FINAL DRAFT
October 18, 2005
for APA Eastern Division Meeting, December 2005
[ca.8200 words]

Philosophy as Naive Anthropology:
Comment on Bennett and Hacker

Bennett and Hacker's *Philosophical Foundations of Neuroscience* (Blackwell, 2003), a collaboration between a philosopher (Hacker) and a neuroscientist (Bennett), is an ambitious attempt to reformulate the research agenda of cognitive neuroscience by demonstrating that cognitive scientists and other theorists, myself among them, have been bewitching each other by misusing language in a systematically "incoherent" and conceptually "confused" way. In both style and substance, the book harks back to Oxford in the early 1960's, when Ordinary Language Philosophy ruled, and Ryle and Wittgenstein were the authorities on the meanings of our everyday mentalistic or psychological terms. I myself am a product of that time and place (as is Searle, for that matter), and I find much to agree with in their goals and presuppositions, and before turning to my criticisms, which will be severe, I want to highlight what I think is exactly right in their approach–the oft-forgotten lessons of Ordinary Language Philosophy.

> Neuroscientific research, ...abuts the psychological, and clarity regarding the achievements of brain research presupposes clarity regarding the categories of ordinary psychological description–that is, the categories of sensation and perception, cognition and recollection, cogitation and imagination, emotion and volition. To the extent that neuroscientists fail to grasp the contour lines of the relevant categories, they run the risk not only of asking the wrong questions, but also of misinterpreting their own experimental results. (p115).

Just so.[1] When neuroscientists help themselves to the ordinary terms that compose the lore I have dubbed "folk psychology,"[2], they need to proceed with the utmost caution, since these

[1] My purpose in *Content and Consciousness*, in1969 was "to set out the conceptual background against which the whole story must be told, to determine the constraints within which any satisfactory theory must evolve. (p *ix*) . . . . [to develop] the notion of a distinct mode of discourse, the language of the mind, which we ordinarily use to describe and explain our mental experiences, and which can be related only indirectly to the mode of discourse in which science is formulated. (p*x*)"

[2]Although earlier theorists–e.g., Freud–spoke of *folk psychology* with a somewhat different meaning, I believe I was the first, in "Three Kinds of Intentional Psychology," (1978), to propose its use as the name for what Hacker and Bennett call "ordinary psychological description." They insist that this is not a *theory*, as do I.

terms have presuppositions of use that can subvert their purposes and turn otherwise promising empirical theories and models into thinly disguised nonsense. A philosopher–an expert on nuances of meaning that can beguile the theorist's imagination–is just the right sort of thinker to conduct this important exercise in conceptual hygiene.

I also agree with them (though I would not put it their way) that "The evidential grounds for the ascription of psychological attributes to others are not inductive, but rather criterial; the evidence is logically good evidence." (p82). This puts me on their side against, say, Fodor.[3]

So I agree wholeheartedly with the motivating assumption of their book. I also applaud some of their main themes of criticism, in particular their claim that there are unacknowledged Cartesian leftovers strewn everywhere in cognitive neuroscience and causing substantial mischief. They say, for instance:

> Contemporary neuroscientists by and large take colours, sounds, smells and tastes to be 'mental constructions created in the brain by sensory processing. They do not exist, as such, outside the brain," [quoting Kandel *et al.*, 1995] This again differs from Cartesianism only in replacing the mind by the brain." (p113)

Here they are criticizing an instance of what I have called "Cartesian materialism," (*Consciousness Explained*, 1991), and they are right, in my opinion, to see many cognitive neuroscientists as bedazzled by the idea of a place in the brain (which I call the Cartesian Theater) where an inner show of remarkable constructions is put on parade for a (material) *res cogitans* sitting in the audience.

More particularly, I think they are right to find crippling Cartesianism in Benjamin Libet's view of intentional action, and in some of the theoretical work by Stephen Kosslyn on mental imagery. I also join them in deploring the philosopher's "technical" term, *qualia*, a poisoned gift to neuroscience if ever there was one, and I share some of their misgivings about the notorious "what is it like" idiom first explored by Brian Farrell (1950) and made famous by Thomas Nagel (1974). Introspection, they say, is not a form of inner vision; there is no mind's eye. I agree. And when you have a pain, it isn't like having a penny; the pain *isn't a thing* that is in there. Indeed. Although I don't agree with everything they say along the paths by which they arrive at all these destinations, I do agree with their conclusions. Or more accurately, they agree with my conclusions, though they do not mention them[4]

[3] See my discussion of this in "A Cure for the Common Code" in *Brainstorms*, and more recently in "Intentional Laws and Computational Psychology," (section 5 of "Back from the Drawing Board,") in Dahlbom, ed., *Dennett and his Critics*, 1993.

[4] The list is long. See, in addition to the work cited in the earlier footnotes, my critiques of work on imagery, qualia, introspection, and pain in *Brainstorms* (1978). I am not the only theorist whose work anticipatory to their own is overlooked by them. For instance, in their

More surprising to me than their failure to acknowledge these fairly substantial points of agreement is the fact that the core of their book, which is also the core of their quite remarkably insulting attack on me[5], is a point that I myself initiated and made quite a big deal of back in 1969. Here is what they call the mereological fallacy:

> We know what it is for human beings to experience things, to see things, to know, or believe things, to make decisions, to interpret equivocal data, to guess and to form hypotheses. But do we know what it is for a *brain* to see or hear, for a *brain* to have experiences, to know or believe something? Do we have any conception of what it would be for a *brain* to make a decision?

