Topics in Cognitive Science 1 (2009) 231–236
Copyright © 2009 Cognitive Science Society, Inc. All rights reserved.
ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01015.x

# The Part of Cognitive Science That Is Philosophy

Daniel C. Dennett

*Center for Cognitive Studies, Tufts University*

Received 31 January 2009; received in revised form 11 February 2009; accepted 11 February 2009

**Abstract**

There is much good work for philosophers to do in cognitive science if they adopt the constructive attitude that prevails in science, work toward testable hypotheses, and take on the task of clarifying the relationship between the scientific concepts and the everyday concepts with which we conduct our moral lives.

*Keywords:* Philosophy in cognitive science; Conceptual confusion in cognitive science; Philosophy and hypothesis-generation; Qualia; Neural correlates of consciousness; Libet

I like Andrew Brook's distinction between philosophy *in* cognitive science and philosophy *of* cognitive science. In principle, you can do one without the other, and some work in the field can be quite clearly given just one of these labels, but in general I think that good philosophical contributions to cognitive science must work on both projects at once: it is a mistake to think, for instance, that philosophy *of* cognitive science is ''just for philosophers'' or ''just for the lay public'' a collection of antidotes to faulty inferences about the import of contemporary scientific work. Cognitive scientists themselves are often just as much in the grip of the sorts of misapprehensions and confusions as outsiders succumb to, and being down in the trenches sometimes makes them even more susceptible.

My parade case would be the large literature *by scientists* in the wake of Benjamin Libet's work on (what he thought was) the consciousness of intention. Here the scientists' ''lay'' intuitions have misled them into dubious public pronouncements about the import of the work (for quotes from Michael Gazzaniga, William Calvin, and V. S. Ramachandran, see my *Freedom Evolves*, Dennett [2003], pp. 230–31) and also distorted the analysis of some ingenious follow-up experiments (e.g., Haggard, 2008; Haggard & Eimer, 1999; Lau & Passingham, 2007; Soon, Brass, Heinze, & Haynes, 2008). All of these experiments rely on subjects making a most unnatural judgment of simultaneity the import of which is not carefully analyzed, because of the presumption that the right question to ask is, *When does*

Correspondence should be sent to Daniel C. Dennett, Center for Cognitive Studies, Tufts University, Medford, MA 02155. E-mail: daniel.dennett@tufts.edu

*the subject become aware of the intention to act?* This leads inexorably to the ''discovery'' that subjects are shockingly tardy in getting the news about what their bodies are about to do. This interpretation tacitly presupposes that until a subject *can say* that he or she has a particular intention, the subject is in no position to endorse, evaluate, review, or modify the intention in question. This creates the illusion of an ominous temporal bottleneck, with the Conscious Agent impatiently waiting (in the Cartesian Theater) for news from the rest of the brain about what projects are underway. I must add that the literature on the topic by philosophers includes some that is equally ill considered. However, since better work is on the way, there is no need to dwell on past confusions. Here the two varieties distinguished by Brook strike me as merging into a single enterprise: there are philosophical problems with the cognitive science itself, not just philosophical problems about how to understand or interpret cognitive science.

There is no dearth of reasons why philosophers are regarded askance (at best) by many in the scientific community, and they are familiar enough so that I will just acknowledge them in passing. Philosophers often comically misjudge their competence to evaluate concepts, arguments, theories with which they have only the most passing acquaintance, and they are prone to coming up with what might be called bumblebee deductions—proving from ''first principles'' that bumblebees cannot possibly fly. I do not exempt myself from this verdict. For instance, I used to claim that any theory that had to postulate ''grandmother neurons'' was dead in the water, but that is because I did not understand the possibilities of ''coarse coding'' and redundancy, which do indeed make it possible to rehabilitate some models that have elements that look an awful lot like grandmother neurons (e.g., Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). I once dismissed any theory that ''replaced the little man in the brain with a committee'' as conceptually bankrupt—until I realized that this was indeed a path, perhaps the royal road, to getting rid of the little man altogether. So live by the sword, die by the sword. If philosophers are going to make arguments with conclusions about fundamental constraints, possibilities, and impossibilities of theories or models of mind, they had better be ready to be shown they have been tricked by their own oversimplifications.

