Further Comments on Ahn et al.

Michael Strevens

In what follows I supplement my published reply (Strevens 2001) to Ahn et al.'s commentary (Ahn et al. 2001) on my paper concerning psychological essentialism (Strevens 2000).

The supplementary comments come in three parts. First, I ask whether my alternative to psychological essentialism, which I call the *minimal hypothesis*, or *minimalism* for short, can explain certain experimental results of Ahn and her coworkers. Second, I comment on Ahn et al.'s handling of my charge that Barbara Malt's work on the concept of water militates against psychological essentialism. Third, I remark on a deep theoretical difference between psychological essentialism and my minimalist approach to the psychology of natural kind terms.

1. Explaining Ahn's Experiments

In this section I take up a challenge thrown out by Ahn et al. as an objection to minimalism, a challenge to give a minimalist explanation of certain experimental data obtained by Ahn and her coworkers.

Ahn has produced similar results in studies concerning a range of different category types; to provide as intense a challenge to minimalism as possible, let me consider a particularly nice experiment involving biological kinds. Ahn (1998) told her subjects of a biological kind K that had a certain property A, which in turn caused another property B. One of the scenarios, for example, concerned the orchid genus *Coryanthes*, which secretes a substance, eucalyptol, that helps to attract a certain kind of bee. Here K is the orchid genus *Coryanthes*, A is the presence of eucalyptol, and B is the fact of bee visitation. The causal structure, then, may be represented as follows: $K \rightarrow A \rightarrow B$.

After hearing the facts about the kind K, subjects were told of two specimens:

1. A specimen that has A but lacks B, and

2. A specimen that has B but lacks A.

Subjects were then asked how likely the specimens were to belong to the biological kind K. They were far more likely to classify specimen (1) as a member of K than specimen (2). Thus, for example, they were more likely to say that an orchid that secretes eucalyptol but attracts no bees is a member of *Coryanthes*, than to say the same for a specimen that does not secrete eucalyptol but that, for some other reason, nevertheless attracts bees.

In a clever control, some of the subjects heard the story the other way round, so that the orchid attracts bees, which in turn causes the presence of the eucalyptol, brought in by the bees' visits. Subjects who heard this second kind of story were more likely to classify an orchid that was visited by bees but lacked

eucalyptol as a member of *Coryanthes*. Thus the identity of the properties *A* and *B* seems not to matter to the categorization; what matters is the relative position of *A* and *B* in the causal hierarchy. The property that is causally closer to the kind *K* is regarded as a more reliable indicator of *K* membership.

Ahn interprets these results in the context of a weights-and-features model of categorization, reminiscent in some ways of prototype theory, reaching the conclusion that those features are weighted more heavily that are causally more central, that is, that are closer to *K*. Ahn et al. challenge minimalism to explain this weighting scheme. Now, minimalism does not make use of a weights-and-features model, so it will not and should not explain Ahn's weighting scheme. But it does provide a very good explanation of what really matters, the data about categorization decisions.

The story is very simple. If you have a causal law of the form "*K* causes *A*, which in turn causes *B*", and you wish to use that law to make an inference back up the causal chain from *B* or *A* to *K*, you must have reason to believe that *B* or *A* is present because of the causal law in question. Or at the very least, you had better not have a reason to believe that the presence of *A* or *B* was *not* brought about by the law. This, according to Strevens (2000), is why older children and adults do not classify a painted raccoon as a skunk: they know that skunkhood causes skunk appearances, but they also know that this particular specimen's skunk appearances were caused by painting, and hence not in accordance with the skunkhood law. Thus they know not to use the law to infer skunkhood. A parallel case involving the orchid would be as follows: if

an orchid is known to have eucalyptol but it is also known that the eucalyptol was placed there by human experimenters, and therefore was not secreted by the flower, the eucalyptol will not be counted as evidence for membership of *Coryanthes*.

Now consider Ahn's two specimens. In the first, eucalyptol is present but bee visits do not occur. Can subjects use the causal law that *Coryanthes* secretes eucalyptol to make a categorization? They can, since absence of bee visits provides no reason to think that the eucalyptol was not produced in accordance with the law. In the second story, bee visits are present but eucalyptol is absent. Can the subjects use the causal law that *Coryanthes* secretes eucalyptol, which in turn attracts bees, to make a categorization? They cannot, since the absence of eucalyptol shows that this particular law is *not* what explains the bee visits. Because eucalyptol is absent, something else must be causing the bee visits, and Ahn's subjects have no reason to think that this something else is connected with membership of *Coryanthes*.

