

Hume and cognitive science: The current status of the controversy over abstract ideas

MARK COLLIER

Stanford University, Stanford, CA, U.S.A. (E-mail: mcollier@stanford.edu)

Abstract. In Book I, Part I, Section VII of the *Treatise*, Hume sets out to settle, once and for all, the early modern controversy over abstract ideas. In order to do so, he tries to accomplish two tasks: (1) he attempts to defend an exemplar-based theory of general language and thought, and (2) he sets out to refute the rival abstraction-based account. This paper examines the successes and failures of these two projects. I argue that Hume manages to articulate a plausible theory of general ideas; indeed, a version of his account has defenders in contemporary cognitive science. But Hume fails to refute the abstraction-based account, and as a result, the early modern controversy ends in a stalemate, with both sides able to explain how we manage to speak and think in general terms. Although Hume fails to settle the controversy, he nevertheless advances it to a point from which we have yet to progress: the contemporary debate over abstract ideas in cognitive science has stalled on precisely this point.

Key words: abstract ideas, categorization, cognitive science, Hume, memory

In Book I, Part I, Section VII of the *Treatise*, Hume attempts to put an end, once and for all, to the controversy over abstract ideas. The controversy, as he sees it, involves a dispute over “*abstract or general ideas, whether they be particular or general in the mind’s conception of them*” (Hume 1739, p. 17). We must choose, then, between two alternatives: either general ideas are represented in the mind by abstract and indeterminate conceptions, or they are represented by particular exemplars with completely determinate properties (1739, p. 18). In order to settle this controversy, Hume attempts to accomplish two separate tasks in this section. The first is constructive: Hume sets out, against the conventional wisdom of the time, to defend the plausibility of the exemplar account. The second is destructive: Hume attempts to eliminate, or at least render superfluous, the abstraction-based account. This paper tells the tale of these two projects.

Hume’s defense of the exemplar alternative

Philosophers have traditionally rejected the exemplar alternative, Hume maintains, on the grounds that it assumes an “infinite capacity in the mind.”

The abstract idea of a man represents men of all sizes and all qualities; which ’tis concluded it cannot do, but either by representing at once all possible sizes and all possible qualities, or by representing no particular one at all. Now it having been esteemed absurd to defend

the former proposition, as implying an infinite capacity in the mind, it has been commonly infer'd in favour of the latter; and our abstract ideas have been suppos'd to represent no particular degree either of quantity or quality (1739, p. 18).

If one denies abstraction, so the objection goes, one is forced to claim that we can represent “at once” an indefinitely long disjunction of particular instances. But such a feat is well beyond the cognitive powers of finite creatures such as ourselves; only an infinite mind like God can “survey an entire range of particulars at once” (Winkler 1989, p. 35). Thus, the traditional “infinite capacity” objection amounts to a *reductio ad absurdum* of the exemplar account; we must accept the doctrine of abstraction, then, since there is no viable alternative.

One of Hume’s primary goals in the *Treatise* 1.1.7 is to defend the exemplar account from this traditional refutation. Hume admits that our finite human minds cannot survey an infinite range of particulars. But he denies that the exemplar alternative requires that we be able to do so.

[T]ho’ the capacity of the mind be not infinite, yet we can at once form a notion of all possible degrees of quantity and quality, in such a manner at least, as, however imperfect, may serve all the purposes of reflexion and conversation (1739, p. 18).

Thus, the infinite capacity objection misses its mark: the exemplar alternative need not account for the unbounded understanding of a divine mind, but only the limited and imperfect apprehension of finite creatures such as ourselves.

Nevertheless, one might respond, even our limited and imperfect comprehension poses a serious problem for defenders of the exemplar account. After all, consider the extension of a complex general idea such as MAN. Mature human adults will have become acquainted with a vast number of different men over the course of their lives. Even a simple general idea such as WHITE picks out an immense variety of shades, ranging from the pure white of freshly fallen snow to the dirty white of a bow tie that has been lying on the floor for several years (Price 1953, p. 14). So the question is: how can defenders of the exemplar account explain, without appeal to abstraction, how we manage to represent, at once, such a bewildering variety of instances?

