

**Comments on Alex Barber,  
"Concepts, Conceptions, and Cognitive Change"**

Presented at the 2001 Congress of the Canadian Philosophical Association, May 2001.

**Sam Scott ([sscott@ccs.carleton.ca](mailto:sscott@ccs.carleton.ca)), Carleton University.**

## **0. Summary**

The purpose of Barber's paper is to suggest that developmental psychologists who are committed to the popular *theory-theory* of conceptual content should adopt a new variant known as *entrenchment* theory, that will be immune to some of the philosophical challenges that plague the theory-theory.

The theory-theory of concepts says that a mental representation expresses some concept when: a) that mental representation occurs in a well-defined set of stored propositions (beliefs), and b) this set of stored propositions are, as Barber puts it, "characteristic" of the concept. (p. 2) The unrestricted conception-conception says that every single belief a person has that involves some representation is relevant in determining which concept that representation expresses. So the theory-theory is a restriction of the conception-conception because it identifies a subset of core beliefs involving a representation R, namely the theoretical ones (whatever that may mean), and claims that only these theoretical beliefs, and not any others, determine the identity of the concept expressed by R.

Entrenchment theory also places a restriction on which beliefs are relevant to the content of a representation, but it does so in a way that is different in two respects. First, and most importantly, rather than determining whether a belief is part of a theory, we use

the level of entrenchment of the representation in the belief to determine whether it is central or not. Second, while the theory-theory encourages us to make a binary distinction between beliefs that do or do not bear on the identity of a concept, entrenchment (I believe) necessarily makes a fuzzy distinction. Those beliefs in which a representation R is more entrenched have a stronger bearing on the identity of the concept expressed by R. Those beliefs in which R is less entrenched have a weaker bearing.

R is more entrenched in belief A than belief B if giving up or revising belief A would be more *epistemically traumatic* than giving up or revising belief B. Giving up a belief will be more epistemically traumatic the more far reaching the implications for evidence are. Barber remarks that, at least for natural kind concepts, demonstrative and linguistic beliefs will tend to be more entrenched than beliefs about theoretical origin or structure. This pattern may be reversed for more purely theoretical concepts. (p. 14)

## **1. Genuine Disagreement**

According to the theory-theory, two people can have the exact same concept if they share the same theory involving it – that is, if they believe the same set of core theoretical propositions. As long as they have this basis of agreement, they can disagree over their non-central beliefs and those disagreements will be *genuine*. If they disagree in their core beliefs then the disagreement is only an *apparent* disagreement because they are actually talking about different concepts. It might seem that any realistic theory must involve so many core beliefs that genuine disagreement could almost never happen, but at least we have a well-defined way of making the genuine/apparent distinction.

The situation is not so simple for entrenchment theory. It seems to me that for two people to have *exactly the same* concept, and hence have something to genuinely disagree

about, they must (paradoxically) agree on *all* their beliefs involving the concept, not just a subset (that is, unless there are some beliefs that involve the concept but in which the concept is not at all entrenched, but this possibility seems contrary to the idea of entrenchment.) Furthermore, their concepts must be entrenched to the same degree in each belief otherwise the introduction of a new belief could have different consequences for the two people's concepts. For instance, suppose Janet and Michael both believe that snow is the material residue of fairy breath, but giving up this belief would be much more epistemically traumatic for Janet than for Michael. Maybe Janet believes she has had first-hand experience with (what appear to be) fairies breathing (what appears to be) snow, while Michael has only heard about this phenomenon from Janet. Then when faced with some contradictory evidence (e.g. that fairies do not exist), Michael will be more willing than Janet to accept this as a new belief and give up his old belief that snow is fairy breath residue. Janet's similar belief was more central to her SNOW concept, hence more difficult to give up and more determinative of the content of her concept.

So if I am right about this, entrenchment makes it much less likely that two people can have exactly the same concept and impossible that they can genuinely disagree about it. More beliefs must match, and the degree of entrenchment of the concepts in each belief must match. Barber suggests that instead of worrying about when two concepts are exactly the same, we can get by with a notion of concept similarity that will be determined in a "partially holistic" manner. (p. 8) But I think we are owed an account of how this is to work in at least two respects. First, we need a metric of similarity that takes into account both the degree to which beliefs are shared and the relative levels of entrenchment of the concept in each of the shared (or non-shared) beliefs. Second, given

such a metric, we would need to know how it could be used to judge apparent vs. genuine disagreement. Almost all disagreements will now, strictly speaking, be non-genuine. But will disagreements between individuals with more similar concepts now be somehow *more genuine* than disagreements between individuals with less similar concepts?

## **2. Demonstrative Beliefs**

I am also concerned about the centrality of demonstrative beliefs to natural kind concepts under entrenchment theory. It seems to me that the formation of new demonstrative beliefs is part of what needs to be explained by a theory of concepts, so we need to be careful about using demonstrative beliefs as examples of beliefs that could fix the content of a concept. There are two questions that I think need to be considered: 1) Demonstrative beliefs are generally the result of applying a concept rather than a defining part of its content, so how do new demonstrative beliefs get formed when the other beliefs about the concept are wrong? and 2) In cases of accidental misrepresentation involving natural kind concepts, how do we know that it's the newly-formed demonstrative beliefs that need to be overturned?

1.) Take the example of Janet, who believes both that snow is hot and that that stuff (points) is snow. The content of her SNOW concept is supposed to come from the set of beliefs that involve this concept. So it is presumably these beliefs that are being used to categorize things in the world as snow or not snow. That is, it is her stored beliefs about snow, including the belief that snow is hot, that are being consulted to form the demonstrative belief. But how could Janet ever come to form this demonstrative belief about actual snow if the belief that snow is hot forms part of the content of her SNOW concept? It's possible that she got the demonstrative belief by being told by a trusted

source, in which case there is no problem. But a theory of concepts needs to be able to account for the possibility of a person realizing she is wrong without being told – something that happens all the time. So we have to figure out a way that Janet could have come by the demonstrative belief herself. One possibility is that she could have used only some of her beliefs involving snow (i.e. part of the content of her SNOW concept) to roughly categorize some stuff as snow, only later to discover that the stuff was cold. At this point, because it would be more epistemically traumatic to revise the demonstrative belief, she changes her mind about snow being hot.

2) But if that story about the formation of new demonstrative beliefs is true, it makes conceptual content far too unstable to handle cases of accidental misrepresentation, where it is appropriate to give up the demonstrative belief and retain the theoretical ones. For example, suppose I think I see a dog in the distance, and then hear it meow and realize it is a cat. Why was it that in this case my newly formed demonstrative belief was less entrenched than my other beliefs about dogs? Why don't I feel compelled to add the belief that dogs sometimes meow to my list of beliefs about dogs?

## **Conclusion**

I do think entrenchment is a promising alternative to theory-theory, but I have some concerns related to genuine disagreement and the role of demonstrative beliefs in the theory:

1. How can genuine disagreement be possible?
- 2a. How can new demonstrative beliefs be formed?
- 2b. How does accidental misrepresentation ever get noticed?