They answer with a ringing NO!

discussion of mental imagery, they reinvent a variety of Zenon Pylyshyn's points without realizing it. Bennett and Hacker are not the first conceptual analysts to frequent these waters, and most, if not quite all, of their points have been aired before and duly considered in literature they do not cite. I found nothing new in their book.

[5] Their Appendix devoted to attacking my views is one long sneer, a collection of silly misreadings, ending with the following: "If our arguments hold, then Dennett's theories of intentionality and of consciousness make no contribution to the philosophical clarification of intentionality or of consciousness. Nor do they provide guidelines for neuroscientific research or neuroscientific understanding." (p435) But there are no arguments, only declarations of "incoherence". At the APA meeting at which this essay was presented, Hacker responded with more of the same. It used to be, in the Oxford of the 60's, that a delicate shudder of incomprehension stood in for an argument. Those days have passed. My advice to Hacker: If you find these issues incomprehensible, try harder. You've hardly begun your education in cognitive science.

> It makes no sense to ascribe psychological predicates (or their negations) to the brain, save metaphorically or metonymically. The resultant combination of words does not say something that is false; rather, it says nothing at all, for it lacks sense. Psychological predicates are predicates that apply essentially to the whole living animal, not to its parts. It is not the eye (let alone the brain) that sees, but *we* see *with* our eyes (and we do not see with our brains, although without a brain functioning normally in respect of the visual system, we would not see.) (p72)

This is at least close kin to the point I made in 1969 when I distinguished the personal and sub-personal levels of explanation. *I* feel pain; my brain doesn't. *I* see things; my eyes don't. Speaking about pain, for instance, I noted:

> An analysis of our ordinary way of speaking about pains shows that no events or processes could be discovered in the brain that would exhibit the characteristics of the putative 'mental phenomena' of pain, because talk of pains is essentially non-mechanical, and the events and processes of the brain are essentially mechanical." (p91).

We have so much in common, and yet Bennett and Hacker are utterly dismissive of my work. How can this be explained? As so often in philosophy, it helps to have someone say, resolutely and clearly, what others only hint at or tacitly presuppose. Bennett and Hacker manage to express positions that I have been combating *indirectly* for forty years but have never before been able to confront head on, for lack of a forthright exponent. Like Jerry Fodor, on whom I have relied for years to blurt out vividly *just* the points I wish to deny–saving me from attacking a straw man–Bennett and Hacker give me a bold doctrine to criticize. I've found the task of marshaling my thoughts on these topics in reaction to their claims to be illuminating to me, and I hope to others as well.

The Philosophical Background

In this section, I am going to speak just of Hacker, leaving his co-author Bennett out of the discussion, since the points I will be criticizing are clearly Hacker's contribution. They echo, often in the same words, claims he made in his 1990 book, *Wittgenstein Meaning and Mind.* (Blackwell, 1990), and they are strictly philosophical.

When Hacker lambastes me, over and over, for failing to appreciate the mereological fallacy, this is a case of teaching your grandmother to suck eggs. I am familiar with the point, having pioneered its use. Did I, perhaps, lose my way when I left Oxford? Among the philosophers who have taken my personal level/sub-personal level distinction to heart, at least one–Jennifer

Hornsby–has surmised that I might have abandoned it in my later work.[6] Did I in fact turn my back on this good idea? No.[7] On this occasion it would be most apt to cite my 1980 criticism of Searle's defense of the Chinese Room intuition pump:

> The systems reply suggests, entirely correctly in my opinion, that Searle has confused different levels of explanation (and attribution). *I* understand English, my brain doesn't–nor, more particularly, does the proper part of it (if such can be isolated) that operates to 'process' incoming sentences and to execute my speech act intentions." (p429. 1980 *Behavioral and Brain Sciences*, vol 3.[8])

(This claim of mine was summarily dismissed by Searle, by the way, in his reply in BBS. I'll be interested to see what he makes of the personal level/subpersonal level distinction in its guise as the mereological fallacy.[9])

The authoritative text on which Hacker hangs his conviction about the mereological fallacy is a single sentence from St. Ludwig:

> It comes to this: Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees, is blind; hears, is deaf; is conscious or unconscious. [*Phil.Inv*, parag281.]

Right here is where Hacker and I part company. I am happy to cite this passage from Wittgenstein myself; indeed I take myself to be *extending* Wittgenstein's position: I see that robots and chess-playing computers and, yes, brains and their parts, *do* "resemble a living human being (by behaving like a human being)"–and this resemblance is sufficient to warrant an adjusted use of psychological vocabulary to characterize that behavior. Hacker does not see this, and he and Bennett call all instances of such usage "incoherent," insisting again and again that they "do not make sense." Now who's right?

Let's go back to 1969 and see how I put the matter then:

[6] Hornsby, 2000. Hacker's obliviousness to my distinction cannot be attributed to myopia; in addition to Hornsby's work, it has also been discussed at length by other Oxford philosophers: *e.g.*, Davies, 2000, Hurley, *Synthese*, 2001, and Bermudez, "Nonconceptual content: From perceptual experience to subpersonal computational states" *Mind and Language*, 1995.

[7] See also "Conditions of Personhood" in *Brainstorms,* (1978*).*

[8] See also the discussion of levels of explanation in *Consciousness Explained*, (1991).

