The Campaign for Concepts

TANIA LOMBROZO *University of California, Berkeley*

When it comes to the advertising campaign for concepts, most philosophers and psychologists agree: "Concepts: You can't explain thought without 'em!" But that's where agreement typically ends. Once you get beyond the slogans, philosophers and psychologists interested in conceptual representation have generated a variety of proposals that share little more than a commitment to the value – and sometimes necessity – of a theory of concepts. In his new book, *Doing without Concepts*, Edouard Machery takes a critical look at both disciplines' claims about concepts, and comes to a startling conclusion: We've been advertising a product that doesn't exist.

Concepts, often characterized as the building blocks of thought, are central to both philosophy and psychology. Within philosophy, claims about concepts have a diverse pedigree, but one thriving tradition identifies concepts with mental representations or *types* of mental representations. This tradition makes the most contact with psychology, where concepts are posited as a psychological kind invoked to explain judgments about category membership, the meaning of words, and aspects of reasoning. Philosophers often focus on questions of individuation and reference, where psychologists aim to characterize the mechanisms involved in concept acquisition, categorization, and inference, but the hope on both sides has been to converge on a common and robust theory of concepts that can explain the nature of thought. So claiming that we should do without concepts is a strong claim indeed.

Importantly, Machery does not deny the existence of mental representations. In fact, he argues that specific proposals about concepts, such as prototype, exemplar, and theory theories, each identify representational structures that support meaningful inductive generalizations, and hence merit status as

Dialogue **50 (2011), 165–177.**
© Canadian Philosophical Association/Association canadienne de philosophie 2011
doi:10.1017/S0012217311000175

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

psychological kinds. What he does deny is that the concept of "concept" is itself a psychological kind. These ambitious claims comprise *the heterogeneity hypothesis*, with the final tenet calling for the elimination of concepts altogether. Machery proceeds by providing evidence for prototypes, exemplars, and theories – three distinct representational structures – and arguing for their mutual existence. He then suggests that the representational structures share little in common, and do not *as a class* satisfy the criteria for a particular conception of what it takes to be a natural kind, namely supporting meaningful inductive generalizations. If concepts are not a natural kind, they have no role in a mature theory of human cognition.

The arguments for these conclusions, which comprise the bulk of the text, are pulled off with interdisciplinary finesse. The book is cognitive science fusion at its best: the clearly articulated arguments of philosophy with a generous serving of classic and contemporary empirical results. For garnish, Machery offers brief but trenchant observations on topics that arise along the way, such as concept individuation and the ecological validity of research on categorization. These qualities make the book a pleasure to read – whether or not one ultimately agrees.

While I find much to praise and endorse in Machery's book, I'll focus on a few points of disagreement. Specifically, I'll attempt a defense of concepts, in part because I'm genuinely sympathetic to a more unified account of conceptual representation and in part because I think it's useful to see how far a unified theory can get. After introducing the heterogeneity hypothesis (section 2), I'll consider whether the evidence for prototypes, exemplars, and theories requires the kind of pluralism Machery endorses. And after introducing the final tenet of the heterogeneity hypothesis, which I refer to as *the elimination thesis* (section 3), I'll consider whether a different concept of concept might eschew elimination, and what the consequences of denying concepts amount to. But before turning to these claims, a point of definition (section 1): what are concepts, such that they don't exist?

Defining Concepts

Machery begins by proposing a definition of "concept." Specifically, he offers the following as an articulation of commitments that often underlie psychological claims:

> The psychologist's definition: "A concept of x is a body of knowledge about x that is stored in long-term memory and is used by default in the processes underlying most, if not all, higher cognitive competences when these processes result in judgments about x." (*DwC*, p.12)

So, for example, my concept of chocolate is a bundle of beliefs about chocolate that is usually consulted in judgments about membership in the category "chocolate," in inductive or deductive reasoning about chocolate, in drawing

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

analogies about chocolate, and so on. Although Machery considers philosophers' definitions and also ways to link the philosophical and psychological projects, he ultimately rejects a simple reconciliation. He instead adopts the psychological project with the stated goal of developing "a new picture of the organization of our knowledge in long-term memory" (*DwC*, p. 52). Psychologist's definition in hand, Machery sets forth to examine what, if anything, counts as a concept. He ultimately concludes that prototypes, exemplars, and theories successfully satisfy the definition, but do not as a class form a psychological kind.