Hume responds to this challenge by drawing on the resources of his science of human nature. He agrees with Bishop Berkeley that general ideas are represented in the mind by particular instances, which we conceive with all their “particular circumstances and proportions” (1739, p. 20).¹ But Hume supplements Berkeley’s theory of general ideas with an additional component. When we hear a general term, according to Hume, it not only calls to mind a particular exemplar, but it also revives each of the various instances to which the term has been applied in the past (1739, p. 21). Thus, according to Hume’s

version of the exemplar account, a general idea is represented in the mind not only by a particular exemplar, but also by a network of related instances.

Hume maintains that this additional component, which has become known as the “revival set”, can be explained in terms of the principles of association.²

A particular idea becomes general by being annex'd to a general term; that is, to a term, which from a customary conjunction has a relation to many other particular ideas, and readily recalls them in the imagination (1739, p. 22).

When a singular term has been constantly conjoined with a particular idea, according to Hume, they become associated in the faculty of the imagination, such that the next time we hear the term, the idea will be automatically called to mind. Similarly, when a general term is correlated over time with several different particular ideas, a “one word-many idea” association becomes established in the imagination (Zartmann 1975, p. 257). The next time we hear the general term, then, it will revive each of the instances to which it has been applied.

This is not to say, of course, that we become consciously aware of the various contents of the revival set. Hume recognizes that there is a strict limit to the number of ideas that we can entertain at any one time. Nevertheless, we have access to the rest of the instances in the revival set, as Aaron puts it, “at a lower plane, as it were, than the fully conscious plane of the adult human mind” (Aaron 1941, p. 122).

Thus when the term Horse is pronounced, we immediately figure to ourselves the idea of a black or a white animal, of a particular size or figure: but as that term is also usually applied to animals of other colors, figures, and sizes, these ideas, though not actually present to the imagination, are easily recalled; and our reasoning and conclusion proceed in the same way, as if they were actually present (Hume 1748, p. 158n).

Hume is able to explain, without abstraction, how the mind can represent a large variety of instances: even though our conscious attention is extremely limited in scope, we nevertheless have access to the remainder of the revival set at the level of custom.³

Hume’s argument for the impossibility of abstract ideas

Hume defends the exemplar alternative by showing that it does not make any implausible assumptions about the capacities of the mind. Indeed, his version appeals to nothing more than elementary principles of association. But Hume has not yet achieved his desired goal of putting an end to the controversy over abstract ideas. In order to accomplish this more difficult task, he has to show not only that the exemplar account offers a possible explanation of the

nature of general ideas, but that it offers the only possible explanation. Thus, in what we might call his negative phase, Hume attempts to turn the tables on his opponents and prove that it is really the doctrine of abstraction, with its appeal to indeterminate mental representations, which is “utterly impossible” (1739, p. 18).

Hume presents three arguments for the impossibility of abstract ideas. The first is that the very notion of an abstract general idea is simply incoherent. The doctrine of abstraction requires that we form indeterminate representations; but it is “evident at first sight” that we cannot mentally represent a quality such as LENGTH without imagining a particular line, nor can we represent a general idea such as LINE without imagining a particular length (1739, pp. 18–19). This argument only has force, however, if one assumes that all ideas are images, which proponents of abstraction need not do; moreover, even if they did, there are those who maintain that reasonable sense can be made out of the notion of an indeterminate or “sketchy” mental image (Bennett 1971, p. 41; Jolley 1999, p. 51).

Hume’s second argument for the impossibility of abstract ideas is that they violate the Copy Principle. Proponents of abstract ideas accept that all of our sensory impressions are determinate in character. Ideas are mere “copies” of impressions, though, and do not differ from them in content, but only in their respective degrees of force and vivacity. Considerations of parity, therefore, require defenders of abstract ideas to hold that all of our ideas are determinate as well.

An idea is a weaker impression; and as a strong impression must necessarily have a determinate quantity and quality, the case must be the same with its copy or representative (1739, p. 19).

