

This work originally appeared in:

Dennett, Daniel. 1980. Reply to Professor Stich, (reply to S. Stich, Headaches: a Critical Notice of Brainstorms), *Philosophical Books XXI*: 73-6.

This is Daniel C. Dennett's final draft before publication. It has been modified to reflect the pagination of the published version of the work.

## PHILOSOPHICAL BOOKS

Vol. XXI No. 2 April 1980

### HEADACHES<sup>1</sup>

I. In philosophy of psychology, and I imagine in almost any field, there are some books that must be read by everyone who wants to keep abreast of current developments. Also, there are some books that are just plain fun to read. Daniel C. Dennett's *BRAINSTORMS* (Harvester Press, 353 pp., £13.50; Bradford Books, \$18.95) sits squarely in the intersection of these two classes. The views he defends are challenging and important. They are sure to be a major focus of attention, perhaps the major focus of attention, in philosophy of psychology during the next decade. Dennett's prose is lively and graceful. His examples, always apposite, are fascinating and colourful, and his expositions of recent research in psychology and artificial intelligence are models of their kind - clear, informative and accurate. The book is a collection of seventeen essays, ten of which have previously been published elsewhere. It is not, however, simply a collection of heterogeneous papers. Several central themes run throughout the book: the nature of mental states, the conceptual foundations of psychology, and "the relationship between our vision of ourselves as responsible, free, rational agents, and our vision of ourselves as complex parts of the physical world of science" (p. x). Indeed, Dennett urges that the essays taken together "express a theory of the mind".

The essays were written to be heard or read independently, and each is self-contained. Thus, Dennett claims, the reader can plunge in where he will with reasonable hope of understanding as much of the overall theory as is needed for the essay at hand. With a single caveat, I think this is true. Though the reader can choose freely what to read second, he would be well advised to start out with the brief Introduction. It is there, and only there, that Dennett attempts to put the pieces together, to say how his various arguments and conjectures fit into a single theory of the mind. The Introduction is all the more important because of Dennett's disconcerting tendency to soft-pedal his more controversial conclusions. In more than one essay he begins by posing a fascinating question, and follows it with a rich mix of argument, example, and report on recent science. But when we reach the last page, we are left wondering what

---

<sup>1</sup> My research has been supported by the U.S.-U.K. Educational Commission and by the American Council of Learned Societies.

answer he is urging for the question that started us off (cf. chs. 8, 9 and II). We must check back with the Introduction to see what all that argument was an argument for - though, as we shall soon see, there are anomalies in Dennett's view that even the Introduction does not explain away. It is a measure of his achievement that this elusiveness about the bottom line does not loom as a major shortcoming; indeed it is often hardly noticeable. Getting there is more than half the fun.

2. At its best, Dennett's mixture of scientific exposition and philosophical interpretation can produce gems. One of these is "Artificial Intelligence as Philosophy and as Psychology" (ch. 7). The essay announces itself modestly as a travel guide for philosophers contemplating a visit to the strange and alien domain of Artificial Intelligence (AI). And this it surely is. It is hard to imagine that a clearer or more informative sketch of what the Artificial Intelligentsia is up to could be given in a brief essay. But contrary to the expectations engendered by the introductory paragraph, there is much more here than a brief Baedeker for AI-land. There is also a crystal-clear statement of a venerable, though often vague, philosophical worry, and an argument that AI shows the worry to be unwarranted. Dennett labels the worry Hume's Problem, though it is more salient in the writings of authors with behaviourist sympathies like Wittgenstein, Malcolm and Skinner. The worry is that there is something profoundly confused about invoking such notions as cognitive maps, memory traces, and internal representations, in a putatively explanatory psychological theory. For, the argument goes, nothing is a representation or a map simpliciter. Something can be a map or a representation only for or to someone. A map or a representation is useless in guiding behaviour unless there is someone to interpret it. It cannot interpret itself. But then a psychological theory which postulates internal representations and their ilk must also postulate a homunculus to interpret them. For particularly clear statements of this worry, see B. F. Skinner, "Behaviorism at Fifty", in T. W. Wann, ed., *Behaviorism and Phenomenology* (Univ. of Chicago Press, 1964), pp. 79-80; also Norman Malcolm, *Memory and Mind* (Cornell Univ. Press, 1977), ch. IV and pp. 156-64.