[9] At the APA meeting at which this essay was presented Searle did not get around to commenting on this matter, having a surfeit of objections to lodge against Bennett and Hacker.

> In one respect the distinction between the personal and sub-personal levels of explanation is not new at all. The philosophy of mind initiated by Ryle and Wittgenstein is in large measure an analysis of the concepts we use at the personal level, and the lesson to be learned from Ryle's attacks on 'para-mechanical hypotheses' and Wittgenstein's often startling insistence that explanations come to an end rather earlier than we had thought is that the personal and sub-personal levels must not be confused. The lesson has occasionally been misconstrued, however, as the lesson that **the personal level of explanation is the *only* level of explanation when the subject matter is human minds and actions**. In an important but narrow sense this is true, for as we see in the case of pain, to abandon the personal level is to stop talking about pain. In another important sense it is false, and it is this that is often missed. The recognition that there are two levels of explanation gives birth to the burden of relating them, and **this is a task that is not outside the philosopher's province.** . . . there remains the question of how each bit of the *talk* about pain is related to neural impulses or talk about neural impulses. This and parallel questions about other phenomena need detailed answers even after it is agreed that there are different sorts of explanation, different levels and categories. (*Content and Consciousness*, p95-96)

This passage outlines the task I have set myself during the last 35 years. And the bold-faced passages mark the main points of disagreement with Hacker, for my path is not at all the path that he has taken. He gives his reasons, and they are worth careful attention:

> [A] Conceptual questions antecede matters of truth and falsehood. . . . Hence conceptual questions are not amenable to scientific investigation and experimentation or to scientific theorizing. (B&H, p2)

One can wonder about the first claim. Are not answers to these conceptual questions either true or false? No, according to Hacker:

> [B]" What truth and falsity is to science, sense and nonsense is to philosophy." (B&H, p6)

So when philosophers make mistakes, they produce nonsense, never falsehoods, and when philosophers do a good job we mustn't say they get it *right* or speak the *truth* but just that they make sense.[10] I am inclined to think that Hacker's [B] is just plain *false*, not nonsense, but be that

[10] For a philosopher who eschews truth and falsehood as the touchstone of philosophical propositions, Hacker is remarkably free with unargued bald assertions to the effect that so-and-so is mistaken, that such-and-such is wrong, and the like. These *obiter dicta* are hard to interpret without the supposition that they are intended to be true (as contrasted with false). Perhaps we are to understand that only a tiny fraction of his propositions, the specifically philosophical propositions, "antecede" truth and falsehood while the vast majority of his sentences are what they appear to be: assertions that aim at truth. And as such, presumably, they are subject to empirical

as it may, Hacker's second claim in [A], in spite of the "hence", is a *non sequitur*. Even if conceptual questions do "antecede" matters of truth and falsity, it might well behoove anybody who wanted to get clear about what the good answers are to investigate the relevant scientific inquiries assiduously. This proposal, which Hacker identifies as Quinian naturalism, he dismisses with an irrelevancy: "we do not think that empirical research can solve any philosophical problems, any more than it can solve problems in mathematics." (B&H, p414) Well of course not; empirical research doesn't *solve* them, it *informs* them and sometimes *adjusts* or *revises* them, and then they sometimes *dissolve*, and sometimes they can then be solved by further philosophical reflection.

Hacker's insistence that philosophy is an *a priori* discipline that has no continuity with empirical science is the chief source of the problems bedeviling this project, as we shall see:

> [C] How can one investigate the bounds of sense? Only by examining the use of words. Nonsense is often generated when an expression is used contrary to the rules for its use. The expression in question may be an ordinary, non-technical expression, in which case the rules for its use can be elicited from its standard employment and received explanations of its meaning. Or it may be a technical term of art, in which case the rules for its use must be elicited from the theorist's introduction of the term and the explanations he offers of its stipulated use. Both kinds of terms can be misused, and when they are, nonsense ensues–a form of words that is excluded from the language. For either nothing has been stipulated as to what the term means in the aberrant context in question, or this form of words is actually excluded by a rule specifying that there is no such thing as (e.g. that there is no such thing as 'east of the North Pole'), that this is a form of words that has no use. (B&H, p6).

This passage is all very reminiscent of 1960 or thereabouts, and I want to remind you of some of the problems with it, which I had thought we had figured out many years ago–but then, we didn't have this forthright version to use as our target.

> How can one investigate the bounds of sense? Only by examining the use of words. . . .

Notice, first, that, no matter what any philosopher may say, examining the use of words is an empirical investigation, which often yields everyday garden-variety truths and falsehoods, and is subject to correction by standard observations and objections. Perhaps it was a dim appreciation of this looming contradiction that led Hacker, in his 1990 book, to pronounce as follows:

> Grammar is autonomous, not answerable to, but presupposed by, factual propositions. In this sense, unlike means/ends rules, it is arbitrary. But it has a kinship to the non-arbitrary. It is moulded by human nature and the nature of the world around us. (1990, p148)

Let grammar be autonomous, whatever that means. One still cannot study it without asking

confirmation and disconfirmation.

questions–and even if you only ask *yourself* the questions, you still have to see what you say. The conviction that this method of consulting one's (grammatical or other) intuitions is entirely distinct from empirical inquiry has a long pedigree (going back not just to the Oxford of the 1960s, but to Socrates) but it does not survive reflection.