In short, Ahn's experiments are evidence of subjects' reasoning about categorization in a typically causal way, just as described in Strevens (2000). It is an open question whether psychological essentialism explains her results so easily.

2. Water and H_2O

In Strevens (2000), section 5.3, I argued that experiments carried out by Barbara Malt (1994) show that people's belief that water is H_2O plays no role in their de-

cision to categorize substances as being water or non-water. Because, it would seem, if people have any essentialist belief, it is the belief that H_2O is the essence of water, this result militates strongly against psychological essentialism, remembering that psychological essentialism holds not only that people have beliefs about essences, but also that these beliefs play a central role in categorization and other inferences about natural kinds.

Ahn et al. reply to this argument as follows. They concede that the Malt experiments provide reason to think, as I have asserted, that the belief that water is H_2O is not important to people's water categorizations, but they suggest that we have, in fact, two concepts of water, the everyday concept and a concept of *pure water*. They argue that our H_2O-involving essentialist belief is that the essence of *pure water* is H_2O; it is no refutation of essentialism, then, to show that this belief plays no role in our decisions as to whether or not various substances belong to the category of everyday water.

But what, then, do people believe about the essence of everyday water? Ahn et al. seem to have two possible options in replying to this question, neither very appealing given their views. First, they can claim that people do not believe that everyday water has an essence. This does not seem to be a strategic move for defenders of essentialism. Second, they can claim that people do believe that everyday water has an essence, but that people think that this essence has nothing to do with H_2O. Good luck to them, but I doubt that this view will survive more than a few moments of empirical scrutiny.

3. The Myth of Depth

I will finish with a few remarks on what I have come to see as perhaps the greatest theoretical difference between minimalism and essentialism; indeed, between minimalism and just about every other theory of categorization that psychology has taken seriously.

Call a belief that members of a category K are members of K because they possess certain properties a *category rationalizing belief*, or a *categorizing belief* for short. A definition of K is a categorizing belief, but so is a belief that, say, objects are members of K because they share a hidden essence. What distinguishes a categorizing belief from other beliefs about a kind is that it is a belief not just about the members of a kind, but about the nature of the kind itself.

There are two sorts of theories of categorization: those according to which categorization is driven by a categorizing belief, and the rest. I call these respectively *deep* and *shallow* theories of categorization.[1] When I say that categorization is driven by a categorizing belief, I mean that categorization normally happens in one of two ways: either a categorizing belief is directly consulted to determine whether or not an object belongs to a kind, or a categorizing belief is used to construct an everyday heuristic that is consulted to determine kind

1. Note that this distinction has nothing to do with the role of observable versus unobservable properties in categorization. A classical theory that defines all natural kinds in terms of their characteristic appearances, such as that advanced by John Locke in his *Essay on Human Understanding*, is a deep theory. What is deep, on a deep theory, is the categorizer's understanding of the nature of their categories.

membership. In theories that propose the second sort of categorization, the categorizing belief may not be consulted on every categorization, but it is the final authority on category membership. Deep theories, that is, theories in which categorization is driven by a categorizing belief, include the classical theory, the kind of augmented classical theory proposed by Rey (1983), and, I think, the essentialist theory.

It is not clear to me whether prototype theories of concepts are deep. They are, I think, if all the apparatus of prototype-driven categorization—the feature sets, the weightings, and the rules relating the feature sets and weightings to category membership—are explicitly represented by the categorizer. When this is so, the categorizer will have explicit beliefs about what determines category membership; these are categorizing beliefs. It must be admitted, however, that much of the literature on prototype theory has a tenor which seems at odds with deep theories of categorization.

What makes minimalism special, I think, is that it is self-consciously shallow: it explicitly denies that categorization is driven by a categorizing belief. At first, this may sound a very unpromising strategy. How can humans put objects into categories unless they have some belief as to why particular objects belong in particular categories?