Now homunculi have long had a bad press. Explanations invoking them are said to be circular or to threaten infinite regress; they explain how a big man does something by postulating a little man inside to guide him. But if homunculi are symptomatic of a "disease of thinking" (cf. Malcolm, *op. cit.*, p. 102), then AI is a terminal case, since "homunculus talk is ubiquitous in AI" (123). It is the burden of Dennett's argument that homunculi stand unjustly accused. The reason they have had such a bad press is that it has unfairly been presumed that they are smarter than they are. If we explain a subject's behaviour by postulating a homunculus with all the talents of the subject himself, then we have indeed made no explanatory progress. Progress comes when we postulate stupid homunculi. "If one can get a team or committee of relatively ignorant, narrow- minded, blind homunculi to produce the intelligent behavior of the whole,

this is progress." (123) Further progress can be made if we can explain the doings of each of these stupid homunculi by postulating teams of still stupider ones. "Eventually this nesting of boxes within boxes lands you with homunculi so stupid ... that they can be, as one says, 'replaced by a machine'." (124) Now this is just the strategy of AI as Dennett portrays it. The dumbest homunculi are simply the and-gates, or-gates and similar functional atoms of digital computers. And what AI theorists are up to is organizing "armies of such idiots" (124) to simulate intelligent behaviour. The philosophical lesson of AI is not that people are really computers whose hardware is soft. Rather it is that there is nothing suspect about a psychological theory invoking internal representations and homunculi to interpret them, so long as none of the homunculi "duplicate entire the talents they are rung in to explain" (123) and so long as there is a reasonable expectation of ultimately reaching a level of homunculi so stupid that what they do could be done (not by a machine but) by a nerve. The point is a profoundly important one, and Dennett's case for it is clear and compelling.

3. Dennett's collages of psychological fact and philosophical interpretation do not always turn out so well. "Why You Can't Make a Computer that Feels Pain" (ch. II) is an example. The question that starts the essay off is one of the philosophical chestnuts of the computer age: Could we design a robot that would feel pain? Having posed the question and intimated that he will argue for a negative answer, Dennett sets off on an intellectual rollercoaster ride. He has insightful things to say about the strategy of computer simulation and intriguing suggestions on the distinction between artificial and synthetic. Next we plunge into a cram course on the psychology and neurophysiology of pain, supplemented with some of Dennett's own speculations on what these sciences might ultimately find. All this is carried off with Dennett's usual flair. The science is presented with such verve and clarity that one wishes we could clone an army of Dennetts to rewrite all our science textbooks. And the speculation goes a long way toward making empirical speculation by philosophers respectable again.

Some of what Dennett has to report is, without exaggeration, spell-binding. He relates a curious and troubling episode involving the paralytic drug curare. The drug "acts directly on all the neuromuscular junctions... to produce total paralysis and limpness of all the voluntary muscles. It has no central effect except for a slight enhancement effect on activity in the cortex." (209) But in the 1940s some doctors mistakenly came to believe that curare was a general anaesthetic, and they used it as such for major surgery. The patients of course evidenced no pain during surgery, though they complained bitterly afterwards. However, perhaps because most of the patients were children, the doctors refused to take their complaints seriously. This horrifying history leads Dennett to ponder the following puzzle: "Suppose that one were to add to curare a smidgin of amnestic, a drug that (we will hypothesize) has no effect on experience or memory

during n hours after ingestion but thereafter wipes out all memory of those n hours." (210) Patients operated on under the influence of the imagined brew would presumably suffer excruciating pain, though they would "not later embarrass their physicians with recountings of agony" (210). But now, Dennett asks, how do we know that general anaesthetics in use today are not really curare-cum-amnestic? Driving home the point that this is not merely a philosopher's puzzle, Dennett reports that curare is routinely used as one ingredient in general anaesthetics today. Still more ominously, anaesthetists will sometimes administer scopolamine to get them off the hook when they think the patient may have been awake during surgery. And scopolamine is "the strongest and most reliable" known amnestic!