This can be readily seen if we compare this style of philosophy with anthropology, a manifestly empirical inquiry that can be done well or ill. If one chooses second-rate informants, or doesn't first get quite fluent in their language, one is apt to do third-rate work, For this reason, some anthropologists prefer to do one or another form of *autoanthropology*, in which you use yourself as your informant–perhaps abetted by a few close colleagues as interlocutors. The empirical nature of the enterprise is just the same.[11] Linguists, famously, engage in a species of this autoanthropology, and they know a good deal, at this point, about the pitfalls and risks of their particular exercises in teasing out grammatical intuitions regarding their native tongues. It is well known, for instance, that it is very difficult to avoid contaminating your intuitions about grammaticality with your own pet theoretical ideas. Some linguists, in fact, have been led to the view that theoretical linguists are, or should be, disqualified as informants, since their judgments are not naive. Now here is a challenge for Hacker and like-minded philosophers: How, precisely, do they distinguish their inquiry from autoanthropology, an empirical investigation that apparently uses just the same methods, and arrives at the same sorts of judgments.[12]

Anybody who thinks that philosophers have found a method of *"grammatical"* inquiry that is somehow immune to (or orthogonal to, or that "antecedes") the problems that can arise for that anthropological inquiry owes us an apologia explaining just how the trick is turned. Bald assertions that *this is what philosophers do* only evade the challenge. My colleague Avner Baz reminds me that Stanley Cavell[13] has made an interesting move towards meeting this burden: Cavell claims that the philosopher's observations about what we would say are more akin to *aesthetic* judgments. As Baz puts it, "you present your judgment as *exemplary*–you talk *for* a community" [personal communication], and this is fine as far as it goes, but since the anthropologist is also engaged in

[11] In *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness,* (2005) I describe some strains of contemporary philosophy of mind as *naive aprioristic autoanthropology* pp31-35). Hacker's work strikes me as a paradigm case of this.

[12] Notice that I am not saying that autoanthropology is always a foolish or bootless endeavor; I'm just saying that it is an empirical inquiry that yields results–when it is done right– about the intuitions that the investigators discover in themselves, and the implications of those intuitions. These can be useful fruits of inquiry, but it is a further matter to say under what conditions any of these implications should be taken seriously as guides to the truth on any topic. See *Sweet Dreams* for more on this.

[13] *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy,* 1979, second edn1999.

finding *the best, most coherent* interpretation of the data gathered (Quine's point about the principle of charity, and my point about the rationality assumption of the intentional stance), this normative or commendatory element is already present–but bracketed–in the anthropologist's investigation. The anthropologist cannot make sense of what his informants *say* without uncovering what they *ought to say* (in their own community) under many conditions. What is deliberately left out of the anthropologist's enterprise, however, and what needs defense in the philosopher's enterprise, is a *justification* for the following claim: This is what these people do and say, *and you should do the same*. As we shall see, it is Hacker's failure to identify the community he is speaking for that scuttles his project.

Back to [C]:

> . . . Nonsense is often generated when an expression is used contrary to the rules for its use. . . .

It is long past time to call a halt to this sort of philosophical pretense. Ryle notoriously claimed to identify "category mistakes" by appeal to the "logic" of existence claims, but let's face it: that was a bluff. He had no articulated logic of existence terms to back up his claims. In spite of the popularity of such talk, from Ryle and Wittgenstein and a host of imitators, no philosopher has *ever* articulated "the rules" for the use of any ordinary expression. To be sure, philosophers have elicited judgments of deviance by the hundreds, but noting that "we wouldn't say thus-and-so" is not expressing a rule. Linguists use an asterisk or star to make the same sort of point, and they have generated thousands of starred sentences such as

*An acorn grew into every oak.
*The house was rats infested.

But as any linguist will assure you, drawing attention to a judgment of deviance, even if it is part of a large and well described pattern of deviance, is not the same as uncovering the rules that govern those cases. Linguists have worked very hard for over forty years to articulate the rules of English syntax and semantics, and have a few modest corners in which they can plausibly claim to have elicited "the rules." But they also have encountered large areas of fuzziness. What about this sentence?

?*The cat climbed down the tree. [an example from Jackendoff]

Is this nonsense that violates "the rules" of the verb *to climb*? It's hard to say, and it may be that usage is changing. Such examples abound. Linguists have learned that something may sound a bit odd, smell a bit fishy, but still not violate any clear rule that anybody has been able to compose and defend. And the idea of *rules* that are ineffable is too obscurantist to be worth discussion. Philosophers' intuitions, no matter how sharply honed, are not a superior source of evidence in this manifestly empirical inquiry.

Back to [C]. Hacker goes on to divide the lexicon in two:

> The expression in question may be an ordinary, non-technical expression, in which case the rules for its use can be elicited from its standard employment and received explanations of its meaning. Or it may be a technical term of art, in which case the rules for its use must be elicited from the theorist's introduction of the term and the explanations he offers of its stipulated use.