Before moving on to the heart of Machery's arguments, a quibble and a qualm. The quibble has to do with identifying concepts with bodies of knowledge. If concepts are bodies of knowledge, in terms of what are the bodies of knowledge articulated? The options, as far as I can tell, are threefold: they are articulated in terms of other concepts (bodies of knowledge all the way down?), in terms of a special subset of primitive concepts (so some concepts aren't bodies of knowledge?), or in terms of sub-conceptual representations (why aren't *those* the concepts?). Each option opens a can with its own species of invertebrate. Why not identify concepts with constituents of bodies of knowledge, and leave the worms for another day? I'll return to this point after introducing the elimination thesis in section 3.

The qualm is subtle but substantive. Machery begins with a definition of concepts, and goes on to consider what, if anything, satisfies that definition. But on my reading of both the philosophical and psychological literatures, concepts are the *explanans*, not the *explanandum*. That is, concepts are posited to explain aspects of cognition: productivity, systematicity, judgments of category membership, typicality effects, and so on. Concepts are not an independent phenomenon in need of explanation, except insofar as theories of concepts have inadvertently reified them. To illustrate the point, consider atoms. In modern chemistry, atoms initially were posited as the indivisible units of matter, invoked to explain why elements react in ratios of whole numbers, why chemical methods could not break down basic substances beyond a certain point, and so on. One could investigate atoms by examining what, if anything, satisfies a definition such as "indivisible units of matter" (answer: not atoms). Alternatively, and I would argue more fruitfully, one could ask: What *does* explain why elements react in ratios of whole numbers? When it comes to concepts, why not start with the real explananda: the phenomena associated with conceptual thought, such as productivity, systematicity, judgments of category membership, and so on?

To be fair, many psychologists do treat concepts as the explanandum, and Machery's project is ultimately a productive one despite my quibble and my qualm. But the proposed definition and the decision to begin with an answer rather than a question bear on what Machery ultimately rejects when he advises psychology to do without concepts: he is rejecting the idea that the definition picks out a real psychological kind. He has not rejected the possibility of

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

answering the questions concepts are posited to explain: how thought can be productive, how judgments of category membership are made, and so on. One could, for example, deny that thought is productive, that the mind represents categories, or that we can ever understand the nature of thought. (I'm not suggesting these claims are wise, only that they are possible.) Might beginning with the phenomena associated with conceptual thought lead to a definition of concepts that does pick out a psychological kind? While Machery does not proceed in this direction, his own proposal – the heterogeneity hypothesis – suggests how he would respond (hint: "No."). It's to this proposal that I now turn.

The Heterogeneity Hypothesis

According to the heterogeneity hypothesis, there are several, distinct kinds of representations in long-term memory that function as concepts – that is, that are used by default in the processes underlying higher cognitive competences. At minimum, there are three: prototypes, exemplars, and theories. Machery provides clear and compelling presentations of each kind of representation, including the evidence typically taken to support the existence of that representational kind. For the purposes of this review, a short paragraph explaining all three representational kinds will have to suffice, but readers who wish to learn more will benefit from Machery's exposition (see also Murphy, 2002, for a psychologist's perspective).

Briefly, "prototypes" summarize categories by representing the features category members tend to possess. Membership in a category is assessed in terms of the similarity between the item being categorized and the prototypes of candidate categories. In contrast, "exemplars" are representations of *particular* category members. The category representation, then, is a set of exemplars, and category membership can be determined by computing similarity to this set. Finally, "theories" are intuitive sets of beliefs that involve causal-explanatory content. For example, my theory of mind specifies relationships between beliefs, desires, and actions, and can be used to predict and explain people's behaviour. Some advocates of this so-called "theory theory" identify concepts directly with theories, while others identify concepts with theoretical terms within theories (e.g., see Murphy and Medin versus Carey, 1985). For Machery, to advocate theories is to claim that some of the bodies of knowledge that underlie higher cognitive competencies support explanations.

