over a single-system account.

Machery cites a series of articles by Smith and Minda to support the idea that exemplars are insufficient (Minda & Smith 2001; Smith & Minda 1998; Smith et al. 1997). In these articles, Smith and Minda presented evidence that they claimed challenged the predictions of the exemplar-based models of classification and that supported predictions of prototype models. The basis for these researchers' claims was that the prototype models provided better quantitative fits to certain sets of individual subject categorization data than did the context model. However, Nosofsky and Zaki (2002) subsequently noted that in all of these quantitative-fit comparisons, Smith and colleagues considered the predictions from only a constrained version of the exemplar model which did not allow the model to capture the levels of deterministic responding that are often evidenced by individual subjects (e.g., Maddox & Ashby 1993). When a response-rule parameter was allowed to vary in the model, an exemplar model consistently outperformed the prototype model. Machery does not cite the debate that followed the original Smith and Minda articles and therefore leaves readers with the incorrect impression that the exemplar-prototype debate was settled in favor of a model that required both representations.

Similarly, in Chapter 6, Machery cites a series of dot-pattern studies (Smith 2002; Smith & Minda 2001; 2002) that he claims provide support for the existence of a prototype process. In these studies, Smith and Minda fit models to data from a particular version of the classic dot-pattern paradigm (Knowlton & Squire 1993). In the Knowlton-Squire version of the dot-pattern paradigm, during a study phase, participants are shown high-level distortions of a single pattern of nine dots arranged in a fixed shape. In a test that follows, participants judge the category membership of various new dot patterns that were derived from this category prototype but distorted at different levels (i.e., the prototype itself, low-level distortion, high-level distortions, and random patterns). The classic result is that the prototype is classified as a member of the category with the highest probability, followed by the low distortions, high distortions, and random patterns. Although both exemplar and prototype models predict the ordering of this typicality gradient, Smith and Minda claimed that only the prototype model could capture that steepness of the gradient. However, Zaki and Nosofsky (2004; 2007) demonstrated that this steepness was at least in part attributable to confounded properties of the stimulus set. This paradigm was simply not diagnostic in terms of telling apart the models. Machery's claim that these dot-pattern studies in some way provided evidence for the existence of a prototype system and an exemplar system is simply not warranted.

We do not claim that a single-system exemplar account of the data is the correct account of concept learning. Indeed, the idea that observers might use different systems to represent categories is highly plausible. And we note that we have only addressed a small portion of the data in Machery's book. In evaluating evidence, we would, however, prefer to see a more careful treatment of the cost of complexity. Is a more complex model warranted by the data? We have no doubt that a triple-process or even a quadruple theory of categorization (see Machery 2009, p. 118, for a candidate fourth system) would predict a vast number of phenomena in the field. After a certain point, however, the important question is what would a complex model *not* predict? Many researchers are moving in a direction where models are penalized for being more complex (e.g., Myung et al. 2000; Myung & Pitt 2009; Navarro et al. 2004) by developing sophisticated metrics of complexity that go beyond traditional approaches of simply penalizing models for additional degrees of freedom. Machery seems to be pulling us in the other direction, and we should resist.

# Author's Response

## The heterogeneity of knowledge representation and the elimination of *concept*

doi:10.1017/S0140525X10000932

Edouard Machery
*Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh PA 15260.*
**Machery@pitt.edu**
**www.pitt.edu/~machery/**

**Abstract:** In this response, I begin by defending and clarifying the notion of concept proposed in *Doing without Concepts* (Machery 2009) against the alternatives proposed by several commentators. I then discuss whether psychologists and philosophers who theorize about concepts are talking about distinct phenomena or about different aspects of the same phenomenon, as argued in some commentaries. Next, I criticize the idea that the cognitive-scientific findings about induction, categorization, concept combination, and so on, could be explained by positing a single kind of concept, and I insist that many categories (substances, types of events, etc.) are represented by distinct coreferential concepts that belong to very different kinds of concept. This is followed by an assessment of the hybrid theories of concepts offered by commentators, according to which categories, substances, and types of events are represented by hybrid concepts made of several parts. Finally, I defend the proposal that it may be useful to eliminate *concept* from the theoretical vocabulary of psychology.

### R1. Introduction

While writing *Doing without Concepts* (Machery 2009; henceforth, *DwC*), I had several goals. One of them was to clarify the notion of concept used in cognitive science and to regiment its use. I also wanted to put an end to useless controversies between philosophers and psychologists about what concepts are. Even more important, I wanted to make a case for a picture of knowledge representation that has emerged from 30 years of cognitive-scientific research on the topic. According to that picture (which I called the *Heterogeneity Hypothesis*), cognitive competencies are often subserved by several distinct processes: There are many ways to categorize, draw inductions, and so forth. These processes are defined over distinct kinds of concepts, which have very little in common. Thus, the class of concepts divides into several heterogeneous kinds. Finally, I proposed a radical reform: It may be useful for cognitive scientists to eliminate the theoretical term *concept* from their theoretical vocabulary.

Judging by the commentaries elicited by the book and the Précis, it seems that the scientific community in cognitive science shares some of these views, although few seem willing to accept the whole picture I proposed. In this response, I defend this picture. I am grateful for the thoughtful and challenging commentaries, which have highlighted some significant limits of the argument made in *DwC*, attracted my attention to some literature I unfortunately ignored while writing this book, and cast some doubts on the strength of particular arguments put forward in support of my views.

Here is how I proceed in my response. After providing some clarifications in section R2, I turn to the characterization of the notion of concept used in cognitive science in section R3 and then examine the relation between the philosophy and the psychology of concepts in section R4. In section R5, I respond to the commentaries that defend the received view – that is, the view that there is only a single kind of concept (e.g., exemplars or prototypes). In section R6, I examine whether concepts that belong to different kinds (specifically, prototypes, sets of exemplars, and theories) can be coreferential. Section R7 focuses on the view that the class of concepts is even more heterogeneous than pictured in *DwC*. In section R8, I defend my criticism of hybrid theories of concepts, before examining in section R9 three responses to my proposal to eliminate *concept* from the theoretical vocabulary of cognitive science.

### R2. Clarifications

Before addressing the substantive criticisms made by the commentators, I should briefly clarify a few misunderstandings. **Hayes & Kearney** misinterpret my concerns with the term *concept*. I do not recommend that *concept* be eliminated from cognitive scientists' classificatory scheme because of the context-sensitivity of knowledge retrieval from long-term memory. (By contrast, **Vlach, Krogh, Thom, & Sandhofer** [Vlach et al.] recommend eliminating *concept* on precisely these grounds – more on this in section R3.) Indeed, I concur with Hayes & Kearney that behavioral and neuroimaging evidence shows that, while knowledge retrieval is indeed context-sensitive, some knowledge is also retrieved in a context-insensitive manner (Machery, forthcoming; see discussion in sect. R3), as I explained in two distinct places of *DwC* (sects. 1.4.1 and 8.1.1). In fact, I propose to identify concepts with those bodies of knowledge that are retrieved in a context-insensitive manner. I also agree with Hayes & Kearney that models of cognitive processes that assume the existence of such context-insensitive bodies of knowledge are compatible with the contextual variation observed in experimental tasks. So, what are, for me, the grounds for eliminating *concept*? Instead of being concerned with flexibility and context-sensitivity, I suggest eliminating *concept* because, failing to pick out a natural kind, this term does not fulfill its classificatory function (see sects. R9.1 and R9.2), and because keeping this term would have some important drawbacks and very few benefits (see sects. R9.4 and R9.5).

**Vlach et al.** object to my criticism of the anti-representationalist approaches in cognitive science, but they seem to misunderstand the claim made by these approaches (illustrated, e.g., by Rodney Brooks's work and by some versions of the dynamical systems theory), namely, that cognitive science should explain behavior without ascribing to organisms representations of their environment. Thus, contrary to what Vlach et al. seem to believe, anti-representationalist approaches to cognitive science do not merely reject the idea of enduring representations – namely, representations stored in long-term memory; they reject the idea that organisms have *any kind of* representations, *including representations constructed on the fly.*

## R3. What are concepts?

One of the goals of *DwC* was to clarify and regiment the slippery notion of concept used in cognitive science. According to the characterization C (introduced and defended in Chapter 1), a concept of *x* is a body of knowledge about *x* that is used by default in the processes underlying our higher cognitive competencies when these result in judgments about *x*. **Blanchard**, **Lalumera**, **Margolis & Laurence**, **Markman**, **Rey**, **Vlach et al.**, and **Weiskopf** raise some questions about this characterization of the notion of concept.

**Lalumera** objects that C is an incomplete characterization of the notion of concept in cognitive science, but her own discussion belies this claim. Although the notion of a higher cognitive competence was left vague, it is clear that the examples of competencies that, according to her, I ignored (viz., generalization and induction) straightforwardly fall under this notion. Indeed, induction is one of the examples I have repeatedly used to illustrate what higher cognitive competencies are.

**Blanchard** doubts that the notion of default is an appropriate way of cashing out the notion of concept used in cognitive science. Noting that many experiments meant to support the theory theories of concepts are not run under time pressure, he concludes that either theory theorists do not view their work as bearing on concepts – which, I agree with Blanchard, is dubious – or the notion of default fails to capture the notion of concept they use.

**Blanchard**'s objection is an excellent occasion to clarify the notion of default (see also Machery, forthcoming). On my view, three properties are characteristic of the bodies of knowledge retrieved by default: speed, automaticity, and context-independence. The bodies of knowledge retrieved by default come to mind quickly; their retrieval does not depend on one's intentional control (although they may also be intentionally retrieved); and they are retrieved in all contexts. Of these three properties context-independence is the essential one, while speed and automaticity are likely effects of context-independence: Because a body of knowledge is retrieved in a context-insensitive manner, its retrieval from memory might be simpler and thus faster, and it can be automatized. On my view, being retrieved quickly – for example, being retrieved under time pressure in an experimental context – is evidence, but not a necessary condition, for being retrieved by default. Thus, experimental tasks that do not involve time pressure can be used to examine people's concepts.

**Rey** criticizes my alleged identification of concepts with the bodies of knowledge used under time pressure (see my response to **Blanchard** above), and he proposes to identify concepts with those bodies of knowledge we would use if we were to reflect. However, if we were to follow this proposal, psychologists working on concepts would be unable to explain how people categorize, make inductions, understand words, and so on, in most circumstances.

Like Barsalou (1993) and Prinz (2002), **Vlach et al.** hold that concepts are temporary constructs in working memory on the grounds that there is "a great variability of the knowledge that is brought to bear in a particular task." I concur with Vlach et al. that the knowledge brought to bear on a given task is influenced by context; indeed, I explicitly acknowledge this variability in *DwC*, and the useful references they give in their commentary provide further evidence of this variability. However, as shown in *DwC* and in the Précis, several remarkable behavioral and neuroimaging studies also support the claim that some bodies of knowledge are retrieved by default (Barsalou 1982; Weisberg et al. 2007; Whitney et al. 1985). Vlach et al. do not address this body of evidence and provide no reason to doubt the conclusion it seems to support.

Other findings provide further evidence for the claim that some bodies of knowledge are retrieved by default. Although their goal was to highlight the context-sensitivity of knowledge retrieval, Hoenig et al. (2008) have shown that brain areas involved in representing actions or physical manipulations are automatically activated in a feature verification task when the question concerns the appearance of a tool, while, conversely, the brain areas involved in representing visual appearances are automatically activated when the question concerns the movement associated with a natural object. James and Gauthier (2003) have similarly shown that auditory areas and areas dedicated to the processing of movements are activated by the mere visual presentation of novel objects (greebles) when participants have respectively been trained to recognize the sounds of these objects and their movements. In both experiments, knowledge that is not relevant to the tasks at hand is accessed in a context-insensitive manner.