I am tempted to assert that Hacker is just wrong (but not speaking nonsense) when he implies that the hallmark of a technical term is that it is "introduced" by a theorist who "stipulates" its use. Either that, or he is defining "technical term" so narrowly that many of the terms we would ordinarily agree to be technical wouldn't be so classified by him--and "technical term" is a technical term whose use he is stipulating here and now. Let Hacker have his definition of technical terms, then, narrow though it is. None of the terms that are the focus of the attacks in the book are technical in this sense, so they must be "ordinary, non-technical" terms–or they must be mongrels, a possibility that Hacker briefly considers in his 1990 book and dismisses:

> If neurophysiologists, psychologists, artificial-intelligence scientists, or philosophers wish to change existing grammar, to introduce new ways of speaking, they may do so; but their new stipulations must be explained and conditions of application laid down. What may not be done is to argue that since we know what 'to think', 'to see,' or 'to infer' mean and know what 'the brain' means, therefore we must know what 'the brain thinks, sees, and infers' means. For we know what these verbs mean only in so far as we have mastered their existing use, which does not license applying them to the body or its parts, save derivatively. Nor may one cross the new 'technical' use with the old one, as, for example, neuroscientist typically do in their theorizing. For this produces a conflict of rules and hence incoherence *in the neuroscientists' use* of these terms. (1990, p148-9).

This last claim–and it is also at the very heart of the 2003 book–is question-begging. If Hacker were able to *show us* the rules, and show us just how the new uses conflict with them, we might be in a position to agree or disagree with him, but he is just making this up. He has no idea what "the rules" for the use of these everyday psychological terms are. More tellingly, his insistence on an *aprioristic* methodology systematically blinds him to what he is doing here. Let him be *right*[14] in his conviction that he has an *a priori* method that gives him "antecedent" insight into the meanings of his ordinary psychological terms. He still needs to confront the burden of showing how his prolegomenon or stage-setting avoids the pitfall of what we might call conceptual myopia: treating *one's own* (possibly narrow and ill-informed) concepts as binding on others with different agendas and training. How, indeed, does he establish that he and those whose work he is criticizing are speaking the same language? That is surely an empirical question, and his failure to address it with sufficient care has led him astray. What he has done, in fact, is not good philosophy but bad anthropology: he went to cognitive science to "examine the use of words" and failed to notice that

[14] Can a philosopher like Hacker be *right* even if not aiming at the truth?

he himself was bringing *his* ordinary language into alien territory, and that *his* intuitions didn't necessarily apply. When he calls *their* usage "aberrant," he is making a beginner's mistake.

The use of psychological predicates in the theorizing of cognitive scientists is indeed a particular *patois* of English, quite unlike the way of speaking of Oxford philosophy dons, and it has its own "rules". How do I know this? Because I've done the anthropology. (You have to be a Quinian naturalist to avoid making these simple mistakes.) There is a telling passage in which Hacker recognizes this possibility but exposes his inability to take it seriously:

> Is it a new discovery that brains also engage in such human activities? Or is it a linguistic innovation, introduced by neuroscientists, psychologists and cognitive scientists, extending the ordinary use of these psychological expressions for good theoretical reasons? Or, more ominously, is it a conceptual confusion? (p70-1)

Hacker opts for the third possibility, without argument, while I say it's the first two together. There *is* an element of discovery. It is an empirical fact, and a surprising one, that our brains–more particularly, *parts* of our brains–engage in processes that are *strikingly like* guessing, deciding, believing, jumping to conclusions, etc. And it is *enough* like these personal level behaviors to warrant stretching ordinary usage to cover it. If you don't study the excellent scientific work that this adoption of the intentional stance has accomplished, you'll think it's just crazy to talk this way. It isn't.

In fact this is what inspired me to develop my account of the intentional stance. When I began to spend my time talking with researchers in computer science and cognitive neuroscience what struck me was that they unselfconsciously, without any nudges or raised eyebrows, spoke of computers (and programs and subroutines, and brain parts and so forth) *wanting* and *thinking* and *concluding* and *deciding* and so forth. What are the rules? I asked myself. And the answer I came up with are the rules for adopting the intentional stance. The *factual* question is: do people in these fields speak this way, and does the intentional stance capture at least a central part of "the rules" for how they speak? And the (factual) answer is Yes.[15] There is also, I suppose, a political

[15] Presumably Bennett, a distinguished neuroscientist, has played informant to Hacker's anthropologist, but then how could I explain Hacker's almost total insensitivity to the subtleties in the *patois* (and the models and the discoveries) of cognitive science? Has Hacker chosen the wrong informant? Perhaps. Bennett's research in neuroscience has been at the level of the synapse, and people who work at that sub-neuronal level are approximately as far from the disciplines of cognitive science as molecular biologists are from field ethologists. There is not much communication between such distant enterprises, and even under the best of circumstances there is much miscommunication–and a fair amount of mutual disrespect, sad to say. I can recall a distinguished lab director opening a workshop with the following remark: "In our lab we have a saying: if you work on one neuron, that's neuroscience; if you work on two neurons, that's psychology." He didn't mean it as a compliment. Choosing an unsympathetic informant is, of

question: do they have any right to speak this way? Well, it pays off handsomely, generating hypotheses to test, articulating theories, analyzing distressingly complex phenomena into their more comprehensible parts, and so forth.

Hacker also discovers this ubiquitous use of intentional terms in neuroscience and he's shocked, I tell you, shocked! So many people making such egregious conceptual blunders! He doesn't know the half of it. It is not just neuroscientists; it is computer scientists (and not just in AI), cognitive ethologists, cell biologists, evolutionary theorists . . . . all blithely falling in with the game, teaching their students to think and talk this way, a linguistic pandemic. If you asked the average electrical engineer to explain how half the electronic gadgets in your house worked, you'd get an answer bristling with intentional terms that commit the mereological fallacy–if it is a fallacy.