What makes the heterogeneity hypothesis radical is not the representational kinds it accepts, but the commitments it denies: alternatives Machery dubs the received view, scope pluralism, competence pluralism, and hybrid theories. I'll consider each of these alternatives in turn. But to be clear, my aim is to articulate the central tenets of the heterogeneity hypothesis by contrasting them with plausible alternatives; I do not attempt to reproduce the arguments and evidence that accompany Machery's own development of the hypothesis.

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

The received view is a commitment to unification. Despite conceptual diversity, the received view maintains that concepts share important properties in common, including their basic structure and their role in reasoning. To the extent psychologists regard prototype, exemplar, and theory-theories as mutually exclusive, they betray an assumption to the received view that one kind of representation will prevail as a unified account of conceptual representation.

While not all psychologists endorse the received view, the burden of proof has been on detractors. And detractors there are. Within psychology, two kinds of pluralism have been proposed. For a *scope* pluralist, concepts of different kinds involve different structures. For example, one might endorse definitions as the representational kinds underlying mathematical concepts, but prototypes as the representations underlying artifact concepts. There will still be a single concept of triangle or of wrench, but triangle and wrench need not be represented by a common representational kind.

For a *competence* pluralist, triangles and wrenches can have multiple representations: prototypes, exemplars, and mental billboards with flashing neon signs. The critical idea is that which representation is involved in a given judgment will depend on the cognitive competence for which the concept is invoked. Perhaps categorization involves your triangle prototype, while inductive reasoning involves your triangle theory.

The heterogeneity hypothesis endorses the one-to-many mapping between concepts and representational kinds of a competence pluralist. That is, there can be multiple representations of dog, of triangle, and of wrench. You can have a dog prototype, dog exemplars, and a dog theory. Where Machery rejects competence pluralism is in its assumption of a one-to-one mapping between representational kinds and cognitive competences. Thus, according to the heterogeneity hypothesis, a concept of x involves multiple bodies of knowledge that belong to distinct representational kinds, *and there is a many-to-many mapping between representational kinds and the higher cognitive competences for which they are invoked*. Sometimes you categorize with your dog exemplar; other times with your dog theory. Machery is open to the possibility that there's some systematicity – with particular tasks and contexts leading to the use of particular representations – but that systematicity won't take the form of a one-to-one mapping at the level of cognitive competences.

A final foil for Machery comes in the form of hybrid theories of concepts. Hybrid theories can take several forms, but share the claim that concepts involve a combination of representational structures. For example, your concept of "dog" might include a prototype *and* a theory. Advocates for hybrid theories have tended to regard these distinct representational structures as *parts* of a single concept rather than as distinct concepts, and thereby differ from the heterogeneity hypothesis's bolder endorsement of multiple concepts. But what hinges on this distinction? Machery points out that hybrid theorists rarely offer an explicit argument for regarding the hybridized components as part of a single concept, but he fills in this gap by suggesting that for a hybrid theorist,

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

the components are necessarily linked to each other and coordinated in judgment. By positing distinct concepts, the heterogeneity hypothesis does not share these commitments. In fact, the heterogeneity hypothesis explicitly denies a particular kind of coordination, as it posits that "coreferential concepts have very few properties in common" (*DwC*, p. 52).

The decisive evidence for Machery's brand of pluralism would come from a demonstration that a single person at a single point in time possesses multiple bodies of knowledge about x, where (a) those bodies of knowledge belong to distinct representational kinds (e.g., prototypes, exemplars, theories), (b) different representational kinds are invoked for the very same competence (e.g., categorization), (c) a given representational kind (e.g., the prototype) is invoked for more than one competence (e.g., categorization *and* deduction), and (d) the distinct representational kinds about x are not necessarily linked or meaningfully coordinated.