Furthermore, the variability of the knowledge used in different contexts is consistent with the fact that knowledge retrieval from long-term memory is partly context-insensitive. Indeed, **Hayes & Kearney** discuss several models that account for the variability while assuming that some bodies of knowledge are retrieved by default (see also sect. 1.4.1 of *DwC*).

Both **Margolis & Laurence** and **Weiskopf** concur that defining concepts as the constituents of thoughts provides a better characterization of the notion of concept than C. Before discussing this proposal, note that, contrary to what Weiskopf claims, I do not deny that concepts *are* the constituents of thoughts. Rather, I insist that concepts should not be *defined* as the constituents of thoughts (sect. 1.4.3 of *DwC*).

**Margolis & Laurence** and **Weiskopf** propose to clarify the notion of a constituent of thought by appealing to the language-of-thought hypothesis. However, it would be a mistake to build an empirical hypothesis as controversial as this in the very notion of concepts since the validity of the work of cognitive scientists working on concepts (e.g., their account of the bodies of knowledge retrieved when we categorize or draw an induction as well as their accounts of the categorization or induction processes) does not hang on the truth of the language-of-thought hypothesis.

Finally, there is little benefit to defining concepts as constituents of thoughts since this definition does not cast light on cognitive scientists' research. While **Margolis & Laurence** mention numerous psychologists who do characterize concepts as constituents of thoughts (Carey, etc.), they do not show that this characterization plays any role in these psychologists' actual empirical work on concepts.

**Weiskopf** mentions the need to explain the productivity of human thought (see also Hill 2010). This suggests that he does not simply define concepts as constituents of thoughts. Instead, for him, a body of knowledge is a concept only if it can combine *freely* with *any other* body of knowledge (i.e., only if it meets Evans's [1982] generality constraint). First, it is not clear that productivity is among the explananda of a psychological theory of concepts, but I will not press this point here (Machery 2010). Rather, I propose that Weiskopf's definition is arbitrary. If concepts are defined this way, organisms that are able to combine *all* bodies of knowledge freely (as humans do according to Weiskopf) have concepts, whereas organisms that are able to combine *many* such bodies, *but not all*, do not, even if the bodies of knowledge in the former and latter kinds of organism are very similar.

Here is another way of illustrating the arbitrariness of **Weiskopf**'s proposal. Suppose we humans have some bodies of knowledge that are used in categorization, induction, and the like, in almost the same way as our other bodies of knowledge. For example, they could be prototypes, and they could be used in similarity-based processes. But, while the other bodies of knowledge can be combined freely, these prototypes can only be combined with a limited number of other prototypes (perhaps because they belong to a modular cognitive system). Weiskopf would conclude that these prototypes are not concepts since they cannot combine freely with all the other bodies of knowledge, but this is the wrong conclusion to draw. Instead, the similarities between these prototypes and the other bodies of knowledge should lead us to conclude that they are concepts and to reject Weiskopf's proposal that a body of knowledge counts as a concept only if it can combine with every other body of knowledge.

**Rey** argues that my characterization of concept entails that two individuals or even a single individual at two different times cannot have the same concept. However, it is incorrect that a single individual at two different times cannot have the same concept, since a given concept can remain the same although its parts (viz., the elements of knowledge that are by default retrieved from long-term memory) change, exactly as objects remain the same despite their parts changing (Machery 2010). It is true that, given my characterization of concept, different individuals are likely to have different concepts (e.g., different concepts of dogs), but I think that this is a clear virtue, since this explains why they categorize differently and make different inductions.

Finally, **Markman** argues that it is a mistake to characterize psychological constructs, such as the notion of concept, by means of particular cognitive competencies and experimental tasks since adequate psychological constructs should explain performance in a range of tasks; and he regrets that I did not pay attention to a broader range of competences and tasks. I agree with Markman's argument. Although *DwC* examined only three cognitive competencies and focused on a few psychological tasks, the proposed characterization of concept – namely, C – is consistent with Markman's point.

## R4. Philosophers and psychologists on concepts

Another goal of *DwC* was to show that, although philosophers and psychologists use the term *concept* and develop theories of concepts, psychological and philosophical theories of concepts are really about different things. As a result, many criticisms of psychological theories of concepts by philosophers (and vice versa) are empty. **Edwards**, **Margolis & Laurence**, **Rey**, and **Schneider** disagree with this claim. **Keil** also seems to assume that philosophical and psychological theories of concepts are about the same thing, but does not press this point.

**Edwards** and **Margolis & Laurence** are right to claim that the fact that psychologists and philosophers have different explanatory interests does not entail that they are theorizing about different things, because they could be theorizing about different aspects of the same phenomenon.

I deny the univocity of *concept* across philosophy and psychology on the following grounds. Generally, in philosophy of science, a candidate explication of a scientific term (e.g., *ether*, *force*) is taken to be a failure if it entails that what scientists say when they use this term is typically mistaken. This is in fact a commonsense idea: If my interpretation of what someone means by a particular word entails that most what she says when she uses this word is false, my interpretation is probably mistaken. Applied to the notion of concept in cognitive science, a satisfactory explication of this notion should not entail that psychologists' claims about concepts are literally false. C, but not **Margolis & Laurence**'s characterization, meets this constraint. (A similar objection applies to **Rey**'s and **Edwards**'s characterizations.) Since Margolis & Laurence propose that the bodies of knowledge psychologists have been focusing on (viz., prototypes, exemplars, etc.) belong to the processing structure of concepts (whatever that is), they are bound to claim that psychologists are literally mistaken when they say that concepts *are* prototypes or exemplars (etc.).

**Margolis & Laurence** take the mutual influence of philosophers and psychologists to be evidence that psychologists and philosophers are talking about the same thing when they theorize about concepts, but I am not swayed by this argument. First, I am more impressed by how often psychologists and philosophers talk past each other when they exchange arguments about concepts than by how useful these exchanges have been. Second, the extent to which philosophers and psychologists have fruitfully influenced each other is perfectly consistent with the idea that they do not theorize about the same thing when theorizing about concepts. Cell biologists and physicists working on quantum mechanics do not theorize about the same thing, but the former are usefully appealing to the theories and findings of the latter (e.g., Collini et al. 2010).

Because I hold that philosophers and psychologists who theorize about concepts are not talking about the same thing, I find **Edwards**'s and **Rey**'s critiques puzzling. Clearly, it is important to understand better how concepts refer, but this is not one of the explananda of the theories developed by psychologists theorizing about concepts. Furthermore, as argued in section 2.4 of *DwC*, the kind of arguments exchanged by philosophers, which often rely on intuitions about what concept an individual in a particular circumstance possesses or fails to possess, or about whether two individuals have the same belief (e.g., Burge 1979), seems unlikely to be appropriate for determining how concepts refer.

**Schneider** notes that psychologists' findings about how we categorize, draw inductions, and so forth, can be coopted by some philosophical theories of concepts – those that individuate concepts by means of the functional role of mental symbols. While correct, this observation does not undermine the claim that psychologists and philosophers tend to theorize about different things when they theorize about concepts. Similarly, while cell biologists use biochemists' findings, biochemists and cell biologists are not developing theories about the same processes.

## R5. Rejecting the heterogeneity of the class of concepts

To my surprise, few commentaries objected to the claims that there are several kinds of concepts, and that prototypes, exemplars, and theories are among these kinds. **Strohminger & Moore** and **Rey** seem even to have found this claim entirely unsurprising. However, I worry that this lack of resistance is due to a confusion between two different claims: (1) that prototypes, exemplars, and theories are three distinct kinds *of concept*; and (2) that our long-term memory includes prototypes, exemplars, and theories. These two claims are not identical since one could grant that we have, say, exemplars and theories, but insist that concepts are prototypes (e.g., Hampton 2001) or that we have prototypes and exemplars, but insist that concepts are really theories. While Claim 2 is indeed not very controversial, Claim 1 is less obviously correct.

**Hayes & Kearney**, **Harnad**, and **Hampton** are the only ones to raise doubts about the heterogeneity of concepts. Harnad rejects the idea that the bodies of knowledge we use in higher cognitive tasks are prototypes or exemplars. Rather, sensorimotor processes play a central role in cognition: They enable living creatures to navigate the world by making object identification and appropriate action possible. There is no doubt that such sensorimotor processes exist since recognition must involve them, but these processes cannot underwrite the bulk of cognition. Like behaviorists and like some roboticists (see Machery 2006b, on Harvey), Harnad underestimates the complexity of higher cognition when he proposes that the processes that can explain perceptual and motor processing can scale up to higher cognition. Appealing to words, as Harnad does, does not help much since words have to be understood and mapped onto bodies of knowledge in long-term memory.

**Hampton** contends that the distinction between exemplars and prototypes breaks down for at least some categories since the exemplars of superordinate categories, such as the category of vehicles, are prototypes of subordinate categories, such as the category of cars (see also Malt, forthcoming). However, superordinate categories could be represented by representations of *particular* cars, planes, bikes, and so on, rather than by prototypes of cars, plans, bikes, et cetera. This is naturally an empirical question, and Hampton might well be right. Supposing he is right, it would nonetheless be a mistake to say, as he does, that the *exemplars* of superordinate categories are *prototypes* of the subordinate categories. Rather, using the term *exemplar* to refer only to representations of *individuals*, one should say that superordinate categories are represented by sets of prototypes of subordinate categories.

In any case, **Hampton**'s interesting comment brings to the fore a shortcoming of *DwC* – the neglect of superordinate concepts and the focus on basic-level categories (e.g., dogs). It might be that the heterogeneity of conceptual representations is even larger than proposed in *DwC*, with superordinate categories being represented differently from basic-level and subordinate categories. Importantly, far from undermining the crucial message of the book, this outcome would reinforce it: It is hopeless to look for a theory of conceptual representations that applies to all default bodies of knowledge.

**Hayes & Kearney** defend the received view. They contend that exemplar models of a range of phenomena outperform prototypes models (see also **Zaki & Cruz**), and that the role of theories can naturally be included in exemplar models (on this latter point, see also **Yermolayeva & Rakison**). But some of the studies they themselves cite in fact undermine this claim. Far from providing evidence that categorization can be explained by means of a single kind of default bodies of knowledge, Allen and Brooks (1991) provided evidence that, in at least some circumstances, we have two distinct representations of a single category (a rule and a set of exemplars) and that these representations can lead to conflicting categorization judgments. Smith et al. (1998) have replicated these findings, and provided evidence that two neural networks are involved in each categorization judgment. Furthermore, exemplar models seem able to explain the empirical findings about concepts only if one takes exclusively into consideration the category-learning studies that involve artificial stimuli such as patterns of points. Research on the knowledge of expert physicians (reviewed in Norman et al. 2006) shows, for instance, that during their training in medical school, physicians acquire different types of bodies of knowledge that are largely independent from one another.

Further, for **Hayes & Kearney**, the role of theories in cognition is indirect: They influence which exemplars are learned in category-learning tasks or which exemplars are retrieved from long-term memory (which explains perhaps why Hayes & Kearney prefer to speak of prior knowledge instead of theories). However, a less partial review of the literature suggests that the use of theories is not so limited, and that causal theories are directly used to categorize and to make inductions (Chs. 6 and 7 of *DwC*).