But this argument, like the first, relies upon a premise that defenders of abstraction need not accept. Locke, for example, rejects the claim that all ideas are copies of impressions. All *particular* ideas are copies of sensations, according to Locke, but these ideas serve in turn as the raw material from which abstract ideas (not themselves copies) are actively constructed (1690, p. 159).

Hume’s third and final attempt to refute the doctrine of abstract ideas appeals to ontological considerations. Both sides in the controversy are nominalists who hold that “everything in nature is individual” (1739, p. 19). But conceivability entails possibility, and thus if abstract or indeterminate objects cannot exist in the world, it follows that abstract and indeterminate ideas cannot exist in the mind, either: “if this therefore be absurd in *fact and reality*, it must also be absurd *in idea*” (1739, p. 19). But once again, Hume attacks a straw man. Proponents of abstraction need not commit themselves to the claim that abstract ideas pick out abstract or indeterminate objects in the

world. Locke, for example, maintains that abstraction is nothing but a “fiction and contrivance of the mind” (1690, p. 596). The absurdity of abstract or indeterminate objects, therefore, does not entail the absurdity of abstract ideas.

Hume’s argument for the superfluosity of abstract ideas

In the *Essay Concerning Human Understanding*, Locke argues that abstract ideas are necessary in order to explain our ability to perform three cognitive tasks. First, we must posit abstract ideas in order to account for our capacity to understand the meaning of general terms. According to Locke’s semantics, singular terms such as “Fido” and “Edinburgh” get their meaning by standing for particular ideas in the mind, but general terms such as “Dog” or “City” do not represent any particulars, and so they must get their meaning by standing for abstract ideas (1690, p. 412). Second, abstract ideas are necessary in order to explain how we carry out geometrical proofs. If we have a particular triangle in mind when we carry out our demonstrations, we could never be sure that our results apply to *all* triangles. Thus, we must perform our proofs with the abstract idea of TRIANGLE in mind; since this idea represents the properties shared by all triangles, the proof can generate universal results (1690, p. 643). Finally, abstract ideas are necessary in order to explain how we sort objects into kinds.

That then which general words signify, is a sort of things; and each of them does that, by being a sign of an abstract idea in the mind, to which idea, as things existing are found to agree, so they come ranked under that name; or which is all one, be of that sort (Locke 1690, p. 414).

For example, once we have acquired the abstract idea of GOLD, we can consult this idea when classifying a new sample of metal; we simply compare its features with those of the abstract idea (e.g. yellow, malleable, soluble in *aqua regia*), and classify it as gold if we find a match.

Hume agrees with Locke that each of these capacities stand in need of explanation. He does not deny that we can communicate via general terms, nor that we can conduct geometrical proofs, nor that we can sort objects into kinds. He merely denies that we must posit abstract ideas in order to do so. His first point is that abstract ideas are not necessary to explain the semantics of general terms. Since there is a “one word-many idea” association in the imagination, we do not need to posit abstract ideas in order to explain how a general term can stand for multiple particulars. The meaning of a general term, then, does not derive from the particular exemplar in the mind of the speaker, but from the network of instances in the revival set (Zartmann 1975, pp. 257–258).

Nor do we need abstract ideas in order to explain how we carry out geometrical proofs. Hume recognizes that if we were to base our proofs about triangles on a single exemplar, we would be prone to false generalizations. According to Hume's theory of general ideas, however, the general term "triangle" not only calls to mind a particular exemplar, but it also puts us in a state of "readiness" to consider the other instances in the revival set (1739, pp. 20–21).

[A]fter the mind has produc'd an individual idea, upon which we reason, the attendant custom, reviv'd by the general or abstract term, readily suggests any other individual, if by chance we form any reasoning, that agrees not with it. Thus shou'd we mention the word, triangle, and form the idea of a particular equilateral one to correspond to it, and shou'd we afterwards assert, *that the three angles of a triangle are equal to each other*, the other individuals of a scalenum and isocetes, which we overlook'd at first, immediately crowd in upon us, and make us perceive the falsehood of this proposition, tho' it be true with relation to that idea, which we had form'd (1739, p. 21).