Machery reviews a range of suggestive findings. He offers evidence that different experimental results are best accounted for by prototypes, exemplars, or theories, where the findings vary in terms of scope (the domain of concepts involved) and competence (the kind of judgment involved). However, the evidence falls short of establishing the simultaneous co-existence of different representations for x, or of establishing that those representations are not meaningfully coordinated. The absence of critical experiments is not an oversight on Machery's part, but a genuine gap in the experimental literature. While Machery has not decisively eliminated alternative forms of pluralism, he has, arguably, shifted the burden of proof. Why endorse unity in the face of so much heterogeneity?

I'll argue that we need not abandon unification. At least not yet. In support, I offer recent work suggesting that prototypes and exemplars can be subsumed under a single representational system that coordinates representations and their roles in supporting cognitive competences. This alternative picture would challenge all four empirical commitments of Machery's proposal. I'll first say something about how to unify prototypes and exemplars, and then move on to the trickier business of theories.

Unifying Prototypes and Exemplars

Prototype and exemplar theories historically have been competitors. Where prototype theories propose a single summary representation for an entire category (one representation for N encountered members), exemplar theories propose sets of representations of particulars (N representations for N members). And as Machery (rightly) points out, while both kinds of models rely on similarity, they tend to employ different similarity metrics. However, Machery also (rightly) notes that there's no reason a prototype theory couldn't adopt an exemplar theory's similarity metric, and vice versa. Recent theoretical and empirical work suggests that prototype and exemplar theories can be brought into even closer alignment. In particular, some have advocated a more unified view

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

according to which prototypes and exemplars are endpoints on a continuum that specifies something like "abstraction" (Vanpaemel and Storms; see also Rosseel; Love, Medin, and Gureckis; Griffiths *et al*.). On this view, a single mechanism for learning and inference coordinates a concept's location on this prototype-exemplar continuum.

In one respect this view is *more* heterogeneous than the heterogeneity hypothesis, for it recognizes a continuum of representational structures that lie between a single prototype and N exemplars: a structure with 2 representations for N members, another with 3 representations for N members, and so on. (Consider a category like "dog," which may involve distinct summary representations for "small lap dog" versus "large guard dog.") But in another important respect, this proposal is *less* heterogeneous than the heterogeneity hypothesis, for it posits a single representational structure for a given concept at a given time. And this is what the evidence supports.

Consider an experiment by Smith and Minda, in which participants learned two novel categories consisting of strings of letters. The category structures were constructed such that each category had a prototypical sequence of letters, but some category members were more similar to the prototype from the *opposite* category. Initially participants misclassified these items, but with experience their performance improved. Smith and Minda found that a prototype model predicted participants' early judgments better than did an exemplar model, but that later judgments were better predicted by an exemplar model. More revealing, Griffiths *et al*. developed a computational model capable of shifting along the prototype-exemplar continuum as evidence about category structure accumulates. Their model was able to capture the prototype-to-exemplar switch in the Smith and Minda data, as well as the transitional stages in which an intermediate representational structure seemed to underlie judgments.

If participants have a single representational structure at a given time, that's bad news for the heterogeneity hypothesis. Studies like that of Smith and Minda can't rule out the possibility that additional representational structures were simultaneously available, but they aren't required to explain the data. Moreover, evidence that representations shift along a prototype-exemplar continuum in response to evidence about category structure suggests that prototype and exemplar representations participate in the same processes, and that prototype-like components and exemplar-like components *are* intricately coordinated, contra Machery's rejection of hybrid theories.

Unifying Theories: The Negative Argument

Assimilating theories to this prototype-exemplar continuum would be a coup for unification. Unfortunately, I don't think it's quite so easy. The "theories" (or "theory-theory") perspective encompasses a broad range of claims and phenomena. Often, arguments *for* theories are in fact arguments *against* prototypes and exemplars. It's thus useful to distinguish the theory-theory as a

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

negative argument from its positive claims about the nature of conceptual representation.