Finally, in *DwC*, I argued that simplicity can be used to choose between scientific hypotheses in a domain of inquiry (e.g., in psychology) only if past evidence inductively supports the belief that in this domain simpler hypotheses tend to be better supported than more complex hypotheses. By contrast, **Zaki & Cruz** argue that simplicity should always be preferred on the grounds that models with more free parameters fit better a given set of data points than models with less free parameters; and they conclude that, everything else being equal, we should prefer a theory that posits a single kind of concept to a theory that posits several distinct kinds of concept. Although the question cannot be resolved in a few words, their argument should be resisted for two reasons. First, it is not necessarily the case that simpler models fit better than more complex ones when fit is evaluated by cross-validation since models with more parameters can overfit. Second, model fitting provides a poor, albeit common (e.g., Forster & Sober 1994), analogy for understanding the use of simplicity as a criterion for theory choice. More complex theories are not necessarily better supported than simpler theories since they often have empirical implications that simpler theories simply do not have. For instance, the heterogeneity hypothesis – but not (e.g.) prototypes theories – predicts that in at least some cases people's categorization (or induction) judgments are going to be slower or less reliable because, for example, the prototype-based and theory-based categorization (or induction) processes conflict with one another (for consistent evidence, see Allen & Brooks 1991; Kulatanga-Moruzi et al. 2001; Regehr et al. 1994; Smith et al. 1998; and see sects 5.1.5, 6.6, and 7.1.5 of *DwC*).

## R6. Can prototypes, exemplars, and theories be coreferential?

While I proposed that we often have several coreferential concepts – for example, we might have a prototype of dogs, a set of exemplars about dogs, and a theory of dogs – **Hampton**, **Jacobson**, and **Margolis & Laurence** doubt that prototypes, exemplars, and theories can corefer.

In *DwC*, I intentionally refrained from proposing a theory of how bodies of knowledge such as prototypes, sets of exemplars, and theories denote (see **Edwards**'s and **Rey**'s commentaries), and this is not the place to propose one. However, for present purposes, it is sufficient to note that, according to several influential theories of reference, a prototype, a set of exemplars, and a theory can be coreferential. Consider, for instance, Fodor's informational semantics (e.g., Fodor 1990). According to this view, roughly, a concept refers to the property that it is nomologically linked to. As argued convincingly by Prinz (2002), informational semantics can naturally be combined with prototype theories: A prototype of dogs refers to dogs because its occurrence (i.e., its retrieval from long-term memory) is nomologically linked to the presence of dogs in the cognizer's environment. Informational semantics can similarly be combined with theory theories and with exemplar theories. Thus, it is perfectly possible for a prototype, a set of exemplars, and a theory to be coreferential.

**Jacobson** and **Margolis & Laurence** assert, mistakenly, that a prototype, a set of exemplars, and a theory cannot be coreferential because they fail to distinguish reference and categorization (on this distinction, see sect. 2.2.2 of *DwC*). The fact that objects can be miscategorized (e.g., a wolf can be misclassified as a dog) shows that reference differs from categorization. Indeed, categorization judgments could not be mistaken if every object that is categorized as an *x* (e.g., as a dog) really belonged to the extension of the concept of *x* (e.g., the concept of dog). Because reference differs from categorization, the fact that prototype-based and theory-based categorization processes can occasionally classify some objects differently (see the examples in sect. 3.3 of *DwC* and in Machery & Seppälä, forthcoming) does not entail that a prototype and a theory cannot be coreferential.

## R7. Should the heterogeneity hypothesis be broadened?

**Couchman, Boomer, Coutinho, & Smith** (**Couchman et al.**), **Dove**, and **Khemlani & Goodwin** argue that the heterogeneity hypothesis fails to capture the extent to which concepts form an heterogeneous kind, and they propose to extend the idea that concepts divide into very different kinds in two distinct directions. I am sympathetic to this kind of proposal: One of the important questions to be addressed by future research on knowledge representation is whether prototypes, exemplars, and theories exhaust the fundamental kinds of default bodies of knowledge (see the conclusion of *DwC*).

**Dove** proposes that, in addition to distinguishing prototypes, exemplars, and theories, it is necessary to distinguish at least two types of format: amodal and modal (the Representational Heterogeneity Hypothesis). Contra Barsalou, Prinz, and others, Dove agrees with me that not all concepts have a perceptual format (Dove 2009; Machery 2006c; 2007; forthcoming). However, following the dual-coding tradition (Paivio 1991), he also holds that *some* concepts have such a format. While the evidence reviewed by Dove is surely suggestive, I remain to be convinced. It is again important to keep in mind the distinction (which is apparently not challenged by Dove) between concepts and the knowledge used in higher cognition: Concepts are just a subset of the knowledge used in higher cognition (see sect. 1.1 of *DwC* and sect. R3 here for a discussion of how to draw the distinction). There is no doubt that we use perceptual representations to solve some tasks, and it is plausible indeed, as Dove argues, that dedicated cognitive systems are used for this purpose. However, this does not entail that these representations are concepts since they might only be used in particular circumstances, in a context-sensitive manner.

**Khemlani & Goodwin** propose to add two other kinds to the fundamental kinds of concepts: rules and generic representations. Goodwin's work about conceptual illusions provides some striking evidence that people are able to learn rules that determine category membership. However, as I explained in section 4.1.4 of *DwC*, the concern with the rule-based approach to concepts is not that people are unable to learn and apply rules, but rather that natural categories, outside of contrived laboratory conditions, do not have the definitions that rule-based accounts assume. In contrast to rules, it is extremely plausible that people store some knowledge about generics, as

Prasada's work (among others) suggests. In *DwC*, following many psychologists, I took knowledge of generics and knowledge of causal relations to be constitutive of theories, but, unfortunately, I did not defend this proposal. Khemlani & Goodwin want to distinguish representations of generic information from theories apparently on the grounds that generic knowledge cannot be identified with causal knowledge. However, because generic knowledge and causal knowledge about some *x's* (e.g., dogs) are plausibly intertwined, it would be a mistake to hold that generic knowledge and theories form distinct kinds of concepts. In the terminology of *DwC* (sect. 3.3.1; see also sect. R8 here), they seem to be linked (or connected) and coordinated. As a consequence, they should be thought of as parts of the same concept, our theory of dogs, rather than as two distinct concepts. In this sense, theories are true hybrids: They are made of distinct types of knowledge, perhaps used in distinct processes, that are linked and coordinated.

## R8. Hybrid theories

Proponents of hybrid theories of concepts agree with some, but not all, tenets of the heterogeneity hypothesis. They propose that the heterogeneous bodies of knowledge about a given category (e.g., dogs) are not distinct concepts, but rather the parts of a single concept. **Keil**, **Gonnerman & Weinberg**, **Margolis & Laurence**, and **Scarantino** defend hybrid theories against the attack mounted in section 3.3 of *DwC*. As I understood the notion of part, if two bodies of knowledge A and B are parts of the same concept, using A enables the use of B (A and B are linked or connected), and A and B must not lead to incompatible judgments that are taken to be equally authoritative (A and B are coordinated). Keil, Gonnerman & Weinberg, Margolis & Laurence, and Scarantino all reject *Coordination* as a necessary condition for two bodies of knowledge to be parts of the same concept. Although Coordination may finally turn out to be an inappropriate way to characterize the notion of conceptual parthood, I will defend it here. Note that, if Coordination is rejected, an alternative characterization of the notion of conceptual parthood should be provided (which Keil fails to do).

Before discussing their key arguments, let me clarify Coordination a bit. Coordination does not state that the parts of a hybrid concept cannot underlie incompatible judgments. This would be an inaccurate way of cashing out the notion of a part of a concept since some well-known hybrid theories of concepts (e.g., Osherson & Smith 1981) assume that the judgments underwritten by prototypical information can be defeated by (and can thus be incompatible with) the judgments underwritten by theoretical or definitional information. What Coordination excludes is that the parts of a concept give rise to incompatible judgments *that are taken to be equally authoritative.*

**Keil** and **Gonnerman & Weinberg** propose two different arguments purporting to show that Coordination cannot be a necessary condition for distinct bodies of knowledge to count as parts of the same concept. Keil rightly notes that predicates such as "tall" can lead to apparent contradictions: Someone can be tall with respect to some standard and not tall with respect to another standard. He infers that, if Coordination were a necessary condition for conceptual parthood, we would have to conclude, absurdly, that all these judgments involve distinct concepts. However, what is going on in these cases is clearly quite different from what is going on when people both agree and disagree with "Tomatoes are vegetables." In the former case, when one both agree and disagree with the claim that, say, John is tall, it is not because "tall" is ambiguous: The meaning of "tall" is the same in "John is tall" and "John is not tall." Instead, one can both agree and disagree with the claim that John is tall because two distinct standards are applied (compared with Tom Cruise, John is tall, but, compared with Shaquille O'Neal, John is not tall). By contrast, one can both agree and disagree with the claim that tomatoes are vegetables because "vegetable" is ambiguous: The meaning of "vegetable" (i.e., the concept this word expresses) is not the same in "Tomatoes are vegetables" and in "Tomatoes are not vegetables."

**Gonnerman & Weinberg** note insightfully that the distinct exemplars of, for example, dogs (which, according to me, form a concept of dogs) seem to violate Coordination, which shows that Coordination cannot be a necessary condition for distinguishing concepts. Exemplars of, say, dogs can indeed underlie incompatible judgments – such as, inconsistent categorization judgments – since different exemplars might be retrieved from long-term memory in different circumstances, but it does not follow that these judgments are taken to be equally authoritative (although, to my knowledge, exemplar theorists have said little about this question). Perhaps the judgment underwritten by the larger number of exemplars is taken to defeat the judgment underwritten by the smaller number of exemplars. Although evidence is lacking to evaluate this proposal, it might thus be that exemplars do not violate Coordination.

**Scarantino** proposes to replace Coordination with weaker conditions, which are satisfied by prototypes, exemplars, and theories, and concludes that these count as parts of concepts rather than as distinct concepts. The main problem with Scarantino's conditions is that they are neither sufficient nor necessary for two bodies of knowledge to be parts of the same concept. Consider, for instance, his first proposal: Two bodies of knowledge are coordinated when the knowledge stored in one of them influences the acquisition of the knowledge stored in the other. A first problem with this proposal is that genuinely distinct concepts meet this condition. For instance, when one forms a body of knowledge about a new animal species, our theoretical body of knowledge about animals in general is likely to influence this acquisition process. A second problem is that the parts of a single concept need not meet this condition. For instance, the elements of knowledge about the typical properties of dogs are parts of the same prototype of dogs, but my knowledge about a typical property of dogs needs not influence the acquisition of my knowledge about another typical property.

Finally, **Keil** argues that hybrid theories of concepts can account for the findings that I argued undermine them – namely, the fact that people are willing to endorse apparent contradictions such as "Tomatoes are vegetables" and "Tomatoes are not vegetables," particularly when such sentences are prefaced with hedges such as "in a sense" (see Machery & Seppälä [forthcoming] for some relevant findings and discussion). There are two issues that need to be distinguished here. First, would people be willing to endorse apparent contradictions such as "Tomatoes are vegetables" and "Tomatoes are not vegetables" if they were not prefaced by hedges such as "in a sense"? I suspect that this is the case, but there is no clear evidence for this claim. Second, supposing that people would agree that tomatoes are vegetables and that they are not vegetables even without such edges, hybrid models would then be compatible with people's judgments only if Coordination were rejected. But, if hybrid theorists reject Coordination, they then need to explain why distinct bodies of knowledge about, say, tomatoes are parts of the same concept of tomato instead of being distinct concepts of tomato.

## R9. Eliminativism

Most commentators reject the eliminativist conclusion put forward in *DwC*, even when they agree with the existence of three different types of concepts.