It is not a fallacy. We don't attribute *fully-fledged* belief (or decision, or desire –or pain, heaven knows) to the brain parts–that *would* be a fallacy. No, we attribute an attenuated sort of belief and desire to these parts, belief and desire stripped of many of their everyday connotations (about responsibility and comprehension, for instance). Just as a young child can *sort of* believe that her daddy is a doctor (without full comprehension of what a daddy or a doctor is[16]), so a robot–or some part of a person's brain–can *sort of* believe that there is an open door a few feet ahead, or that something is amiss over there to the right, and so forth. For years I have defended such uses of the intentional stance in characterizing complex systems ranging from chess-playing computers to thermostats, and in characterizing the brain's subsystems at many levels, The idea is that when we engineer a complex system (or reverse engineer a biological system like a person or a person's brain), we can make progress by breaking down the whole wonderful person into sub-persons of sorts, agentlike systems that have *part* of the prowess of a person, and then these *homunculi* can be broken down further into still simpler, less person-like agents, and so forth–a *finite*, not infinite, regress that bottoms out when we reach agents so stupid that they can be replaced by a machine. Now perhaps all my attempts at justifying and explaining this move are mistaken, but since Bennett and Hacker never address them, they are in no position to assess them.

Here is how I put it in a paper Hacker cites several times (though not this passage):

One may be tempted to ask: Are the subpersonal components *real* intentional systems? At

course, a recipe for anthropological disaster. [Added after the APA meeting: Bennett confirmed this surmise in his opening remarks; after reviewing his career of research on the synapse, he expressed his utter dismay with the attention-getting hypotheses and models of today's cognitive neuroscientists, and made it clear that the thought it was all incomprehensible. With an informant like Bennett, it is no wonder that Hacker was unable to find anything of value in cognitive neuroscience.

[16] See my *Content and Consciousness*, p183.

> what point in the diminution of prowess as we descend to simple neurons does *real* intentionality disappear? Don't ask. The reasons for regarding an individual neuron (or a thermostat) as an intentional system are unimpressive, but not zero, and the security of our intentional attributions at the highest levels does not depend on our identifying a lowest level of real intentionality. ("Self-portrait," in Guttenplan, listed by H & B as "Dennett, "Daniel C. Dennett," 1994)

The homunculus *fallacy*, by attributing the *whole* mind to a proper part of the system, merely postpones analysis and thus would generate an infinite regress since each postulation would make no progress. Far from it being a *mistake* to attribute hemi-semi-demi-proto-quasi-pseudo intentionality to the mereological parts of persons, it is precisely the enabling move that lets us see how on earth to get whole wonderful persons out of brute mechanical parts. That is a devilishly hard thing to imagine, and the poetic license granted by the intentional stance eases the task substantially.[17] From my vantage point, then, Hacker is comically naive, for all the world like an old-fashioned grammarian scolding people for saying "ain't" and insisting *you can't say that!* to people who manifestly *can* say that and know what they mean when they do. Hacker had foreseen this prospect in his 1990 book, and described it really very well:

> If all this [cognitive science] is to be taken at face value, it seems to show, first, that the grammatical remark that these predicates, in their literal use, are restricted to human beings and what behaves like human beings is either wrong *simpliciter* or displays 'semantic inertia' that has been overtaken by the march of science, for machines actually do behave like human beings. Secondly, if it makes literal sense to attribute epistemic and even perceptual predicates to machines which are built to simulate certain human operations and to execute certain human tasks, it seems plausible to suppose that the human brain must have a similar abstract functional structure to that of the machine design. In which case, surely it must make sense to attribute the variety of psychological predicates to the human brain after all. . . . .(1990, p160-61)

Exactly. That's the claim. How does he rebut it? He doesn't. He says "Philosophical problems stem from conceptual confusion. They are not resolved by empirical discoveries, and they cannot be answered, but only swept under the carpet, by conceptual change." (p161) Since Hacker's philosophical problems are becoming obsolete, I suppose we might just sweep them under the carpet, though I'd prefer to give them a proper burial.

[17] To take just one instance, when Hacker deplores my "barbaric nominal 'aboutness'" (p422) and insists that "Opioid receptor are no more *about* opioids than cats are about dogs or ducks are about drakes." (p423) he is of course dead right: the elegant relation between opioids and opioid receptors isn't fully-fledged aboutness (sorry for the barbarism), it is mere proto-aboutness (ouch!) but that's *just* the sort of property one might treasure in a mere part of some mereological sum which (properly organized) could exhibit *bona fide, echt*, philosophically sound, paradigmatic . . . intentionality.

The neuroscientific details

When Bennett and Hacker undertake to examine the neuroscientific literature, there is scant variety to their critique. They quote Crick and Edelman and Damasio and Gregory and many others saying things that strike them as “incoherent” because these scientists commit the so-called mereological fallacy.