Theory-theorists often point to the gaps and presuppositions in prototype and exemplar theories. For example, what determines the features of category members that are encoded in prototypes and exemplars? Most prototype and exemplar theories take an initial vocabulary of features for granted. A theory-theorist might invoke intuitive theories to help explain why "self-animated" is a feature in representing living things, while "rhymes with 'chocolate'" is not. More commonly, theory-theorists question the existence of a master metric for similarity, which prototype and exemplar theories both require. In support, theory-theorists marshal an impressive set of findings suggesting that evaluations of global similarity do not govern categorization, that similarity judgments can vary with context, and that any two objects can be similar or dissimilar in an infinite number of respects, only a small subset of which will be relevant for a given judgment (e.g., Murphy and Medin; Rips; Heit and Rubenstein; Goodman; see Murphy, 2002 for review).

Not all theory-theorists would agree, but one way to characterize theories is as mechanisms for specifying the similarity relation that's relevant for a given judgment (how this is supposed to occur is a mystery I won't dwell on here). So if a combined prototype/exemplar model provides the machinery, it's theories that constrain the input and set the operating parameters. However, taking theory-theorists' negative arguments to heart isn't so much a reason for pluralism as a vote of no confidence in prototypes/exemplars. At minimum, the prototype/exemplar story is incomplete.

Unifying Theories: The Positive Argument

What of the theory-theorists' positive claims? Machery summarizes the theory paradigm's two core claims as follows: "concepts are bodies of knowledge that underlie explanations" and "concepts are organized by domains" (*DwC*, p. 103). Within psychology, theory theorists additionally invoke the idea that intuitive theories share important properties with scientific theories (e.g., Carey 1985, 2009; Gopnik and Meltzoff). Machery questions the value of this analogy, however, pointing out that scientific theories are themselves enormously diverse and only partially understood. In particular, scientific theories don't share a uniform structure. What they do share with each other and with intuitive theories is a common function: explanation. Theories are supposed to explain phenomena, not merely to describe them. It's for this reason that Machery identifies theories with bodies of knowledge that support explanations, and not in terms of structural properties such as specifying laws or causal mechanisms.

While the theory paradigm relies on explanation as a core notion to distinguish theories from other kinds of knowledge structures, Machery points out that accounts of explanation have been but minimally developed. Philosophers of science have offered several accounts of scientific explanation (not themselves without problems), but "psychologists rely on a folk understanding of

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

explanation" (*DwC*, p. 102), which is itself supported by "folk examples of explanation" (*DwC*, p. 103).

Machery is right to focus on explanation as the key to the theory-theory position, and to underscore an inappropriate reliance on folk intuitions. One of the central aims of my own research has been to fill this gap (Lombrozo, 2006). Work from my lab and others has examined how explanations are evaluated (Lombrozo and Carey, 2006; Keil; Lombrozo, 2007; Bonawitz and Lombrozo) and their role in conceptual representation (Prasada and Dillingham, 2006, 2009; Lombrozo, 2009; Williams and Lombrozo). For example, Lombrozo (2009) finds that the explanation a participant offers for a category feature – either spontaneously or through experimental prompting – predicts which features will be weighted most heavily in subsequent categorization judgments. One interpretation is that explanations determine how similarity between known and candidate category members is evaluated. This finding is thus consistent with the picture sketched at the end of the previous section, in which theories specify the respects in which items must be similar to support a given judgment (in this case, a judgment of category membership).

Once viable psychological theories of explanation have been proposed, the theory-theory will be on firmer footing and its relationship to prototypes and exemplars reassessed. While I'm not optimistic that this reassessment will yield a neat unification between prototype/exemplars and theories, it may be that these two families of representational structures conform to a plausible functional architecture, with prototype/exemplars tracking descriptive information and theories tracking explanatory and normative information.