Fascinating facts, these, and tantalizing puzzles. The essay is a sure-fire tonic for any philosophy of mind course beset with the midterm blahs. But what about the question that started us off? Why couldn't a robot feel pain? And what has all this to do with the question anyway? Here is Dennett's answer, or at least the beginning of his answer: "The chief value of all this somewhat science-fictional flow-charting and compiling of odd phenomena - the reason I have spent so much time on it - is that it serves to drive a wedge of contingency between features that are often thought to be conceptually inseparable, simply because they are usually co-incident." (220) Fair enough, the wedge has been well driven. Pains are ordinarily thought to be abhorrent. But, as Dennett reports, lobotomized subjects and subjects under morphine analgesia can report pains, locate them with typical accuracy, and rank them in terms of greater and less intensity, yet they report that they do not mind the pains. Also, "pains are goal modifiers, but they might not be. That is, we can imagine a person who says he is in pain, locates the pain consistently, is in fact being beaten, writhes, cries, trembles, but is immune to torture." (220) So it looks as if typical examples of pain are cases marked by the confluence of a number of phenomena which need not necessarily occur together. Still, what's the point? Well, our intuitions about pain, the commonsense principles (or folk theory, if you will) which limn the boundaries of the ordinary concept of pain, seem to require that these phenomena must run together. So these intuitions are not consistent with undeniable empirical fact. Moreover, Dennett argues, the intuitions are not consistent with each other either. Therefore "what must be impeached is our ordinary concept of pain. A better concept is called for..." (225)

The thrust of the argument now seems to be coming clear. For the purposes of a serious psychological theory, the concept of pain will not do; it will have to be replaced by one or more new concepts cut to the pattern dictated by science. Is Dennett then an eliminative materialist about pains? A quick check back to the Introduction confirms this hunch. "About the theoretical entities in a mature psychology that eventually supplant beliefs, desires, pains... I am... a type intentionalist.... About other putative mental entities I am an eliminative materialist." (xx) Dennett explains his version of eliminative materialism with the aid of a

fanciful example of a community who are like us save in one respect. "When they are tired they talk of being beset by fatigues, of having mental fatigues, muscular fatigues..." etc. (xix). What is more, the philosophers among them sometimes puzzle over such questions as whether fatigues have a definite location in space and time, whether they are identical with some particular physical process or state in the body, etc. But however entrenched the term 'fatigues' may be in the thought and talk of this imagined society, Dennett maintains that "fatigues are not good theoretical entities" (xx). "The same is true," he holds, "of beliefs, desires, pains, . . . - as all these are ordinarily understood" (xx). In short, there are no such things as beliefs, desires and pains "though it is no easier to convince someone [of this]... than it would be to convince our imaginary people that there are no fatigues. If it can be done at all..., it can only be done by subjecting our intuitions and convictions about particular cases to skeptical scrutiny." (xx-xxi) So the point of all the science and speculation about pain was to convince us that there are no such things as pains.