### *R9.1. Are there generalizations about concepts?*

A key step in the eliminativist argument proposed in *DwC* consists in denying that the class of default bodies of knowledge forms a natural kind on the grounds that few scientifically interesting generalizations are true of this class. **Blanchard**, **Danks**, **Lombrozo**, **Virtel & Piccinini**, and **Yermolayeva & Rakison** challenge this claim.

Inspired by her fascinating work on explanation, **Lombrozo** proposes that prototypes, exemplars, and theories (together with, perhaps, a host of other kinds of knowledge) are used to explain. While she takes this finding to unify concepts, I disagree. Ribosomes and transfer RNA are both involved in the production of proteins out of amino acids, but it does not follow that they form a single kind. In fact, I propose that a different conclusion follows from Lombrozo's work. This work suggests that it might not be possible to characterize the notion of theory by means of the notion of explanation, as I did in *DwC* (following many psychologists), since being used to explain is not a distinctive property of theories. Just like induction, categorization, or concept combination, explanation might be one of the cognitive competencies that are subserved by distinct processes defined over different kinds of default body of knowledge.

**Blanchard**, **Virtel & Piccinini,** and **Yermolayeva & Rakison** challenge the claim that prototypes, exemplars, and theories are used in distinct processes (e.g., distinct categorization processes). If they are right, then generalizations about cognitive processes are true of all concepts, and the class of concepts is a genuine natural kind.

**Blanchard** notes that some evidence for the existence of theories (Luhmann et al. 2006) is compatible with theories being used in similarity-based processes just like exemplars and prototypes (see also **Hayes & Kearney** and **Yermolayeva & Rakison**). However, research on induction (reviewed in sect. 7.1 of *DwC*) shows that the process underlying theory-based induction and the processes using prototypes and exemplars differ. The latter are similarity-based – representations are compared with one another, and their match is evaluated by some similarity measure – while phenomena like the causal asymmetry effect show that the inductive processes using causal knowledge are not based on similarity. **Virtel & Piccinini** hold that there is no evidence that exemplars and prototypes are used in different kinds of cognitive process. However, even if exemplars and prototypes were used in the same kind of process, it would still be the case that theories are used in a different kind of process, and thus that there are no generalizations about how concepts are used in cognitive processes.

**Yermolayeva & Rakison** rightly bemoan the fact that *DwC* paid little attention to developmental psychology, including the acquisition pattern of prototypes, exemplars, and theories (but see the brief discussion of Smith & Minda [1998] in sect. 6.4.4 of the book), and further work on the developmental questions raised in their commentary is called for. However, I find the proposed developmental sequence unconvincing. First, I object to the way *exemplar* is used in Yermolayeva & Rakison's commentary: An exemplar is not just any representation of an individual, it is a representation that is used by default in higher cognition. Thus, the fact that babies acquire representations of individuals (e.g., of their mother) early does not entail that they acquire genuine exemplars: Such representations are exemplars only if they can be used, for example, to categorize in addition to identify the individuals they are about. Second, in some category-learning experiments with adult participants, prototypes seem to be acquired before exemplars (Smith & Minda 1998). Research on medical expert knowledge also suggests that causal theories are acquired at the beginning of physicians' training, *before* physicians acquire any knowledge of particular cases (in the second half of their training) and form prototypes (for review, see Norman et al. 2006). It would thus seem that there is no necessary acquisition sequence, which casts doubts on the idea that a single process underlies the acquisition of prototypes, exemplars, and theories.

**Danks** puts forward a distinct criticism, based on his finding that prototype-, exemplar-, and theory-based formal models of categorization can be seen as distinct graphical models (Danks 2007). However, the fact that different formal models are instances of a more abstract formalism does not entail that the processes described by these models form a unified kind. Lokta-Volterra equations in ecology, Hodgkin and Huxley's model of the action potential in neuroscience, and the Cagan monetary model in economics are all differential equations. Would Danks conclude that they form a unified kind of process that is the object of a unified empirical theory?

### *R9.2. Do concepts form a natural kind?*

**Samuels & Ferreira** acknowledge, perhaps for the sake of the argument, that the class of bodies of knowledge used by default in higher cognition includes several kinds that have little in common (viz., prototypes, exemplars, and theories), but they insist that the bodies of knowledge used by default in higher cognition are a natural kind on the grounds that they form a genuine homeostatic property cluster kind. However, being a homeostatic property cluster kind is not sufficient for being a natural kind. Natural kinds are those kinds that support a large number of scientifically relevant inductions. Because few generalizations are true of them, the bodies of knowledge used by default in higher cognition do not form a natural kind.

**Samuels & Ferreira** rightly note that there is no precise cutting point for distinguishing those kinds that support, respectively, few and many generalizations. However, those kinds that support many generalizations are different from those kinds that support only few generalizations, exactly as white differs from black even if there is no precise cutting point when one moves from white to black through gray (Machery 2005). The bodies of knowledge used by default in higher cognition are an instance of the latter type of kinds and, for this reason, are not a natural kind.

### *R9.3. Concepts as a functional kind*

While **Couchman et al.**, **Khemlani & Goodwin**, **Lalumera**, and **Weiskopf** acknowledge that prototypes, exemplars, and theories form distinct kinds of concept and that there are few generalizations true of all of them, they reject the conclusion that the notion of concept should be eliminated from the theoretical vocabulary of psychology on the grounds that concepts form a functional kind. While concepts are indeed a functional kind – my own explication of the notion of concept, C, is functional – this alone does not settle the issue of whether *concept* should be eliminated.

Being a functional kind is not sufficient for earning one's keep in a scientific classification: Some functional kinds have their place, and others do not. Functional kinds have their place in scientific classificatory schemes either because they are natural kinds or because denoting them fulfills some useful function and does not have any drawback. So, what needs to be shown is either that many generalizations are true of the bodies of knowledge used by default in higher cognition or that denoting this class fulfills some useful function in cognitive science.

### *R9.4. Is the notion of concept useful for cognitive scientists?*

**Hampton**, **Lalumera**, **Strohminger & Moore**, and **Weiskopf** contend that the notion of concept has a useful role to play in cognitive science and that, as a result, eliminating it would be detrimental.

**Hampton** contends that the term *concept* is useful for bringing attention to the way prototypes, exemplars, and theories are organized, but I do not see exactly why this term would be needed. It seems straightforward to ask whether prototypes, exemplars, and theories are coordinated, whether they acquired in any particular developmental sequence, and so on.

**Lalumera** argues that we need the notion of concept to explain why our representation of, say, dogs changes over time: We start with exemplars of particular dogs, then develop a prototype, and finally build a theory. This won't do, however, for – as I have argued at length – there is no such thing as our representation of dogs. Rather, we simultaneously have several distinct representations of dogs, and Lalumera provides no reason to doubt this claim. Furthermore, as noted earlier in my reply to **Yermolayeva & Rakison** (sect. R9.1), it is not the case that prototypes, exemplars, and theories are necessarily acquired in any particular order.

**Weiskopf** provides a different, intriguing reason for keeping the term *concept* in the classificatory scheme of cognitive science: There is an important distinction between those organisms that have this type of body of knowledge and those that do not. However, I doubt that the class of organisms that have concepts would be of interest to comparative psychologists. It will probably be more fruitful to examine which organisms have prototypes, and which processes in which species use these, or to compare the causal theories in humans and the theories (or proto-theories) in other species. That is, my concerns about the usefulness of the notion of concept for psychologists working on human higher cognition carry over to comparative psychology. It might even well be that the term *concept* misleads us in thinking that the class of organisms that have concepts is an interesting class for comparative psychology, exactly as it misleads us in thinking that in the human mind they form an interesting class for cognitive scientists working on higher cognition.

**Strohminger & Moore** note that keeping *concept* within the classificatory scheme of psychology has numerous virtues. Further research might undermine the evidence suggesting that there are very different kinds of body of knowledge used by default in higher cognition, and not eliminating *concept* might keep psychologists aware of this possibility. However, I doubt that caution is needed here because the evidence in support of the existence of distinct kinds of concept seems unlikely to be undermined. Strohminger & Moore also note that the description "bodies of knowledge used by default in higher cognition" is ungainly, and, as a result, unlikely to be adopted by cognitive scientists as a replacement for *concept* (see similar concerns in Mercier 2010). I hope that the benefits that I argue would fall out from eliminating *concept* will convince cognitive scientists that this is a cost worth paying.

### *R9.5. Does the eliminativist argument against "concept" overgeneralize?*

According to **Gonnerman & Weinberg**, **Khemlani & Goodwin**, and **Margolis & Laurence,** my eliminativist argument cannot be valid because, if it were, we would have to eliminate numerous notions that have earned their keep in science in general and in cognitive science in particular, such as the notions of representation, module, algorithm, and nutrient. After all, representations and nutrients are probably no more natural kinds than concepts.

In response, first, I do not hold that *concept* should be eliminated merely because it fails to pick out a natural kind. Rather, in addition to failing to pick out a natural kind, keeping *concept* has numerous drawbacks and few benefits. Second, the case of concepts is very different from the case of, say, representations. Psychologists do not attempt to discover generalizations about representations in general and to encompass these generalizations within a theory of representations, while they do precisely this for concepts. There is thus no theoretical habit to curb in the case of representations, while, if I am right about how knowledge is organized, cognitive scientists' tendencies to develop theories of concepts should be curbed (for some recent attempts, see, e.g., Gallese & Lakoff 2005; Martin 2007; Prinz, forthcoming).

One might wonder whether the use of *concept* really impedes the progress of cognitive science. **Strohminger & Moore** rightly note that I provide little actual evidence in support of this claim. **Couchman et al.** seem to agree with this criticism. I acknowledge that this claim is partly speculative, but it strikes me as plausible. While the use of *concept* in cognitive science and the attempts to develop a unified theory of concepts have not prevented cognitive scientists for making numerous findings about concepts, some important questions have not attracted sufficient attention, such as: How are prototypes, exemplars, and theories used concomitantly? And what happens when they yield incompatible judgments? I am less convinced than Couchman et al. and **Markman** that we have already acquired an extensive knowledge about these questions, although some noticeable work has already been conducted (including by Smith and Markman). I further hypothesize that much more work would be done on these questions if cognitive scientists stopped theorizing about concepts, and started theorizing about prototypes, exemplars, and theories.