John Tienson describes the revival set as a "counterexample machine" (Tienson 1984, p. 323). We can carry out our proofs with particular figures in mind because the revival set will protect us against false generalizations.

Finally, Hume puts his theory of general ideas to work in order to explain classification. Classification can be viewed as the mirror-image of communication: in the case of communication, we perceive a general term and the task is to revive appropriate instances of the class; in the case of classification, we perceive an instance of the class and the task is to revive the appropriate general term. Thus, Hume simply reverses his account of communication in order to explain classification: since there is a "many-to-one" associative mapping of instances to words, the presentation of a previously observed instance will automatically call to mind its associated term. This account handles previously unobserved instances as well: novel items can be sorted into groups according to the degree of resemblance they bear to instances in the revival set (1739, p. 23).

Hume has shown that the exemplar account can explain our capacity to speak and think in general terms. It is not necessary, therefore, to posit abstract ideas. But Hume has failed to eliminate the possibility of abstraction, and thus he cannot conclude that his account is necessary, either. The problem is that both sides of the controversy provide sufficient explanations of the phenomena. For example, suppose that someone utters the general term "man" and we understand what the person means; this might be because we have in mind an abstract idea of man, or it might be because we have in mind an exemplary man and the appropriate revival set. Suppose further that someone shows us a shiny piece of metal and we classify it as "gold"; this might be because we have compared the sample with our abstract idea of gold, or it might be because we have compared it with particular exemplars of gold. The empirical

evidence, in other words, underdetermines the rival theories about the nature of general ideas.

The early modern controversy over abstract ideas, therefore, ends in a stalemate. Neither side can refute the other with *a priori* arguments: just as proponents of abstraction cannot demonstrate the absurdity of the exemplar hypothesis, so Hume cannot prove the impossibility of abstract ideas. But as we have seen, the controversy cannot be settled by appeal to *a posteriori* considerations either, since both accounts provide adequate explanations of our capacity to speak and think in general terms.

The contemporary controversy over abstract ideas in cognitive science

Can our contemporary science of the mind get us out of this stalemate? Cognitive scientists have been investigating the representational format of concepts and categories for well over forty years; surely there is hope that progress has been made on this issue. I will argue in the final section of this paper, however, that the contemporary debate over abstract ideas *recapitulates* the early modern controversy, and as a result, the question of how general ideas are represented in the mind remains open to this day.

In a landmark 1968 paper entitled “On the Genesis of Abstract Ideas”, Michael Posner and Steven Keele argue that it is necessary to posit abstraction in order to explain the behavior of subjects in classification task experiments. Posner and Keele train the subjects in these experiments to sort a series of random dot patterns into two classes.⁴ Once the subjects learn to categorize the patterns appropriately, Posner and Keele present them with the prototypical patterns of each class (which they had not seen), and then measure their reaction times. The significant result is that subjects manage to classify the prototypes as efficiently as the patterns they had seen in the learning phase. Posner and Keele conclude that the subjects not only stored the particular instances on which they were trained, but they also acquired an abstract representation of the “commonalities” of each class (Posner and Keele 1968, p. 354). We must postulate a process of abstraction, they argue, in order to explain how subjects were able to classify the unseen prototype with such ease.

In the late 1970s and early 1980s, however, this conclusion was challenged by contemporary exemplar theorists in cognitive psychology. Exemplar theorists agree with Hume that classes are represented in memory as separately stored instances, each of which retains its idiosyncratic properties (Smith and Medin 1981, p. 145). This means, of course, that there will be a vast number of instances stored in memory. We nevertheless manage to access this information, according to contemporary exemplar theorists, because search and

retrieval takes place in a massively parallel fashion. In Ratcliffe's model of associative memory, for example, an exemplar "resonates" with similar instances in memory in much the same way that "a ringing tuning fork evokes sympathetic vibrations in other tuning forks" (Ratcliffe 1978, p. 59; cf. Hintzman 1986, p. 426).

Exemplar theorists argue that it is unnecessary to posit abstraction in order to account for the Posner and Keele results, since their model also predicts that reaction time to the unseen prototypes would be at least as good as performance on the previously observed patterns.