Or was it? Flipping back to "Why You Can't Make a Computer that Feels Pain" (and now we are less than a page from the end) it looks at first as if we have hit upon the right interpretation: "But if, as I have claimed, the intuitions we would have to honor were we to honor them all do not form a consistent set, there can be no true theory of pain, and so no computer or robot could instantiate the true theory of pain, which it would have to do to feel real pain. Human beings and animals could no more instantiate the true theory of pain (there being none), which lands us with the outrageous conclusion that no one ever feels pain." (228) Now there are some philosophers, and I confess to being one, who take a wicked delight in defending outrageous views. But Dennett, it seems, does not share this vice. For in the very next sentence he pulls the rug out from under the eliminative materialist interpretation I have been trying to foist on him: "But of course we do [feel pain]. Human suffering and pain cannot be whisked out of existence by such an argument." (228) So pain does exist? But then what of Dennett's self-proclaimed eliminative materialism? And what was the point of all that "skeptical scrutiny" if it was not to convince us that there are no such things as pains? Here is what Dennett says: "The parochiality of the concept of pain protects us but not robots (or Martians or at least lower animals) from the skeptical arguments, by fixing the burden of proof: an adequate theory of pain must have normal human beings as instantiations, a demand that presupposes the primacy, but not the integrity, of our ordinary concept of pain. What then is the conclusion? It is that any robot instantiation of any theory of pain will be vulnerable to powerful objections that appeal to well-entrenched intuitions about the nature of pain, but reliance on such skeptical arguments would be short-sighted, for the inability of a robot model to satisfy all our intuitive demands may be due not to any irredeemable mysteriousness about the phenomenon of pain, but to irredeemable incoherence in our concept of pain." (228)

The reader who fails to see how these remarks reconcile the prima facie contradiction in Dennett's view will get no help from this reviewer. But if I cannot explain the contradiction away, perhaps I can explain how it arose. On my view, Dennett's problem with pain is a symptom of a much deeper difficulty with the story he wants to tell about the mind. There are two central themes in that story, themes which pull in opposite directions. Dennett has developed a clever strategy for gluing the themes together. However, the apparent contradiction in his remarks about pain marks one of the places where the glue comes unstuck. Let me elaborate on all this.

4. One of the themes that plays a central role in Dennett's thinking is the eliminative materialist line sketched above. The second is the puzzle of how we are to reconcile "our vision of ourselves as responsible, free, rational agents" with our vision of ourselves as "complex parts of the physical world of science" (x). The problem Dennett is struggling with is familiar enough. Our conception of ourselves as persons seems to require that we are beings "to which states of consciousness are attributed" (269) and to which intentional predicates like 'believes', 'desires', 'fears', 'hopes', etc. can be ascribed. But what if it should turn out that the best psychological theory, the best theory for predicting and explaining our behaviour, had no use for states of consciousness nor for any notion which much resembles our commonsense notions of belief, desire, etc. ? Would it follow, as writers like Thomas Nagel have suggested (PR 79 (1970), pp. 394-403), that we are not persons after all? Would we, perhaps, be forced to agree with writers like Skinner who hold that notions like freedom, dignity and moral responsibility are the heritage of mythical thinking and should be abandoned along with the scientifically useless notions of folk medicine or folk astronomy? The problem is an urgent one for an eliminative materialist like Dennett, since he thinks it is already pretty clear that the notions of folk psychology will not be of much use in scientific psychology. But Dennett is not willing to renounce the concept of personhood nor to undermine the foundations of our moral lives. He would have his cake and eat it too.

The central move in Dennett's attempt to reconcile personhood and scientific psychology is his introduction of the notion of an intentional system. Roughly, an intentional system is a system which, to all appearances, is rational through and through. It believes what it is rational to believe, given its perceptual history and capacities; it desires what it is rational to want, given its needs and beliefs; and it behaves in the most rational way, given its beliefs and desires. In interacting with objects in our world, one of the strategies or "stances" we can adopt is to view the object as an intentional system. To take this stance is to expect that the object will believe what is rational for it to believe, desire what is rational for it to desire, and behave accordingly. There is little point in adopting the intentional-system stance toward a stone, but it is quite sensible to adopt this stance toward a chess-playing computer - Dennett's favourite example. To adopt the intentional-system stance toward the chess-playing

computer is to anticipate that the computer will play as a rational agent would, a rational agent who wants to win. Often, indeed, this is the only practical strategy to adopt in dealing with a chess-playing computer.