## References

**[Letters "a" and "r" appearing before authors' initials refer to target article and response references, respectively.]**


Allen, S. W. & Brooks, L. R. (1991) Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General* 120:3–19. [BKH, arEM]

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [aEM]

Anderson, J. R. & Betz, J. (2001) A hybrid model of categorization. *Psychonomic Bulletin and Review* 8:629–47. [aEM]

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. & Waldron, E. M. (1998) A neuropsychological theory of multiple systems in category learning. *Psychological Review* 105(3):442–81. [ABM]

Ashby, F. G. & Ell, S. W. (2002) Single versus multiple systems of learning and memory. In: *Stevens' handbook of experimental psychology, vol.4: Methodology in experimental psychology, 3rd ed.*, ed. J. Wixted & H. Pashler, pp. 655–92. Wiley. [aEM]

Ashby, F. G. & Maddox, W. T. (2004) Human category learning. *Annual Review of Psychology* 56:149–78. [aEM]

Ashby, F. G. & Maddox, W. T. (2005) Human category learning. *Annual Review of Psychology* 56:149–78. [JJC, BKH]

Barrera, M. E. & Maurer, D. (1981) Recognition of mother's photographed face by the three-month-old infant. *Child Development* 52:714–16. [YY]

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

Barsalou, L. W. (1985) Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11:629–54. [aEM]

Barsalou, L. W. (1987) The instability of graded structures: Implications for the nature of concepts. In: *Concepts and conceptual development: Ecological and intellectual factors in categorization*, ed. U. Neisser, (pp. 101–38). Cambridge University Press. [aEM]

Barsalou, L. W. (1993) Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In: *Theories of memory*, ed. A. C. Collins, S. E. Gathercole & M. A. Conway, pp. 29–101. Erlbaum. [arEM]

Barsalou, L. W. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22(4):577–660. [GD, SH, aEM]

Barsalou, L. W. (2008a) Cognitive and neural contributions to understanding the conceptual system. *Current Directions in Psychological Science* 17:91–95. [aEM]

Barsalou, L. W. (2008b) Grounded cognition. *Annual Review of Psychology* 59:617–45. [aEM]

Barsalou, L. W. (2009) Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London: Biological Sciences* 364:1281–89. [aEM]

Barsalou, L. W. & Hale, C. R. (1993) Components of conceptual representation: From feature lists to recursive frames. In: *Categories and concepts: Theoretical views and inductive data analysis*, ed. I. van Mechelen, J. A. Hampton, R. S. Michalski & P. Theuns, pp. 97–144. Academic Press. [JAH]

Barsalou, L. W., Simmons, W. K., Barbey, A. K. & Wilson, C. D. (2003) Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* 7:84–91. [aEM]

Berndt, R. S., Haendiges, A. N., Burton, M. W. & Mitchum, C. C. (2002) Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics* 15:353–71. [GD]

Binder, J., Westbury, C., McKiernan, K., Possing, E. & Medler, D. (2005) Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience* 17:905–17. [GD]

Bird, H., Howard, D. & Franklin, S. (2003) Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics* 16:113–49. [GD]

Blair, M. & Homa, D. (2003) As easy to memorize as they are to classify: The 5-4 categories and the category advantage. *Memory and Cognition* 31:1293–1301. [JJC]

Blondin-Massé, A., Chicoisne, G., Gargouri, Y., Harnad, S., Picard, O. & Marcotte, O. (2008) How is meaning grounded in dictionary definitions? Paper presented at TextGraphs-3 Workshop – 22nd International Conference on Computational Linguistics, 18 August 2008. [SH]

Bloom, P. (1996) Intention, history, and artifact concepts. *Cognition* 60:1–29. [aEM]

Bloom, P. (2000) *How children learn the meanings of words*. MIT Press. [EL]

Boulenger, V., Hauk, O. & Pulvermüller, F. (2009) Grasping ideas with the motor system: Semantic somatotopy in idiom comprehension. *Cerebral Cortex* 19:1905–14. [aEM]

Boyd, R. (1989) What realism implies and what it does not. *Dialectica* 43:5–29. [RS]

Boyd, R. (1991) Realism, anti-foundationalism and the enthusiasm for natural kinds. *Philosophical Studies* 61:127–48. [aEM, RS]

Boyd, R. (1999) Kinds, complexity and multiple realization. *Philosophical Studies* 95:67–98. [aEM]

Braisby, N. (2005) Similarity and categorisation: Getting dissociations in perspective. In: *Proceedings of the Twenty-Sixth Annual Cognitive Science Society*, ed. K. Forbus, D. Getner & T. Regier, pp. 150–55. Erlbaum. [CG]

Buresh, J. S. & Woodward, A. L. (2007) Infants track action goals within and across agents. *Cognition* 104:287–314. [YY]

Burge, T. (1979) Individualism and the mental. *Midwest Studies in Philosophy* 4:73–121. [rEM]

Burge, T. (1986) Individualism and psychology. *Philosophical Review* 95(1):3–46. [GR]

Butterworth, G., Jarrett, N. & Hicks, L. (1982) Spatiotemporal identity in infancy: Perceptual competence or conceptual deficit? *Developmental Psychology* 18:435–49. [HAV]

Cangelosi, A. & Harnad, S. (2001) The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication* 4(1):117–14. [SH]

Caramazza, A. (1986) On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and Cognition* 5:41–66. [aEM]

Carey, S. (1985) *Conceptual change in childhood.* MIT Press. [TL]

Carey, S. (2009) *The origin of concepts.* Oxford University Press. [EMar]

Carey, S. & F. Xu (2001) Infants' knowledge of objects: Beyond object files and object tracking. *Cognition* 80(1–2):179. [AJJ]

Carlson, G. N. & Pelletier, F. J. (1995) *The generic book.* University of Chicago Press. [SSK]

Carmichael, C. & Hayes, B. K. (2001) Prior knowledge and exemplar encoding in children's concept acquisition. *Child Development* 72:1071–90. [BKH]

Chao, L. L. & Martin, A. (1999) Cortical representation of perception, naming, and knowledge of color. *Journal of Cognitive Neuroscience* 11:25–35. [aEM]

Chaput, H. H. & Cohen, L. B. (2001) A model of infant causal perception and its development. In: *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, ed. J.D. Moore & K. Stenning, pp. 182–87. Erlbaum. [YY]

Chin-Parker, S. & Ross, B. H. (2002) The effect of category learning on sensitivity to within-category correlations. *Memory and Cognition* 30(3):353–62. [ABM]

Churchland, P. M. (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78:67–90. [aEM]
Cohen, B. & Murphy, G. L. (1984) Models of concepts. *Cognitive Science* 8:27–58. [AS]
Collini, E., Wong, C. Y., Wilk, K. E., Curmi, P. M. G., Brumer, P. & Scholes, G. D. (2010) Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature* 463:644. [rEM]
Connolly, A. C., Fodor, J. A., Gleitman, L. R. & Gleitman, H. (2007) Why stereotypes don't even make good defaults. *Cognition* 103:1–22. [SSK]
Couchman, J. J., Coutinho, M. V. C. & Smith, J. D. (in press) Rules and resemblance: Their changing balance in the category learning of humans (*Homo sapiens*) and rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes.* [JJC]
Crepaldi, D., Aggujaro, S., Arduino, L. S., Zonca, G., Ghirardi, G., Inzaghi, M. G., Colombo, M., Chierchia, G. & Luzzatti, C. (2006) Noun-verb dissociation in aphasia: The role of imageability and functional location of the lesion. *Neuropsychologia* 44(1):73–89. [GD]
Danks, D. (2004) *Psychological theories of categorization as probabilistic models.* Technical report CMU-PHIL-157. July 15, 2004. [DD]
Danks, D. (2007) Theory unification and graphical models in human categorization. In: *Causal learning: Psychology, philosophy, and computation*, ed. A. Gopnik & L. Schulz, pp. 173–89. Oxford University Press. [DD, rEM]
Devitt, M. (1981) Designation. Columbia University Press. [GR]
Devitt, M. (1996) *Coming to our senses.* Cambridge University Press. [GR]
Devitt, M. (forthcoming) Experimental semantics. *Philosophy and Phenomenological Research.* [GR]
Diamond, A. (1985) Development of the ability to use recall to guide action, as indicated by infants' performance on A-not-B. *Child Development* 56:868–83. [HAV]
Dove, G. (2009) Beyond perceptual symbols: A call for representational pluralism. *Cognition* 110:412–31. [GD, arEM]
Dretske, F. (1981) *Knowledge and the flow of information*. Blackwell/MIT Press. [KE, GR]
Dunn, J. C. (2008) The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review* 115(2):426–46. [BKH]
Dunn, J. C. & Kirsner, K. (1988) Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review* 95:91–101. [aEM]
Dunn, J. C. & Kirsner, K. (2003) What can we infer from double dissociations? *Cortex* 39:1–7. [aEM]
Edwards, K. (2009) What concepts do. *Synthese* 170:289–310. [KE, aEM]
Edwards, K. (2010) Concept referentialism and the role of empty concepts. Mind and Language 25(1):89–118. [KE]
Elder, C. L. (1994) Higher and lower essential natures. *American Philosophical Quarterly* 31:255–65. [CG]
Evans, G. (1982) *The varieties of reference.* Oxford University Press. [rEM]
Evans, J. S. B. T. (2007) *Hypothetical thinking: Dual processes in reasoning and judgement.* Routledge. [aEM]
Evans, J. S. B. T. & Frankish, K., eds. (2009) *In two minds: Dual processes and beyond.* Oxford University Press. [aEM, JV]
Feeney, A. & Heit, E., eds. (2007) *Inductive reasoning: Experimental, Developmental, and computational approaches.* Cambridge University Press. [aEM]
Feldman, J. (2000) Minimization of Boolean complexity in human concept learning. *Nature* 407:630–33. [SSK]
Feldman, J. (2006) An algebra of human concept learning. *Journal of Mathematical Psychology* 50:339–68. [SSK]
Field, H. (2001) *Truth and the absence of fact*. Oxford University Press. [KE]
Flynn, J. J., Nedbal, M. A., Dragoo, J. W. & Honeycutt, R. L. (2000) Whence the red panda? *Molecular Phylogenetics and Evolution* 17(2):190–99. [AJJ]
Fodor, J. A. (1974) Special sciences (Or: The disunity of science as a working hypothesis). Synthese 28(2):97–115. [KE]
Fodor, J. A. (1975) *The language of thought.* Crowell. [aEM, DAW]
Fodor, J. A. (1987) *Psychosemantics: The problem of meaning in the philosophy of mind.* MIT Press. [KE]
Fodor, J. A. (1990) *A theory of content and other essays.* MIT Press. [rEM, GR]
Fodor, J. A. (1992) *A theory of content and other essays.* MIT Press. [KE]
Fodor, J. A. (1997) Special sciences: Still autonomous after all these years. *Philosophical Perspectives: Mind Causation, and World 2:49–63.* (Noûs *31, Suppl.) [KE]*
Fodor, J. A. (1998) *Concepts, Where cognitive science went wrong.* Oxford University Press. [FK, aEM, GR, SS]
Fodor, J. A. (2003) Is it a bird? Problems with old and new approaches to the theory of concepts. *Times Literary Supplement*, January 17, 2003, pp. 3–4. [aEM]
Fodor, J. A. (2008) *LOT 2: The language of thought revisited.* Oxford University Press. [aEM, SS, DAW]
Forster, M. & Sober, E. (1994) How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science* 45:1–36. [rEM]
Gallese, V. & Lakoff, G. (2005) The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 21:455–79. [arEM]
Garrod, S. & Anderson, A. (1987) Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition* 27:181–218. [ABM]
Garrod, S. & Doherty, G. (1994) Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215. [ABM]
Gelman, R. (2004) Cognitive development. In: *Stevens' handbook of experimental psychology, vol. 3: Memory and cognitive processes*, ed. H. Pashler & D. L. Medin, pp. 533–60. Wiley. [aEM]
Gelman, S. A. (2003) *The essential child: Origins of essentialism in everyday thought.* Oxford University Press. [TB, SSK]
Gelman, S. A. (1988) The development of induction within natural kinds and artifacts categories. *Cognitive Psychology* 20:65–95. [aEM]
Gelman, S. A. (2003) *The essential child. Origins of essentialism in everyday thought.* Oxford University Press. [aEM]
Gelman, S. A. & Markman, E. (1986) Categories and induction in young children. *Cognition* 23:183–209. [aEM]
Gibson, E. J. (1969) *Principles of perceptual learning and development.* Prentice Hall. [HAV]
Giesbrecht, B., Gamblin, C. & Swaab, T. (2004) Separable effects of priming and imageability on word processing in human cortex. *Cerebral Cortex* 14:521–29. [GD]
Gigerenzer, G. & Regier, T. P. (1996) How do we tell an association from a rule? *Psychological Bulletin* 119:23–26. [aEM]
Gigerenzer, G., Todd, P. M. & the ABC Research Group (1999) *Simple heuristics that make us smart.* Oxford University Press. [aEM]
Glenberg, A. M. (1997) What memory is for? *Behavioral and Brain Sciences* 20:1–55. [GD, aEM]
Glenberg, A. M. & Robertson, D. A. (2000) Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language* 43:379–401. [SH]
Glymour, C. (1994) On the methods of cognitive neuropsychology. *British Journal for the Philosophy of Science* 45:815–35. [aEM]
Goldberg, R. F., Perfetti, C. A. & Schneider, W. (2006) Distinct and common cortical activations for multimodal semantic categories. *Cognitive, Affective, and Behavioral Neuroscience* 6:214–22. [GD]
Goodwin, G. P. & Johnson-Laird, P. N. (2010) Conceptual illusions. *Cognition* 114:253–65. [SSK]
Goodwin, G. P. & Johnson-Laird, P. N. (submitted) Models as the representation of Boolean concepts. [SSK]
Gopnik, A. (2003) The theory theory as an alternative to the innateness hypothesis. In: *Chomsky and his critics*, ed. L. Antony & N. Hornstein, pp. 238–54. Blackwell. [aEM]
Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T. & Danks, D. (2004) A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review* 111:1–31. [aEM]
Gopnik, A. & Meltzoff, A. N. (1997) *Words, thoughts, and theories.* MIT Press. [TL, aEM]
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. (2001) An fMRI investigation of emotional engagement in moral Judgment. *Science* 293:2105–108. [aEM]
Griffiths, P. E. (1997) *What emotions really are.* Chicago University Press. [aEM, AS, RS]
Griffiths, T. L., Steyvers, M. & Tenenbaum, J. B. (2007) Topics in semantic representation. *Psychological Review* 114:211–44. [aEM]
Hampton, J. A. (1979) Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior* 18:441–61. [TB, aEM]
Hampton, J. A. (1981) An investigation of the nature of abstract concepts. *Memory and Cognition* 9:149–56. [aEM]
Hampton, J. A. (1993) Prototype models of concept representation. In: *Categories and concepts: Theoretical views and inductive data analysis*, ed. I. Van Mechelen, J. A., Hampton, R. S. Michalski & P. Theuns, pp. 67–95. Academic Press. [aEM]
Hampton, J. A. (1998) Similarity-based categorization and fuzziness of natural categories. *Cognition* 65:137–65. [JAH]
Hampton, J. A. (2001) The role of similarity in natural categorization. In: *Similarity and categorization*, ed. U. Hahn & M. Ramscar, pp. 13–28. Oxford University Press. [rEM]
Hampton, J. A. (2006) Concepts as prototypes. In: *The psychology of learning and motivation: Advances in research and theory, vol. 46*, ed. B. H. Ross, pp. 79–113. Academic Press. [aEM]
Hampton, J. A. (2007) Typicality, graded membership, and vagueness. *Cognitive Science* 31:355–84. [aEM]
Hampton, J. A., Storms, G., Simmons, C. L. & Heussen, D. (2009) Feature integration in natural language concepts. *Memory and Cognition* 37:1721–30. [FK]
Harnad, S. (1990) The symbol grounding problem. *Physica D* 42:335–46. [SH]