The Elimination Thesis

In the concluding chapter of *Doing without Concepts*, Machery advocates concept eliminativism, the final tenet of the heterogeneity hypothesis. I'll refer to his formulation as the elimination thesis:

> Elimination thesis: "The notion of concept ought to be eliminated from the theoretical vocabulary of psychology because it might prevent psychologists from correctly characterizing the nature of the knowledge in long-term memory and its use in cognitive processes." (*DwC*, p. 220)

Machery's brand of eliminitivism does not hinge on whether *something* satisfies the definition of concept. As we've seen, Machery believes that *many things* do, among them prototypes, exemplars, and theories. Rather, Machery considers whether the set of mental representations that satisfy the definition constitutes a "natural kind": a class of entities that supports scientifically relevant inductive generalizations. While prototypes, exemplars, and theories may each individually support such generalizations, Machery claims that the superordinate "concept," encompassing all three (and possibly more) representational kinds, does not.

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

There's an obvious reply, which Machery is quick to address. Surely the class of exemplars, prototypes, and theories supports *some* generalizations. In particular, aren't prototypes, exemplars, and theories all bodies of knowledge consulted by default in judgments concerning the higher cognitive competencies? That is, don't they all satisfy the definition of concept? Moreover, don't they serve a common explanatory role in accounting for phenomena like categorization, inductive reasoning, and so on? Machery writes that this objection "misconceives the nature of natural kinds " (*DwC*, p. 243), as "members of natural kinds have many properties in common *besides those properties that are used to identify them*" (*DwC*, pp. 243-4, emphasis mine). Natural kinds are classes about which we can discovery generalizations; if none exist for concepts, concepts are not a natural kind.

In section 1, I suggested that identifying concepts with the *constituents* of bodies of knowledge rather than with the bodies of knowledge themselves could potentially sidestep some of Machery's concerns. In particular, Machery's key arguments for elimination rest on the failure to find meaningful generalizations over the class that contains prototypes, exemplars, and theories. On a "constituent" definition of concept, concepts would figure in very different kinds of bodies of knowledge – prototypes, exemplars, and theories – but it's not clear that the structure of the constituents for these bodies of knowledge need be as diverse as the bodies of knowledge themselves. Moreover, the rules for how the constituents combine and interact may be common across bodies of knowledge, suggesting that this notion of concept could support meaningful generalizations of the sort that merit status as a psychological kind. If this is right, Machery's decision to identify concepts with bodies of knowledge significantly reduces the scope of his conclusion, as alternative definitions may resist elimination.

Suppose for the moment, though, that we accept Machery's definition; that the heterogeneity hypothesis is correct, and that concepts are not a natural kind. What follows? In the elimination thesis, Machery suggests that eliminating concepts will help foster an accurate characterization of the representations in long-term memory, as well as their role in cognition. He later elaborates that prototype, exemplar, and theory-theorists would do well to recognize the psychological reality of all three kinds, and to focus on developing the best account of each representational kind rather than striving for critical experiments to discredit alternative kinds. Researchers could also focus on characterizing the distinct cognitive processes involving prototypes, exemplars, and theories, including how they underwrite judgments. For example, if prototypes, exemplars, and theories all figure in categorization, what determines which representation and corresponding process is invoked on a given occasion, or how multiple outputs are coordinated when more than one representation is invoked?

I agree with Machery that these endeavours are valuable, and that psychological research and theory could benefit from a greater appreciation of the

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

heterogeneity in conceptual thought. However, this shift in focus seems to follow from Machery's pluralism rather than from his eliminativism. It's Machery's pluralism that motivates the value of characterizing distinct representational kinds and processes; what his eliminativism adds is a rejection of the superordinate "concept." This elimination may have consequences, but I'm not sure what, exactly, those consequences are. For a pluralist, representations can have a hierarchical structure, with "mental representations" at the top, "perceptual representations" and "concepts" below, and multiple conceptual structures subordinate to "concepts," such as prototypes, exemplars, and theories. Does rejecting concepts have consequences for the relationships between conceptual and perceptual representations? Would Machery go beyond his rejection of concepts to suggest that a superordinate like "mental representation" (or "long-term memory representation"?) itself supports no meaningful generalizations? These are substantive questions; Machery's replies would clarify what it really means to do without concepts.