The key to answering this question is to observe that, in order to put an object in a category *K*, a categorizing belief is not always necessary. All that is necessary is some belief or other that involves *K*. This belief may say very little about *K*, and it may say nothing at all about what kinds of things *K*s are. When

I am hiking, for example, and my trial guide says that I am passing through a grove of California-nutmegs, my belief that I am currently passing through a grove of California-nutmegs is sufficient for me to categorize the trees around me as California-nutmegs, even though the belief says absolutely nothing about what kind of trees California-nutmegs are.

Now the hiking story is possible, of course, only because somebody, somewhere, knows a lot more about California-nutmegs than I do. But now suppose that my trail guide contains far more information about California-nutmegs, specifically, all the K-laws that, according to minimalism, a California-nutmeg expert would believe. As a result of reading the guide, I now believe a number of causal laws about California-nutmegs: that California-nutmegs have short spiky needles, that they grow in the Sierra Nevada foothills, and so on. If the guide is sufficiently comprehensive, I am now as good at categorizing California-nutmegs as any of my more arboreally savvy ancestors. Yet I still have no categorizing beliefs about California-nutmegs. I have many beliefs about the causal consequences of being a California-nutmeg, but provided that I refrain from metaphysical speculation, I have no beliefs about what makes a tree a California-nutmeg.

This, I suggest, is the situation in which we are all in. Nature has been sure to provide us with a theory structure to encode the knowledge we need to recognize a California-nutmeg, a skunk, an apple tree, or whatever, when we encounter one, but has not given us a metaphysical belief system to go along with our categorization ability. Minimalism fulfills a valuable function, I think,

by showing how this is possible, how it is possible, that is, for a species such as ours to have rich and powerful categorization abilities that are not driven by categorizing beliefs.

Two comments on this, and then I will finish. First, it is tempting to turn minimalism into a deep theory, that is, a theory in which the beliefs that play a central role in categorization are categorizing beliefs. This can be done by taking the K-laws as constitutive of concepts, so that the K-laws about tigers, for example, in effect define what it is to be a tiger: a tiger is any organism for which the tiger K-laws are true. Ahn et al. attribute this view to me when they write that according to minimalism, "if 'tigers have stripes' is a K-law, then tigers have stripes essentially" (p. 2). But this is not what I intend at all. We do not, I submit, represent any of the K-laws for a kind as being essential for the membership of that kind; indeed, we do not consider *any* of the facts about kinds deployed in everyday categorization as being essential for kind membership.

Second comment: although categorizing beliefs play no role, according to minimalism, in everyday categorization, it must be admitted that people do, at various times, form categorizing beliefs about kinds, for example, beliefs about essences. Typical of these are the belief that tigers are just any animals with a certain kind of DNA, and that water is just H_2O. Minimalism is committed to the view that beliefs such as these play no role in everyday categorization. This commitment provides considerable scope for empirical tests of minimalism against, in particular, essentialism. An example of such test is provided by the Malt data on water and H_2O, mentioned in section 2, and described in greater

detail in Strevens (2000), section 5.3.

It is a well known phenomenon that in many fields, it has proved enormously difficult to find a satisfactory account of what makes a thing a member of a particular kind. In biology, in particular, there is very little consensus about what makes an organism a member of a given species. If our categorization is driven by categorizing beliefs, it is rather hard to see why this should be so. Supposedly, our categorization of things as members of a given kind is based on a pre-existing opinion, however fragmentary, as to the nature of members of that kind. Thus the quality of our categorization skills ought not to exceed the quality of these opinions. Yet we find categorization easy, but we find reaching opinions about the nature of kinds very difficult. How can this be? Simple: such opinions play no role in categorization.

References

Ahn, W. (1998). Why are different features central for natural kinds and artifacts?: The role of causal status in determining feature centrality. *Cognition* 69:135–178.

Ahn, W., C. Kalish, S. A. Gelman, D. L. Medin, C. Luhmann, S. Atran, J. D. Coley, and P. Shafto. (2001). Why essences are essential in the psychology of concepts: Commentary on Strevens (2000). *Cognition* 00:00.

Malt, B. (1994). Water is not H_2O. *Cognitive Psychology* 27:41–70.

Rey, G. (1983). Concepts and stereotypes. *Cognition* 15:237–262.

Strevens, M. (2000). The essentialist aspect of naive theories. *Cognition* 74:149–175.

———. (2001). Only causation matters: Reply to Ahn et al. In press, *Cognition*.