Harnad, S. (2005) To cognize is to categorize: cognition is categorization. In: *Handbook of Categorization*, ed. C. Lefebvre & H. Cohen, pp. 20–42. Elsevier. [SH]

Harnad, S. (2006) Cohabitation: Computation at 70, cognition at 20. In: *Essays in Honour of Zenon Pylyshyn*, ed. D. Dedrick. pp. 245–57. MIT Press. [SH]

Harnad, S. (2007) From knowing how to knowing that: Acquiring categories by word of mouth. Paper presented at the Kaziemierz Naturalized Epistemology Workshop (KNEW), Kaziemierz, Poland, September 2, 2007. [SH]

Harnad, S. (2008) The annotation game: On Turing (1950) on computing, machinery and intelligence. In: *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer*, ed. R. Epstein & G. Peters, pp. 23–66. Springer. [SH]

Hauk, O., Johnsrude, I. & Pulvermüller, F. (2004) Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41:301–07. [aEM]

Heit, E. (1994a) Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:1264–82. [BKH]

Heit, E. (1994b) Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:411–22. [AS]

Heit, E. (2000) Properties of inductive reasoning. *Psychonomic Bulletin and Review* 7:569–92. [aEM]

Heit, E. & Hayes, B. (2008) Predicting reasoning from visual memory. In: *Proceedings of the 29th Meeting of the Cognitive Science Society*, ed. V. Sloutsky, B. Love & K. McCrae, pp. 83–88. Erlbaum. [BKH]

Hill, C. (2010) I love Machery's book, but love concepts more. *Philosophical Studies* 149:411–21. [rEM]

Hilton, D. J. & Slugoski, B. R. (1986) Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review* 93:75–88. [TL]

Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B. & Kiefer, M. (2008) Conceptual flexibility in the human brain: Dynamic recruitment of semantic maps from visual, motor, and motion-related areas. *Journal of Cognitive Neuroscience* 20(10):1799–814. [arEM]

Holcomb, P. J., Kounios, J., Anderson, J. E. & West, W. C. (1999) Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:721–42. [GD]

Homa, D., Sterling, S. & Trepel, L. (1981) Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory* 7:418–39. [JJC]

Horobin, K. & Acredolo, L. (1986) The role of attentiveness, mobility history, and separation of hiding sites on stage IV search behavior. *Journal of Experimental Child Psychology* 41:114–27. [HAV]

Horwich, P. (1998a) *Meaning*. Clarendon Press. [KE, GR]

Horwich, P. (1998b) *Truth*, 2nd edition. Clarendon Press. [KE]

Jackendoff, R. (2002) *Foundations of language: Brain, meaning, grammar, evolution.* Oxford University Press. [EMar]

James, T. W. & Gauthier, I. (2003) Auditory and action semantic features activate sensory-specific perceptual brain regions. *Current Biology* 13:1792–96. [arEM]

Jonides, J., Lewis, R., Nee, D. E., Lustig, C. A., Berman, M. G. & Moore K. S. (2008) The mind and brain of short-term memory. *Annual Review of Psychology* 59:15.1–15.32. [NS]

Kable, J. W., Kan, I. P., Wilson, A., Thompson-Schill, S. L. & Chatterjee, A. (2005) Conceptual representations of action in the lateral temporal cortex. *Journal of Cognitive Neuroscience* 17:1855–870. [aEM]

Keil, F. C. (1989) *Concepts, kinds, and cognitive development.* MIT Press. [TB, FK, aEM, AS, SS]

Keil, F. C. (2006) Explanation and understanding. *Annual Review of Psychology* 57:227–54. [TL]

Keil, F. C., Carter Smith, W., Simons, D. J. & Levin, D. T. (1998) Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition* 65:103–35. [aEM]

Keil, F. C. & Newman, G. E. (2010) Darwin and development: Why ontogeny does not recapitulate phylogeny for human concepts. In: *The making of human concepts*, ed. D. Mareschal, P. Quin & S. Lea. pp. 317–34. Oxford University Press. [FK]

Kelemen, D. (1999) Function, goals, and intention: Children's teleological reasoning about objects. *Trends in Cognitive Sciences* 3:461–68. [TL]

Kellenbach, M. L., Wijers, A. A., Hovis, M., Mulder, J. & Mulder, G. (2002) Neural differentiation of lexico-syntactic categories or semantic features? Event related potential evidence for both. *Journal of Cognitive Neuroscience* 14:561–77. [GD]

Kemp, C. & Tenenbaum, J. (2009) Structured statistical models of inductive reasoning. *Psychological Review* 116:20–58. [BKH]

Khemlani, S., Leslie, S. J. & Glucksberg, S. (submitted) Generics modulate default inferences. [SSK]

Kiefer, M., Sim, E.-J., Liebich, S., Hauk, O. & Tanaka, J. (2007) Experience-dependent plasticity of conceptual representations in human sensory–motor areas. *Journal of Cognitive Neuroscience* 19:525–42. [aEM]

Kim, J. (1992) Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52(1):1–26. [KE, EL]

Kim, J. (1998) *Mind in a physical world: An essay on the mind-body problem and mental causation.* MIT Press. [KE]

Knowlton, B. J. & Squire, L. R. (1993) The learning of categories: Parallel brain systems for item memory and category knowledge. *Science* 262:1747–49. [SZ]

Knowlton, B. J., Squire, L. R. & Gluck, M. A. (1994) Probabilistic classification learning in amnesia. *Learning and Memory* 1:106–20. [ABM]

Kripke, S. (1972/1980) *Naming and necessity*. Harvard University Press. [GR]

Kruschke, J. K. (2005) Category learning. In: *The handbook of cognition*, ed. K. Lamberts & R. L. Goldstone, pp. 183–201. Sage. [BKH]

Kulatanga-Moruzi, C., Brooks, L. R. & Norman, G. R. (2001) Coordination of analytical and similarity based processing strategies and expertise in dermatological diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:563–72. [rEM]

Lakoff, G. (1972) Hedges: A study in meaning criteria and the logic of fuzzy concepts. In: *Papers from the Eighth Regional Meeting, Chicago Linguistic Society, Chicago, IL*, pp. 183–228. Chicago Linguistic Society. [FK]

Laurence, S. & Margolis, E. (1999) Concepts and cognitive science. In: *Concepts, core readings*, ed. E. Margolis & S. Laurence, pp. 3–82. MIT Press. [TL, EMar, aEM]

Laurence, S. & Margolis, E. (2002) Radical concept nativism. *Cognition* 86:25–55. [SS]

Lauritzen, S. L. (1996) *Graphical models.* Oxford University Press. [DD]

Lawler, J. (1973) *Studies in English generics. University of Michigan Papers in Linguistics, vol. 1.* University of Michigan Press. [SSK]

Logan, G. D. (1988) Toward an instance theory of automaticity. *Psychological Review* 95:492–527. [ABM]

Logan, G. D. (2002) An instance theory of attention and memory. *Psychological Review* 109(2):376–400. [ABM]

Lombrozo, T. (2006) The structure and function of explanations. *Trends in Cognitive Sciences* 10(10):464–70. [TL]

Lombrozo, T. (2009) Explanation and categorization: How "why?" informs "what?" *Cognition* 110:248–53. [TL]

Lombrozo, T. (under review) Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. [TL]

Lombrozo, T. & Carey, S. (2006) Functional explanation and the function of explanation. *Cognition* 99(2):167–204. [TL]

Lombrozo, T., Kelemen, D. & Zaitchik, D. (2007) Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science* 18(11):999–1006. [TL]

López, A., Atran, S., Coley, J. D., Medin, D. L. & Smith E. E. (1997) The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology* 32:251–95. [aEM]

Love, B. C., Medin, D. L. & Gureckis, T. M. (2004) SUSTAIN: A network model of category learning. *Psychological Review* 111:309–32. [BKH]

Luhmann, C. C., Ahn, W. & Palmeri, T. (2006) Theory-based categorization under speeded conditions. *Memory and Cognition* 34:1102–11. [TB, rEM]

Machery, E. (2005) Concepts are not a natural kind. *Philosophy of Science* 72:444–67. [arEM, RS]

Machery, E. (2006a) How to split concepts. Reply to Piccinini and Scott. *Philosophy of Science* 73:410–18. [aEM]

Machery, E. (2006b) Review of A. Zilhao, ed.: *Evolution, rationality, and cognition: A cognitive science for the twenty-first century. Notre Dame Philosophical Reviews.* Retrieved from http://ndpr.nd.edu/review.cfm?id=6342. [rEM]

Machery, E. (2006c) Two dogmas of neo-empiricism. *Philosophy Compass* 1:398–412. [arEM]

Machery, E. (2007) Concept empiricism: A methodological critique. *Cognition* 104:19–46. [GD, arEM]