Still, unrepentant, I think we philosophers *can* make just such contributions; at the very least we can sharpen and clarify the alternatives, rendering the suspect claims clear enough to be refuted—which is definitely progress. Wolfgang Pauli famously put down another physicist's model as ''not even wrong''—it was too murky to be a candidate for truth or falsehood—and philosophers can sometimes help bring ideas into focus that otherwise would not be refutable or confirmable but that can still influence the direction of research by providing a sort of prevailing fog that deters people from entering certain regions of likely theory. One of the reasons cognitive science is such a land of plenty for philosophers is that so many of its questions—not just the grand bird's-eye view questions but quite proximal, in-the-lab-now questions—are still ill thought out, prematurely precipitated into forms that deserve critical reevaluation. If philosophy is, as my bumper sticker slogan has it, what you're doing until you figure out just what questions to ask, then there is a lot of philosophy to be done by cognitive scientists these days.

That raises the question: Can't they do it themselves better than carpet-bagging philosophers can? Not necessarily. Many of them certainly think so, but I relish the cases where

once they try it, they come to appreciate just how tricky the issues are and come asking for help. For instance, one of the side benefits for philosophers of the flood of books on consciousness by scientists in the last 15 years or so is the number of them whose authors have tied themselves in rather embarrassing knots (an example to which we will return).

One of the prime lessons I have learned from scientists is that they have a justified impatience, even contempt, for the sorts of purely destructive criticism, the piling on and hooting, that philosophers often try to import from the home discipline. Philosophers can sit in the trees sniping away merrily—but that is not constructive. The background assumption of philosophers always ought to be: if a problem is worth our attention, it is because really smart people find it difficult; we should humbly try to help them sort it all out, asking, not telling, being tentative, not preemptive, in our criticisms. I cringe when I see young philosophers doing a smarty-pants demolition number in front of scientists, a talk that would go down like honey in a room full of philosophers but merely makes the scientists shake their heads in dismay.

If philosophers aren't to do the sort of refutation by counter-example and reductio ad absurdum that their training prepares them for, what is the alternative? To take on the burden of working out an implication or two that is actually testable, if not today, in the near future. I have found that if you can help scientists design experiments, they take you more seriously than they otherwise would do. Conversations with David Premack at a meeting in Minnesota back in 1975 led to his experiments with the chimpanzee Sarah on deception, and then my ''Beliefs about beliefs'' commentary on Premack and Woodruff in *Behavioral and Brain Sciences* in Dennett (1978) helped to spawn the experimental industry in false belief tasks. A few years later, at the Dahlem conference on animal intelligence in 1982, I proposed my ideas about the intentional stance to a room full of ethologists and primatologists and was politely asked if this scheme of mine would help them design experiments. I sat down at lunch with several of them—Robert Seyfarth, Carolyn Ristau, Peter Marler, Sue Savage-Rumbaugh, Hans Kummer, as best I recall—and in an hour or two we designed a variety of experiments that they subsequently went off and did, with good results. (See Dennett [1983] for an early progress report.) More recently, the ''Appendix for Scientists'' in Dennett's *Consciousness Explained* (1991) provided a variety of experimental predictions that have borne fruit: change blindness most prominently but also experiments on ''filling in'' and color phi. Soon there will be experimental results on the ''precognitive carrousel'' effect discussed in that book as well, now that noninvasive brain-scanning technologies permit short-latency tapping of motor cortex activity.[1] And, make no mistake, the experimental work that has resulted has shown how the phenomena in question are more complicated than I had supposed when I first proposed the experiments. That is where the real progress lies, of course, not in the straight confirmation or refutation of a relatively simple ''theory.''

Communication with scientists is sometimes made difficult because scientists overuse the principle of charity. They hear about *qualia*, for instance, and decide that they know what these philosophers' qualia are (and of course qualia exist—anybody who has ever had a toothache knows that!), and when they learn about the controversies about qualia, they think they can solve them, or at least point to their solutions, with a few observations drawn

from recent experimental work. In recent work on consciousness (to return to that topic), a moving example of this is the late Jeffrey Gray, who in 2004 posthumously published the book, *Consciousness: Creeping up on the Hard Problem,* which is full of valuable and important ideas but misses by a fairly wide margin the philosophers' problems it aspires to tackle. Gray had simply assumed—and why not?—that philosophers had something important in mind, something testable, a difference that could make a difference, when they talked about the Hard Problem, or the problem of absent qualia and functionalism. He refused to believe that the issues that engaged the philosophers were as footling as in fact they are. (I tried on several occasions to show him, but he just laughed: ''Oh no, you can't be serious, Dan!'') So he valiantly charged ahead, doing his experiments, demonstrating the role and nature of what *he* called qualia, and thus creeping up on what *he* thought the Hard Problem was. It has now been 5 years since his book was published, and I can locate just one philosophical reaction to it (Kriegel, 2007). Why have philosophers ignored the book? Because it manifestly *did not* address their hard problem and was not about the qualia they took seriously. As Kriegel says, putting it mildly, ''. . . scientists use the label 'Hard Problem' more loosely than philosophers'' (p. 97).