Performance on prototype patterns frequently equals or exceeds performance on old training patterns. This result can be predicted from the context [i.e. exemplar] model because the prototypic pattern always is the pattern having the greatest number of highly similar category exemplars or training patterns (Medin and Schaffer 1978, p. 214).

According to exemplar theory, novel patterns are categorized into the class of exemplars to which they have the highest summed similarity. When Posner and Keele construct the training patterns in their experiment by distorting the prototype according to a fixed statistical rule, they guarantee that the prototype will be the member of the class with the highest number of similar patterns, and thus that it will be the easiest pattern to categorize.

The contemporary debate over abstract ideas in cognitive psychology ends up in the same stalemate as its early modern predecessor. On one side are the prototype theorists, who (like Locke) claim that our capacity for general language and thought can only be explained by abstract ideas; on the other side are the exemplar theorists, who (like Hume) provide an alternative explanation of these capacities in terms of particular instances. The root of the problem, as Larry Barsalou points out, is that both theories make very similar, if not identical, predictions (Barsalou 1990, p. 72; Barsalou 1992, p. 31). Prototype and exemplar theorists agree that in order to categorize a novel pattern, we must perform a summary computation of its similarity to the various classes in memory. They merely disagree over whether this computation is based on its *similarity to the averages* of each class (i.e. the prototype), or its *average similarity* to the exemplars (Reed 1972, p. 387). This is a very subtle difference, and as a result, the two theories are extremely difficult to distinguish at the behavioral level.

One might be tempted to think that the source of the problem lies with the methods of cognitive psychology. Cognitive psychologists attempt to *infer* the representational format of general ideas from behavioral evidence; but in order to settle the debate over abstract ideas, we must *directly observe* the representations that are formed during the classification experiments. In order to accomplish this more difficult task, the objection continues, we must turn to computer modeling. One of the benefits of computer simulation over

human experimentation is that it allows researchers to “directly inspect” the representations that enable the computer to solve the task (Elman et al. 1996, p. 45).

Over the last 20 years, there have in fact been several simulations of the Posner and Keele experiments. Knapp and Anderson (1984), for example, train a connectionist network with distributed memory to classify random dot patterns, and they demonstrate that when the number of training instances is large, the network will develop an abstract representation of the patterns (Knapp and Anderson 1984, p. 616). These results are not conclusive, however, because subsequent connectionist simulations have shown that networks can solve the Posner and Keele classification tasks by relying solely on stored exemplars (Kruschke 1992).

The problem with connectionist simulations at this point in time is that they are too unconstrained to yield a principled resolution of the controversy over abstract ideas. Whether or not the network abstracts information from the training data will depend upon a choice of the modeler. On the one hand, if one constructs a network with a large number of hidden units relative to that of the input patterns, the network will serve as a look-up table and store the learned patterns separately; on the other hand, if one builds a network with a “narrow waist” – that is, a hidden unit space that is proportionally smaller than the input space – then the network will develop a compact representation of the input patterns (Hampton 1997, p. 97).

Conclusion

This is not to say that no progress has been made. The recent debate in cognitive science sheds a good deal of light on the early modern controversy over abstract ideas. First, we can now see that Hume and his opponents were bound to fail in their attempts to refute each other with *a priori* arguments. There is no logical inconsistency in the notion of an abstract idea, nor is there anything demonstrably absurd about the exemplar account. The controversy cannot be settled, therefore, from the philosophical armchair.

Second, it is now clear that the early modern philosophers were on the right track when they regarded the question of how concepts and categories are represented in the mind as an empirical matter, one that must be investigated by putting their theories to the evidential test. It seems entirely proper, then, that the debate over abstract ideas has migrated over the years from philosophy to cognitive science, where it can be pursued with the best available experimental techniques.

Finally, the recent controversy allows us to rethink our evaluation of Hume’s theory of general ideas. There is a consensus among historians that Hume’s account is a muddle. Even a sympathetic critic like Passmore, for example, refers

to the “notorious inadequacy of Hume’s theory of general ideas” (Passmore 1952, p. 40). This criticism, however, now seems unfair: a version of Hume’s associationist-exemplar account has reemerged as one of the most prominent positions in our contemporary science of the mind.