Machery, E. (2009) *Doing without concepts.* Oxford University Press. [TB, JJC, DD, GD, KE, CG, JAH, SH, BKH, AJJ, FK, SSK, EL, TL, EMar, arEM, ABM, GR, RS, AS, SS, NS, JV, HAV, DAW, YY, SZ]

Machery, E. (forthcoming) Reply to Barbara Malt and Jesse Prinz. *Mind and Language.* [rEM]

Machery, E. (2010) Reply to my critics. *Philosophical Studies* 149:429–36. [rEM]

Machery, E. & Seppälä, S. (forthcoming) Against hybrid theories of concepts. *Anthropology & Philosophy*. [arEM]

Maddox, W. T. & Ashby, F. G. (1993) Comparing decision-bound and exemplar models of classification. *Perception and Psychophysics* 53:49–70. [SZ]

Maddox, W. T. & Ashby, F. G. (2004) Dissociating explicit and procedure-learning based systems of perceptual category learning. *Behavioral Processes* 66(3):309–32. [ABM]

Mahon, B. Z. & Caramazza, A. (2008) A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology, Paris* 102:59–70. [aEM]
Mahon, B. Z. & Caramazza, A. (2009) Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology* 60:27–51. [aEM]
Mallon, R. (2006) "Race": Normative, not metaphysical or semantic. *Ethics* 116:525–51. [aEM]
Mallon, R., Machery, E., Nichols, S. & Stich, S. P. (2009) Against arguments from reference. *Philosophy and Phenomenological Research* 79:332–56. [aEM]
Malt, B. C. (1989) An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(4):539–55. [JV]
Malt, B. C. (1994) Water is not $H_2O$. *Cognitive Psychology* 27:41–70. [aEM]
Malt, B. C. (forthcoming) Why we should do without concepts. *Mind and Language.* [rEM]
Malt, B. C. & Sloman, S. A. (2007) Artifact categorization: The good, the bad, and the ugly. In: *Creations of the mind: Theories of artifacts and their representation*, ed. E. Margolis & S. Laurence, pp. 85–123. Oxford University Press. [aEM]
Malt, B. C., Sloman, S. A. & Gennari, S. P. (2003) Universality and language specificity in object naming. *Journal of Memory and Language* 49:20–42. [ABM]
Margolis, E. (1994) A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition* 51:73–89. [aEM]
Margolis, E. (1995) The significance of the theory analogy in the psychological study of concepts. *Mind and Language* 10:45–71. [aEM]
Margolis, E. & Laurence, S. (2006) Concepts. *Stanford Encyclopedia of Philosophy.* Available at http://plato.stanford.edu/entries/concepts/. [aEM]
Margolis, E. & Laurence, S. (2007) The ontology of concepts – Are concepts abstract objects or mental representations? *Noûs* 41(4):561–93. [EMar]
Margolis, E. & Laurence, S., eds. (1999) *Concepts: Core readings.* MIT Press. [SS]
Markman, A. B. & Dietrich, E. (2000) In defense of representation. *Cognitive Psychology* 40(2):138–71. [ABM]
Markman, A. B. & Makin, V. S. (1998) Referential communication and category acquisition. *Journal of Experimental Psychology: General* 127(4):331–54. [ABM]
Markman, A. B. & Ross, B. (2003) Category use and category learning. *Psychological Bulletin* 129:592–613. [BKH, ABM]
Marr, D. (1982) *Vision: A computational investigation in the human representation of visual information.* Freeman. [SSK]
Marshall, J., Pring, T., Chiat, S. & Robson, J. (1996) Calling a salad a federation: An investigation of semantic jargon. Part 1 – nouns. *Journal of Neurolinguistics* 9:237–50. [GD]
Martin, A. (2007) The representation of object concepts in the brain. *Annual Review of Psychology* 58:25–45. [arEM]
Martin, A. & Chao, L. L. (2001) Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology* 11:194–201. [aEM]
Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. & Ungerleider, L. G. (1995) Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* 270:102–05. [aEM]
Mayr, E. (1982) *The growth of biological thought: Diversity, evolution, and inheritance.* Harvard University Press. [JAH]
Medin, D. L., Lynch, E. B. & Solomon, K. O. (2000) Are there kinds of concepts? *Annual Review of Psychology* 51:121–47. [aEM]
Medin, D. L. & Schaffer, M. M. (1978) Context theory of classification learning. *Psychological Review* 85:207–38. [aEM]
Mellet, E., Tzourio, N., Denis, M. & Mazoyer, B. (1998) Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *NeuroReport* 9:803–08. [GD]
Mercier, H. (2010) How to cut a concept? Review of *Doing without concepts* by Edouard Machery. *Biology and Philosophy* 25:269–77. [rEM]
Millikan, R. G. (1984) *Language, thought, and other biological categories: New foundations for realism.* MIT Press. [KE, GR]
Millikan, R. G. (1993) *White Queen psychology and other essays for Alice.* MIT Press. [KE]
Millikan, R. G. (1998) A common structure for concepts of individuals, stuffs, and basic kinds: More mama, more milk, and more mouse. *Behavioral and Brain Sciences* 22:55–65. [EL]
Millikan, R. G. (2000) *On clear and confused ideas: An essay about substance concepts.* Cambridge University Press. [aEM]
Minda, J. P. & Smith, J. D. (2001) Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:775–99. [aEM, SZ]
Morris, C. D., Bransford, J. D. & Franks, J. J. (1977) Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior* 16:519–33. [ABM]
Murphy, D. & Stich, S. P. (1999) Griffiths, elimination and psychopathology. *Metascience* 8:13–25. [aEM]
Murphy, G. L. (2002) *The big book of concepts.* MIT Press. [JJC, JAH, BKH, AJJ, EL, aEM, AS]
Murphy, G. L. & Medin, D. L. (1985) The role of theories in conceptual coherence. *Psychological Review* 92:289–316. [TB, BKH, TL, aEM]
Myung, I. J., Forster, M. R. & Browne, M. W., eds. (2000) Model selection [Special issue]. *Journal of Mathematical Psychology* 44:190–204. [SZ]
Myung, I. J. & Pitt, M. A. (2009) Optimal experimental design for model discrimination. *Psychological Review* 116:499–518. [SZ]
Navarro D. J., Pitt M. A. & Myung, I. J. (2004) Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology* 49:47–84. [SZ]
Nesse, R. M. & Ellsworth, P. C. (2009) Evolution, emotions, and emotional disorders. *American Psychologist* 64:129–39. [NS]
Newell, B. & Dunn, J. C. (2008) Dimensions in data: Testing psychological models using state-trace analysis. *Trends in Cognitive Sciences* 12(8):285–90. [BKH]
Noppeney, U. & Price, C. J. (2004) Retrieval of abstract semantics. *NeuroImage* 22:164–70. [GD]
Norman, G., Eva, K., Brooks, L. & Hamstra, S. (2006) Expertise in medicine and surgery. In: *The Cambridge handbook of expertise and expert performance*, ed. K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman, pp. 339–53. Cambridge University Press. [rEM]
Nosofsky, R. M. (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:104–14. [aEM]
Nosofsky, R. M. (1992) Exemplar-based approach to relating categorization, identification, and recognition. In: *Multidimensional models of perception and cognition*, ed. F. G. Ashby, pp. 363–93. Erlbaum. [aEM]
Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. (1994) Rule-plus-exception model of classification learning. *Psychological Review* 101:266–300. [aEM]
Nosofsky, R. M. & Stanton, R. D. (2005) Speeded classification in a probabilistic category structure: Contrasting exemplar-retrieval, decision-boundary, and prototype models. *Journal of Experimental Psychology: Human Perception and Performance* 31:608–29. [aEM]
Nosofsky, R. M. & Zaki, S. R. (2002) Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 285:924–40. [SZ]
Öğmen, H. (2007) A theory of moving form perception: Synergy between masking, perceptual grouping, and motion computation in retinotopic and non-retinotopic representations. *Advances in Cognitive Psychology* 3(1–2):67–84. [AJJ]
Öğmen, H., Otto, T. U. & Herzog, M. H. (2006) Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Research* 46(19):3234–242. [AJJ]
Osherson, D. N. & Smith, E. E. (1981) On the adequacy of prototype theory as a theory of concepts. *Cognition* 9:35–58. [arEM]
Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A. & Shafir, E. (1990) Category-based induction. *Psychological review* 97:185–200. [aEM]
Paivio, A. (1987) *Mental representations: A dual-coding approach.* Oxford University Press. [GD]
Paivio, A. (1991) Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology* 45:255–87. [rEM]
Palmeri, T. J. & Flanery, M. A. (1999) Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science* 10(6):526–30. [JV]
Peacocke, C. (1992) *A study of concepts.* MIT Press. [aEM, GR]
Peacocke, C. (2008) *Truly understood.* Oxford University Press. [aEM]
Pecher, D., Zeelenberg, R. & Barsalou, L. W. (2004) Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin and Review* 11:164–67. [aEM]
Piaget, J. (1954) *The construction of reality in the child.* Basic Books. [HAV]
Piccinini, G. (forthcoming) Two kinds of concept: Implicit and explicit. *Dialogue.* [JV]
Piccinini, G. & Scott, S. (2006) Splitting concepts. *Philosophy of Science* 73:390–409. [GD, EL, aEM, AS, JV]
Pinker, S. (1997) *How the mind works.* Norton. [EMar]
Plaut, D. C. (1995) Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology* 17:291–21. [aEM]
Plumert, J. M. (2008) Children's thinking is not just about what's in the head: Understanding the organism and environment as a unified system. In: *Advances in child development and behavior*, ed. R. V. Kail, pp. 373–417. Academic Press. [HAV]
Prasada, S. & Dillingham, E. M. (2006) Principled and statistical connections in common sense conception. *Cognition* 99:73–112. [TL]

Prasada, S. & Dillingham, E. M. (2009) Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science* 33:401–48. [SSK]

Prasada, S., Khemlani, S., Leslie, S. J. & Glucksberg, S. (submitted) Conceptual distinctions amongst generics. [SSK]

Prinz, J. J. (2002) *Furnishing the mind: Concepts and their perceptual basis.* MIT Press. [GD, arEM, SS]

Prinz, J. J. (2005) The return of concept empiricism. In: *Handbook of categorization in cognitive science*, ed. H. Cohen & C. Lefebvre. pp. 679–95. Elsevier. [aEM]

Prinz, J. J. (forthcoming) Can concept empiricism forestall eliminativism? A commentary on Machery. *Mind & Language.* [rEM]

Proffitt, J. B., Coley, J. D. & Medin, D. L. (2000) Expertise and category-based induction. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:811–28. [aEM]

Pulvermüller, F. (2005) Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 6:576–82. [aEM]

Pulvermüller, F. & Hauk, O. (2006) Category-specific conceptual processing of color and form in left fronto-temporal cortex. *Cerebral Cortex* 16:1193–1201. [aEM]

Putnam, H. (1967) Psychological predicates. In: *Art, mind, and religion*, ed. W. H. Capitan & D. D. Merrill, pp. 37–48. University of Pittsburgh Press. [EL]

Putnam, H. (1975) The meaning of "meaning". In: *Collected papers, vol. 2*. Cambridge University Press. [GR]

Pylyshyn, Z. W. (1973) What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin* 80:1–24. [SH]

Pylyshyn, Z. W. (2007) *Things and places: How the mind connects with the world.* MIT Press. [EMar]

Quine, W. V. O. (1969) Natural kinds. In: *Ontological relativity and other essays*, ed. W. V. O. Quine, pp. 114–38. Columbia University Press. [aEM]

Regehr, G., Cline, J., Norman, G. R. & Brooks, L. R. (1994) Effect of processing on diagnostic skill in dermatology. *Academic Medicine* I:S34–S36. [rEM]