> Far from being new homonyms, the psychological expressions they use are being invoked in their customary sense, otherwise the neuroscientists would not draw the inferences from them which they do draw. When Crick asserts that “what you see is not what is *really* there; it is what your brain *believes* is there,” it is important that he takes “believes” to have its normal connotations–that it does not mean the same as some novel term “believes*”. For it is part of Crick’s tale that the belief is the outcome of an *interpretation* based on previous *experience* or *information* (and not the outcome of an interpretation* based on previous experience* and information*. (p75)

In fact they are just wrong (but not nonsensical). Crick’s whole tale (and in this instance it is a quite banal and uncontroversial explanation) is intended by Crick to be understood at the sub-personal level. The interpretation in question is *not* of (personal level) *experience* but of, say, *data from the ventral stream,* and the process of interpretation is of course supposed to be a sub-personal process. Another passage in the same spirit:

> Similarly, when [J.Z.] Young talks of the brain’s containing knowledge and information which is encoded in the brain “just as knowledge can be recorded in books or computers”, he means knowledge, not knowledge*–since it is knowledge and information, not knowledge* and information*, that can be recorded in books and computers.

The authors have done nothing at all to establish that there is no concept of knowledge or information that can be encoded in both books and brains. There is a large and complex literature in cognitive science on the concept of information–and on the concept of knowledge (just think of Chomsky’s discussions of “cognizing” in response to an earlier parry in much the same direction as the authors’)–and the authors’ obliviousness to these earlier discussions shows that they are not taking their task very seriously. Many other examples in the same vein could be cited. They have one idea, the mereological fallacy, and they use it wholesale, without any consideration of the details. Each time they quote the offending passage–and they could have found a hundred times more instances of intentional stance attributions to brain subsystems–and then simply declare it nonsense because it commits their fallacy. Not once do they attempt to show that because of making this presumably terrible mistake the author in question is led astray into some actual error or contradiction. Who knew philosophy of neuroscience would be so easy?

Consider their discussion of the fascinating and controversial topic of mental imagery.

First they demonstrate–but I doubt anybody has ever doubted it–that creative imagination and mental imagery are really quite distinct and independent phenomena. Then comes their knockout punch: “A topographically arranged sensory area is not an image of anything; there are no images in the brain, and the brain does not *have* images.” (p183) As one who has argued strenuously for years that we must not jump to the conclusion that “mental imagery” involves actual images in the brain, and that the retinotopic arrays found therein may well not *function* as images in the brain’s processing, I must note that their bald assertion does not help at all. It is simply irrelevant whether “we would say” that the brain *has images.* Whether any of the arrays of stimulation in the brain that manifestly have the geometric properties of images *function* as images is an empirical question, and one that is close to being answered. Philosophical analysis is powerless to settle the issue–except by a deeply reactionary insistence that these image-like data structures, from which information is apparently extracted in much the way we persons (at the personal level) extract information visually from public images, don’t *count* as images. Such obfuscatory moves give philosophy a serious credibility problem in cognitive science.

In fact, there are serious conceptual problems with the ways in which cognitive scientists have spoken about images, and knowledge, and representations, and information, and the rest. But it is hard, detailed work *showing* that the terminology used is being misused in ways that seriously mislead the theorists. The fact is that for the most part, these terms, as they are found in cognitive science, really are “ordinary language”–not technical terms[18] that have been explicitly stipulated within some theory. Theorists have often found it useful to speak, *somewhat* impressionistically, about the information being processed, the decisions being reached, the representations being consulted, and instead of doing what a philosopher might do when challenged about what they meant, namely *defining their terms more exactly*, they instead point to their models and say: “See: here are the mechanisms at work, doing the information-processing I was telling you about.” And the models work. They *behave* in the way they have to behave in order to cash out *that* particular homunculus, so there need be no further cavil about exactly what was being attributed to the system.

But there are also plenty of times when theorists’ enthusiasm for their intentional interpretations of their models misleads them[19]. For instance, in the imagery debate, there have been missteps of over-interpretation–by Stephen Kosslyn, for instance--that need correction. It is

[18] in Hacker’s narrow sense.

[19] This has been an oft-recurring theme in critical work in cognitive science. Classic papers go back to William Woods’ “What’s in a link?” (in Bobrow and Collins, *Representation and Understanding*, 1975) and Drew McDermott’s “Artificial Intelligence meets Natural Stupidity” (in Haugeland, *Mind Design*, 1981) through Ulrich Neisser’s *Cognition and Reality* (1975) and Rodney Brooks “Intelligence without Representation” (*Artificial Intelligence*, 1991). They continue to this day, including contributions by philosophers who have done their homework and know what the details of the issues are.

not that map-talk or image-talk is *utterly* forlorn in neuroscience but that it has to be very carefully introduced, and it sometimes isn't. Can philosophy help? Yes it can, say Bennett and Hacker: "It can explain–as we have explained–why mental images are not ethereal pictures and why they cannot be rotated in mental space." (p405). This wholesale approach is not helpful. What is actually happening in the brain when people are engaged in mental imagery cannot be settled by making the point that the personal level is not the sub-personal level. The theorists already know that; they are not making *that* mistake. They are actually quite careful and subtle thinkers, and some of them *still* want to talk about images functioning as images in the brain. They may well be right.[20] Philosophers such as Hacker may lose interest once the topic is sub-personal,[21] but then they shouldn't make the mistake of criticizing a domain that they know little about.

Sometimes the authors miss the point with such blithe confidence that the effect is quite amusing, as in their stern chastising of David Marr:

> To see is not to discover anything from an image or light array falling upon the retina. For one cannot, in this sense, discover anything from something one cannot perceive (we do not perceive the light array that falls upon our retinae [*sic*], what we perceive is whatever that light array enables us to perceive.(p144).