Another area of philosophical naiveté in the cognitive science of consciousness concerns the quest for the Neural Correlate of Consciousness (NCC). It has seemed obvious to quite a few scientists aspiring to solve the mystery of consciousness that there *has* to be an NCC, the necessary and sufficient conditions, characterized in terms of locatable neural activity, for conscious experiences. How indeed could there not be one, if materialism is true? Consider a parallel question: suppose evolutionary biologists set out to discover the Biological Correlate of Speciation. Suppose they identify a hundred thousand instances in the past where population divergence has occurred. Such divergence is, if not always necessary (sympatric speciation does seem to occur), at least conditionally necessary—a condition without which no speciation would have occurred in most of the instances studied. But probably only a handful of those hundred thousand divergences would have led, in the fullness of time, to speciation. In the other instances, the populations reunited before speciation could occur, or one population died out, leaving no descendants. Evolution is about the amplification of effects that almost never happen.

What, these biologists ask, does the subset that proves eventually to have been speciation events have in common? The answer may be: Nothing! There may be absolutely no distinguishing features of those divergences that eventuate in speciation and those that do not, since whether speciation occurs depends on the happenstances of hundreds or thousands of generations that come into existence after the divergence is complete. Similarly, it can be the case that whether a particular contentful event—a discrimination, a binding, a categorization at time *t*—was conscious can depend on how its descendant effects fared over seconds, not milliseconds (Dennett & Akins, 2008).

There need not be anything mysterian or anti-scientific about the observation that the quest for the NCC is probably a wild goose chase. If the processes that elevate contents to consciousness are like the processes that elevate lineage divergences into speciation events, we would be looking in the wrong place at the wrong time for the hallmarks of consciousness by homing in on the brain processes that occurred at the time of content fixation—at

the time, in other words, when the relevant content just began to make a difference in the subject's cognitive life.

Scientists will probably always be impatient with philosophers' scruples about language, and with the philosophers' insistence that paths be carefully built from the everyday terms (''belief,'' ''dream,'' ''image,'' etc.) to their neuroscientific counterparts and then back to the everyday terms, but philosophers should not cave in to scientists' hyperpragmatic eliminativism. (Besides, they do not live by what they say. They continue to help themselves uncritically to all the connotations and implications of the everyday terms until they come to grief on some quandary or paradox.) It is worth remembering that the main reason everybody—really, just about everybody—is fascinated with, and troubled by, work in cognitive science is that it so manifestly promises or threatens to introduce alien substitutes for the everyday terms *in which we conduct our moral lives*. Will we still have free will? Will we still be conscious, thinking agents who might be held responsible? Does suffering really exist? It is because we truly need good, philosophically sound, scientific answers to *these* questions and not to any substitutes, that philosophers have a very substantial job to do in the ongoing progress of cognitive science.

## Note

1. My account in *Consciousness Explained* (Dennett, 1991, pp. 167–168) of Gray Walter's reported experiments with chronically planted electrodes has been met with skepticism by several readers, and digging in the archives has so far failed to yield a written, let alone published, version of Gray Walter's talk to the Osler Society of Oxford that I attended in 1963 or 1964. It has been suggested by some that he was making it all up, to dazzle the students. Perhaps he was, but the claim he made was eminently plausible and the experiments he described can now be replicated without boring holes in subjects' skulls.

## References

Dennett, D. C. (1978). Beliefs about beliefs (commentary on Premack et al.). *Behavioral and Brain Sciences*, *1*, 568–570.

Dennett, D. C. (1983). Intentional systems in cognitive ethology: The 'Panglossian Paradigm' defended. *Behavioral and Brain Sciences*, *6*, 343–390.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little Brown.

Dennett, D. C. (2003). *Freedom evolves*. New York: Viking Penguin.

Dennett, D. C., & Akins, K. (2008). *The multiple drafts model*. Available at http://www.scholarpedia.org.

Gray, Jeffrey. (2004). *Consciousness: Creeping up on the hard problem*. Oxford, England: Oxford University Press.

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews. Neuroscience*, *9*, 934–946.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental brain research/ Experimentelle Hirnforschung/Expérimentation cérébrale*, *126*(1), 128–133.

Kriegel, Uriah. (2007). Gray matters: Functionalism, intentionalism, and the search for the NCC in Jeffrey Gray's work. *Journal of Consciousness Studies*, *14*, 96–116.

Lau, H. C., & Passingham, R. E. (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *The Journal of neuroscience: The official journal of the Society for Neuroscience*, *27*(21), 5805–5811.

Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102–1107.

Soon, C. S., Brass, M., Heinze, H., & Haynes, J. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543–545.