Conclusions

Machery's provocative book raises a number of important questions, and articulates a viable and valuable position that has been – until now – absent from contemporary debates. If Machery is right, the cognitive science community has its work cut out for it, as does Machery (for example, writing the sequel: *Doing without Concepts: A Guide for Recovering Addicts*). But even if he's wrong, advocates for the received view and for more traditional forms of pluralism will have to address compelling arguments and evidence for heterogeneity and elimination. I've suggested a few ways in which to defend the campaign for concepts: by attempting to unify prototypes, exemplars, and theories, and by focusing on the constituents of bodies of knowledge rather than on the bodies of knowledge themselves. Ultimately these issues will be resolved with new empirical and theoretical work, but in the meantime I hope I've allayed fears of false advertising.[1]

Note

1 Sincere thanks to James Genone, Hannah Ginsborg, and Tom Griffiths for helpful comments and relevant conversations.

References

Bonawitz, L. B. and T. Lombrozo
Under Review "Occam's Rattle: How Simplicity and Probability Constrain Children's Inference."

Carey, S.
1985 *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books, MIT Press.
2009 *The Origin of Concepts*. New York: Oxford University Press.

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

Goodman, N.
1972 "Seven Strictures on Similarity." In *Problems and Projects*, ed. N. Goodman, 437–47. Indianapolis and New York: Bobbs-Merrill.
Gopnik, A. and A.N. Meltzoff
1998 *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.
Griffiths, T. L., A. N. Sanborn, K. R. Canini, and D. J. Navarro
2008 "Categorization as Nonparametric Bayesian Density Dstimation." In *The Probabilistic Mind: Prospects for Rational Models of Cognition*, ed. M. Oaksford and N. Chater. Oxford: Oxford University Press.
Heit, E. and J. Rubinstein
1994 "Similarity and Property Effects in Inductive Reasoning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 411–22.
Keil, F. C.
2006 "Explanation and Understanding." *Annual Review of Psychology* 57: 227–54.
Lombrozo, T.
2006 "The Structure and Function of Explanations." *Trends in Cognitive Sciences* 10: 464–70.
2007 "Simplicity and Probability in Causal Explanation." *Cognitive Psychology* 55: 232–57.
2009 "Explanation and Categorization: How 'Why?' Informs ''What?' " *Cognition* 110: 248–53.
Lombrozo, T. and S. Carey
2006 "Functional Explanation and the Function of Explanation." *Cognition* 99: 167–204.
Love, B. C., D. L. Medin, and T. M. Gureckis
2004 "SUSTAIN: A Network Model of Category Learning." *Psychological Review* 111: 309–32.
Machery, E.
2009 *Doing Without Concepts*. New York: Oxford University Press.
Murphy, G. L.
2002 *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Murphy, G. L. and D. Medin
1985 "The Role of Theories in Conceptual Coherence." *Psychological Review* 92: 289–316.
Prasada, S. and E. M. Dillinghman
2006 "Principled and Statistical Connections in Common Sense Conception. *Cognition* 99: 73–112.
2009 "Representation of Principled Connections: A Window onto the Formal Aspects of Common Sense Conception." *Cognitive Science* 33: 401–48.
Rips, L.
1989 "Similarity, Typicality, and Categorization." In *Similarity and Analogical Reasoning*, ed. S. Vosniadou and A. Ortony, 21–59. Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press

Rosseel, Y.
2002 "Mixture Models of Categorization." *Journal of Mathematical Psychology* 46: 178–210.
Smith, J. D. and J. P. Minda
1998 "Prototypes in the Mist: The Early Epochs of Category Learning." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24: 1411–36.
Vanpaemel, W. and G. Storms
2008 "In Search of Abstraction: The Varying Abstraction Model of Categorization." *Psychonomic Bulletin & Review* 15: 732–49.
Williams, J. J. and T. Lombrozo
2010 "The Role of Explanation in Discovery and Generalization: Evidence from Category Learning." *Cognitive Science* 34: 776–806.

https://doi.org/10.1017/S0012217311000175 Published online by Cambridge University Press