Got it. Since *we* do not perceive the light array that falls upon our retinas, it is obvious that *we* do

[20] Hacker and Bennett say: " . . . it would be misleading, but otherwise innocuous, to speak of maps in the brain when what is meant is that certain features of the visual field can be mapped on to the firings of groups of cells in the 'visual' striate cortex. But then one cannot go on to say, as Young does, that the brain makes use of its maps in formulating its hypotheses about what is visible." (p77) But that is just what makes talking about maps perspicuous: that the brain *does* make use of them *as* maps. Otherwise, indeed, there would be no point. And that is why Kosslyn's pointing to the visible patterns of excitation on the cortex during imagery is utterly inconclusive about the nature of the processes underlying what we call, at the personal level, visual imagery. See Pylyshyn's recent target article in BBS, (April 2002) and my commentary, "Does your brain use the images in it, and if so, how?"

[21] "Philosophers should not find themselves having to abandon pet theories about the nature of consciousness in the face of scientific evidence. They should have no pet theories, since they should not be propounding empirical theories that are subject to empirical confirmation and disconfirmation in the first place. Their business is with concepts, not with empirical judgments; it is with the forms of thought, not with its content; it is with what is logically possible, not with what is empirically actual; with what does and does not make sense, not with what is and what is not true." (p404). It is this blinkered vision of the philosopher's proper business that permits Hacker to miss the mark so egregiously when he sets out to criticize the scientists.

not discover anything. Marr was not an idiot. He understood that. Now, what about Marr's theory of the *sub-personal processes* of vision?

> Moreover, it is *altogether obscure* [emphasis added] how the mind's having access to putative neural *descriptions* will enable the person *to see*. And if Marr were to insist (rightly) that it is the person, not the mind, that sees, how is the transition from the presence of an encoded 3-D model description in the brain to the experience of seeing what is before one's eyes to be explained? To be sure, *that* is not an empirical problem, to be solved by further investigations. It is the product of a conceptual confusion, and what it needs is disentangling. (p147)

I would have said, on the contrary, that it is a philosophical problem to be solved by addressing those who find it "altogether obscure" and leading them to an understanding of just how Marr's theory *can* account for the family of competences that a seer has.[22] Marr was more or less taking it for granted that his readers could work out for themselves how a model of the brain as having a consultable 3-D model of the world would be well on the way to explaining how a creature with just such a brain could see, but if this eluded some readers, a philosopher would probably be a good specialist to explain the point. Just *asserting* that Marr is suffering from a conceptual confusion has, as Russell so aptly put it, all the advantages of theft over honest toil.

> . . . For seeing something is the exercise of a power, a use of the visual faculty–*not* [emphasis added] the processing of information in the semantic sense or the production of a description in the brain. (p147)

This '*not*' is another theft. What has to be explained is the power of the "visual faculty" and that power is explained in terms of the lesser powers of its parts, whose activities include the creation and consultation of descriptions (of sorts). These examples could be multiplied to the point of tedium:

> It makes no sense, save as a misleading figure of speech, to say, as LeDoux does, that it is 'possible for your brain to know that something is good or bad before it knows exactly what it is. (p152)

But who is misled? Not LeDoux, and not LeDoux's readers, if they read carefully, for they can see that he has actually found a very good way to make the surprising point that a specialist circuit in the brain can discriminate something as dangerous, say, or as desirable, on the basis of a swift sort of *triage* that is accomplished *before* the information is passed on to those networks that complete

[22] For an example of such a *type* of explanation, see my simplified explanation of how Shakey the robot tells the boxes from the pyramids (a 'personal level' talent in a robot) by (sub-personally) making line drawings of its retinal images and then using its line semantics program to identify the tell-tale features of boxes, in *Consciousness Explained*.

the identification of the stimulus. (Yes, yes, I know. Only a person–a doctor or a nurse or such–can perform the behavior we call triage; I am speaking “metonymically”. Get used to it.)

In conclusion, what I am telling my colleagues in the neurosciences is that there is no case to answer here. The authors claim that just about everybody in cognitive neuroscience is committing a rather simple conceptual howler. I say dismiss all the charges until the authors come through with some details worth considering. Do the authors offer anything else that might be of value to the neurosciences? They offer no positive theories or models or suggestions about how such theories or models might be constructed, of course, since that would be not the province of philosophy. Their “correct” accounts of commissurotomy and blindsight–for instance–consist in bland restatements of the presenting phenomena, not explanations at all. They are right so far as they go: that’s how these remarkable phenomena appear. Now, how are we to explain them? Explanation has to stop somewhere, as Wittgenstein said, but not here. Bennett and Hacker quote with dismay some of the rudely dismissive remarks about philosophy by Glynn, Crick, Edelman, Zeki and others (pp396-98). On the strength of this showing, one can see why the neuroscientists are so unimpressed.[23]

[23] Bennett and Hacker’s Appendix 1: Daniel Dennett does not deserve a detailed reply, given its frequent misreadings of passages quoted out of context and its apparently willful omission of any discussion of the passages where I specifically defend against the misreadings they trot out, as already noted. I cannot resist noting, however, that they fall for the creationist canard that they presume will forestall any explanations of biological features in terms of what I call the design stance: “Evolution has not *designed* anything–Darwin’s achievement was to displace explanation in terms of design by evolutionary explanations.” (p425) They apparently do not understand how evolutionary explanation works.