Rehder, B. (2003a) A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:1141–59. [aEM]

Rehder, B. (2003b) Categorization as causal reasoning. *Cognitive Science* 27:709–48. [TL]

Rehder, B. (2006) When causality and similarity compete in category-based property induction. *Memory and Cognition* 34:3–16. [TL, aEM]

Rehder, B. & Kim, S. (2006) How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:659–83. [aEM]

Rehder, B. & Murphy, G. L. (2003) A knowledge-resonance (KRES) model of category learning. *Psychological Bulletin & Review* 10:789–94. [BKH]

Rey, G. (1983) Concepts and stereotypes. *Cognition* 15:237–62. [aEM, GR]

Rey, G. (1985) Concepts and conceptions: A reply to Smith, Medin and Rips. *Cognition* 19:297–303. [aEM, GR]

Rey, G. (2009a) Concepts, defaults, and internal asymmetric dependencies: Distillations of Fodor and Horwich. In: *The a priori and its role in philosophy*, ed. N. Kompa, C. Nimtz & C. Suhm. Mentis, pp. 185–203. Mentis. [GR]

Rey, G. (2009b) Review of E. Machery, *Doing without Concepts. Notre Dame Philosophical Reviews.* (Online journal. Epub: 2009.07.15.) Available at http://ndpr.nd.edu/review.cfm?id=16608 [KE, aEM]

Rips, L. J. (1989) Similarity, typicality, and categorization. In: *Similarity and analogical reasoning*, ed. S. Vosniadou & A. Ortony, pp. 21–59. Cambridge University Press. [TB, TL, aEM]

Roediger, H. L. (2008) Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology* 59:225–54. [BKH]

Rosch, E. & Mervis, C. B. (1975) Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology* 7:573–605. [TB, aEM]

Roth, E. M. & Shoben, E. J. (1983) The effect of context on the structure of categories. *Cognitive Psychology* 15:346–78. [aEM]

Rubenstein, A. J., Kalakanis, L. & Langlois, J. H. (1999) Infant preferences for attractive faces: A cognitive explanation. *Developmental Psychology* 35:848–55. [YY]

Russell, B. (1948) *Human knowledge: Its scope and its limits.* Routledge & Kegan Paul. [aEM]

Sabsevitz, D., Medler, D., Seidenberg, M. & Binder, J. (2005) Modulation of the semantic system by word imageability. *NeuroImage* 27:188–200. [GD]

Samuelson, L. K., Schutte, A. R. & Horst, J. S. (2009) The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalizations. *Cognition* 110:322–45. [HAV]

Samuelson, L. K. & Smith, L. B. (1998) Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development* 69:94–104. [HAV]

Sandhofer, C. M. & Doumas, L. A. A. (2008) Order of presentation effects in the learning of color categories. *Journal of Cognition and Development* 9:194–221. [HAV]

Sandhofer, C. M. & Thom, E. E. (2006) Taking the task seriously: Reflections on measures of color acquisition. *Journal of Experimental Child Psychology* 94:344–48. [HAV]

Sato, J. J., Wolsan, M., Minami, S., Hosoda, T., Sinaga, M. H., Hiyama, K., Yamaguchi, Y. & Suzuki, H. (2009) Deciphering and dating the red panda's ancestry and early adaptive radiation of Musteloidea. *Molecular Phylogenetics and Evolution* 53(3):907–22. [AJJ]

Schneider, S. (2009) The nature of symbols in the language of thought. *Mind and Language* 24(4):523–53. [SS]

Schneider, S. (forthcoming) *The language of thought: New philosophical directions.* MIT Press. [SS]

Schwanenflugel, P. J. & Shoben, E. (1983) Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:82–102. [GD]

Segal, G. (2000) *A thin book about narrow content*. MIT Press. [GR]

Shafto, P., Coley, J.D. & Baldwin, D. (2007) Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review* 14:890–94. [BKH]

Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [aEM]

Shepard, R. N. & Cooper, L. A. (1982) *Mental images and their transformations.* MIT Press/Bradford Books. [SH]

Shepard, R. N., Hovland, C. I. & Jenkins, H. M. (1961) Learning and memorization of classifications. *Psychological Monographs: General and Applied* 75(13):1–42. (Whole No. 517). [SSK]

Siegler, R. S. (1994) Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science* 3:1–5. [HAV]

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A. & Barsalou, L. W. (2007) A common neural substrate for perceiving and knowing about color. *Neuropsychologia* 45:2802–10. [aEM]

Simon, H. A. (1995) Machine as mind. In: *Android epistemology*, ed. K. M. Ford, C. Glymour & P. J. Hayes, pp. 23–40. MIT Press. [aEM]

Sloman, S. A. (1993) Feature-based induction. *Cognitive Psychology* 25:231–80. [aEM]

Sloman, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [aEM]

Sloman, S. A. & Lagnado, D. (2005) The problem of induction. In: *The Cambridge handbook of thinking and reasoning*, ed. K. Holyoak & R. Morrison, pp. 95–116. Cambridge University Press. [aEM]

Sloutsky, V. M. & Fisher, A. V. (2008) Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development* 79:639–51. [HAV]

Smith, E. E., Patalano, A. L. & Jonides, J. (1998) Alternative strategies of categorization. *Cognition* 65:167–96. [arEM]

Smith, E. E. & Sloman, S. A. (1994) Similarity- vs. rule-based categorization. *Memory and Cognition* 22:377–86. [TB]

Smith, J. D. (2002) Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science* 13:437–42. [aEM, SZ]

Smith, J. D. & Minda, J. P. (1998) Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1411–36. [JJC, JAH, arEM, SZ]

Smith, J. D. & Minda, J. P. (2000) Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:3–27. [JJC, aEM]

Smith, J. D. & Minda, J. P. (2001) Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:984–1002. [SZ]

Smith, J. D. & Minda, J. P. (2002) Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:800–11. [SZ]

Smith, J. D., Murray, M. J., Jr. & Minda, J. P. (1997) Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory and Categorization* 23:659–80. [SZ]

Smith, L. B. & Samuelson, L. K. (1997) Perceiving and remembering: Category stability, variability and development. In: *Knowledge, concepts, and categories*, ed. K. Lamberts & D. Shanks, pp. 161–96. MIT Press. [aEM]

Smith, L. B., Thelen, E., Titzer, R. & McLin, D. (1999) Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review* 106:235–60. [HAV]

Sobel, D. M. & Kirkham, N. Z. (2007) Interactions between causal and statistical learning. In: *Causal learning: Psychology, philosophy, and computation*, ed. A. Gopnik & L. E. Schulz, pp. 139–53. Oxford University Press. [YY]

Solomon, K. O. & Barsalou, L. W. (2001) Representing properties locally. *Cognitive Psychology* 43:129–69. [aEM]

Solomon, K. O. & Barsalou, L. W. (2004) Perceptual simulation in property verification. *Memory and Cognition* 32:244–59. [aEM]

Spencer, J. P. & Schöner, G. (2003) Bridging the representational gap in the dynamic systems approach to development. *Developmental Science* 6:392–412. [HAV]

Sperber, D. & Wilson, D. (1995) *Relevance: Communication and cognition*, 2nd ed. Blackwell. [EMar]

Sperber, D. & Wilson, D. (1998) The mapping between the mental and the public lexicon. In: *Language and thought: Interdisciplinary themes*, ed. P. Carruthers & J. Boucher, pp. 184–200. Cambridge University Press. [aEM]

Spivey, M. (2007) *The continuity of mind.* Oxford University Press. [ABM]

Strevens, M. (2000) The essentialist aspect of naïve theories. *Cognition* 74:149–75. [DD]

Stich, S. P. (1983) *From folk psychology to cognitive science.* MIT Press. [aEM]

Stich, S. P. (1996) *Deconstructing the mind.* Oxford University Press. [aEM]

Storms, G., De Boeck, P. & Ruts, W. (2000) Prototype and exemplar based information in natural language categories. *Journal of Memory and Language* 42:51–73. [JAH]

Swaab, T., Baynes, K. & Knight, R. (2002) Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research* 15:99–103. [GD]

Tenenbaum, J. B., Griffiths, T. L. & Niyogi, S. (2007) Intuitive theories as grammars for causal inference. In: *Causal learning: Psychology, philosophy, and computation*, ed. A. Gopnik & L. Schulz, pp. 301–22. Oxford University Press. [aEM]

Thelen, E. & Smith, L. B. (1994) *A dynamic systems approach to the development of cognition and action.* MIT Press. [HAV]

Thompson-Schill, S. L. (2003) Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia* 41:280–92. [aEM]

Turing, A. M. (1950) Computing machinery and intelligence. *Mind* 49:433–60. [SH]

Uttal, W. R. (2001) *The new phrenology.* MIT Press. [ABM]

van Geert, P. & van Dijk, M. (2002) Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior and Development* 25:340–74. [HAV]

Van Orden, G. C., Pennington, B. F. & Stone, G. O. (2001) What do double dissociations prove? *Cognitive Science* 25:111–72. [aEM]

Verguts, T. & Fias, W. (2009) Similarity and rules united: Similarity- and rule-based processing in a single neural network. *Cognitive Science* 33:243–59. [YY]

Vigo, R. (2009) Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology* 53:203–21. [SSK]

Vlach, H. A., Sandhofer, C. M. & Kornell, N. (2008) The spacing effect in children's memory and category induction. *Cognition* 109:163–67. [HAV]

Wattenmaker, W. & Shoben, E. (1987) Context and the recallability of concrete and abstract sentences. *Journal of Experimental Psychology* 13:140–50. [GD]

Weisberg, J., van Turrennout, M. & Martin, A. (2007) A neural system for learning about object function. *Cerebral Cortex* 17:513–21. [rEM]

Weiskopf, D. A. (2007) Concept empiricism and the vehicles of thought. *Journal of Consciousness Studies* 14:156–83. [GD]

Weiskopf, D. A. (2009a) Atomism, pluralism, and conceptual content. *Philosophy and Phenomenological Research* 79:130–62. [DAW]

Weiskopf, D. A. (2009b) The plurality of concepts. *Synthese*169(1):145–73. [EL, AS, DAW]

Weiskopf, D. A. (2010) Concepts and the modularity of thought. *Dialectica* 64:107–30. [DAW]

Weiskopf, D. A. (forthcoming) The functional unity of special science kinds. *British Journal for the Philosophy of Science.* [DAW]

Whitney, P., McKay, T., Kellas, G. & Emerson, W. A., Jr. (1985) Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning Memory and Cognition* 20:804–23. [arEM]

Williams, J. J. & Lombrozo, T. (in press) The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science.* [TL]

Wisniewski, E. J. & Medin, D. L. (1994) On the interaction of theory and data in concept learning. *Cognitive Science* 18:221–81. [AS]

Yamauchi, T., Love, B. C. & Markman, A. B. (2002) Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(3):585–93. [ABM]

Yamauchi, T. & Markman, A. B. (1998) Category learning by inference and classification. *Journal of Memory and Language* 39(1):124–48. [ABM]

Yaxley, R. H. & Zwaan, R. A. (2007) Simulating visibility during language comprehension. *Cognition* 105:229–36. [aEM]

Zaki, S. R. & Nosofsky, R. M. (2004) False prototype enhancement effects in dot pattern categorization. *Memory & Cognition* 32:390–98. [SZ]

Zaki, S. R. & Nosofsky, R. M. (2007) A high-distortion enhancement effect in the prototype-learning paradigm: Dramatic effects of category learning during test. *Memory & Cognition* 35:2088–96. [SZ]

Ziff, E. (1972) *Understanding understanding.* Cornell University Press. [aEM]