Notes

1. Hume praises Berkeley’s exemplar theory of general ideas as “one of the greatest and most valuable discoveries that has been made of late years in the republic of letters” (1739, p. 17). By the end of the *Treatise* 1.1.7, though, he has made so many alterations to Berkeley’s proposal that the version he defends ought to be considered his own invention (Aaron 1941, p. 126).
2. The term ‘revival set’ was coined by Don Garrett (1997, p. 24).
3. It is difficult for Hume to express, in his early modern vocabulary, our awareness of the items in the revival set. In the *Treatise*, he maintains that the general term “touches the soul, if I may be allow’d so to speak, and revives the custom which we have acquir’d by surveying them” (1739, p. 20). In contemporary psychology, this point could be captured in terms of the notion of “associative priming” (Ratcliffe and McKoon 1981). The revival set, along similar lines, would be explained in terms of “spreading activation” along the pathways that connect the instances of a category (see Loftus 1974).
4. Posner and Keele construct the training patterns for their experiment by taking three random dot patterns and declaring them to be prototypes; they then use a statistical distortion rule to generate the patterns for their experiment.

References

- Aaron, R. I. 1941. Hume’s theory of universals. *Aristotelian Society Proceedings* 42: 117–140.
- Barsalou, L. W. 1990. On the indistinguishability of exemplar memory and abstraction in category representation. In: T. K. Srull and R. S. Wyer (eds), *Advances in Social Cognition: Content and Process Specificity in the Effects of Prior Experiences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Barsalou, L. W. 1992. *Cognitive Psychology: An Overview for Cognitive Science*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bennett, J. 1971. *Locke, Berkeley, Hume: Central Themes*. Oxford: Clarendon Press.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Garrett, D. 1997. *Cognition and Commitment in Hume’s Philosophy*. Oxford: Oxford University Press.
- Hampton, J. 1997. Psychological representation of concepts. In: M. Conway (ed), *Cognitive Models of Memory*. Cambridge, MA: MIT Press.
- Hintzman, D. L. 1986. “Schema abstraction” in a multiple-trace memory. *Psychological Review* 93: 411–428.
- Hume, D. 1748/1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press.
- Hume, D. 1739/1978. *A Treatise of Human Nature*. Oxford: Oxford University Press.

- Jolley, N. 1999. *Locke: His Philosophical Thought*. Oxford: Oxford University Press.
- Knapp, A. and Anderson, J. 1984. Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 4: 616–637.
- Kruschke, J. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99: 22–44.
- Locke, J. 1690/1975. *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.
- Loftus, E. F. 1974. Activation of semantic memory. *American Journal of Psychology* 86: 331–337.
- Medin, D. L. and Schaffer, M. M. 1978. Context theory of classification learning. *Journal of Experimental Psychology: Human Learning and Memory* 7: 241–253.
- Passmore, J. A. 1952. *Hume's Intentions*. Cambridge, England: Cambridge University Press.
- Posner, M. I. and Keele, S. 1968. On the genesis of abstract ideas. *Journal of Experimental Psychology* 77: 353–363.
- Price, H. H. 1953. *Thinking and Experience*. London: Hutchinson University Library.
- Ratcliff, R. 1978. A theory of memory retrieval. *Psychological Review* 85: 59–101.
- Ratcliff, R. and McKoon, G. 1981. Does activation really spread? *Psychological Review* 88: 454–462.
- Reed, S. K. 1972. Pattern recognition and categorization. *Cognitive Psychology* 3: 382–407.
- Smith, E. E. and Medin, D. L. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.
- Tienson, J. 1984. Hume on universals and general terms. *Nous* 18: 311–330.
- Winkler, K. 1989. *Berkeley: An Interpretation*. Oxford: Oxford University Press.
- Zartman, J. 1975. Hume and “the meaning of a word”. *Philosophy and Phenomenological Research* 36: 255–260.