On Dennett's view, our ordinary notions of belief and desire along with the other notions of commonsense intentional psychology are inextricably interwoven with the notion of an intentional system. To ascribe beliefs and desires to an organism or an object, we must adopt the intentional-system stance toward it. "Rationality is the mother of intention." (19) Though it is difficult to pin him down on the point, I do not think Dennett wants to claim that all the nuances of our ordinary notions of belief and desire are capturable via the notion of an intentional system. Rather, I read him as claiming that the notion of an intentional system captures the basic core of these commonsense notions. However, he clearly holds that the pared-down notions of belief and desire that are characterizable with the aid of the idea of an intentional system are adequate to sustain the concept of personhood. If having beliefs and desires is a necessary condition for being a person, then "intentional-system-characterizable" beliefs and desires will do. Recall, however, that in taking the intentional-systems stance toward an object or an organism we are making no claims of any substance about internal (or external) mechanisms responsible for its behaviour. We are merely deciding that, for the purposes at hand, the object is usefully viewed as an intentional system. We may be forced to abandon the stance if the object refuses to behave rationally. Thus, for example, if we were to embark unknowingly on a chess game with a computer that had been programmed to obey all the rules, but to select its own move randomly among the legal moves available, we would quickly be forced to abandon the intentional stance toward it. If, however, the computer plays as a rational chess player would, it is open to us to view it as an intentional system no matter how the behaviour is actually produced. It is just this fact about the intentional-system stance that enables Dennett to attempt his reconciliation between our view of ourselves as persons and our view of ourselves as parts of the physical world of science. To be a person it is necessary that we have beliefs and desires. And to have beliefs and desires it is necessary and sufficient that we be viewable as intentional systems. But nothing that physiology or scientific psychology could discover about the mechanisms underlying our behaviour could possibly show that we are not viewable as intentional systems, since the claim that we are so viewable entails nothing about these underlying mechanisms. To show that people cannot be viewed as intentional systems would require showing, as in the case of our random-move chess-playing computer, that people's behaviour is not generally predictable by assuming they are rational. But this we already know to be false. We assume the intentional-system stance toward each other all the time, and it works very well indeed, thank you.

We can now see how Dennett proposes to reconcile his eliminative materialism with the demands of morality and personhood. Beliefs and desires are to be eliminated from scientific psychology. They "are not good



theoretical entities" (xx). Nonetheless, we can still view people (along with certain animals and automata) as intentional systems, and thus as having beliefs and desires. Doing so enables us to predict their behaviour adequately for the purposes at hand, and that is all the justification we need for adopting the intentional-systems stance. So the fact that beliefs and desires play no role in a scientific psychology constitutes no threat to our "manifest image" of ourselves as persons.

I am profoundly sceptical about Dennett's proposed reconciliation. My reasons are two. First, I do not think that Dennett's notion of an intentional system really makes sense. In introducing the notion, Dennett relies heavily on a normative idea of rationality. Intentional systems believe what they ought to believe and want what they ought to want. But I do not know what a frog ought to believe or what a person ought to want, nor do I know how we would go about finding out. I suspect that it simply makes no sense to ask. Dennett tries to nail these notions down with evolutionary considerations, but I think this effort is simply confused. The second reason for my scepticism is that I do not think our ordinary notions of belief and desire, or anything much like them, are characterizable in terms of intentional systems. Rather, I suspect, the notion of an intentional system, such as it is, picks out a very special (and small) subset of the systems to which we comfortably attribute beliefs and desires. If Dennett were right that the notions of belief and desire were tied to the idea of full rationality, then we could not attribute a belief to a dog (or a man) without attributing to him belief in all the logical consequences of that belief. Also, it follows from Dennett's view that those psychologists who claim to be studying the ways in which people depart from a normative standard of inference in forming their beliefs must simply be confused (cf., e.g. R. Nisbett and L. Ross, *Human Inference* (Prentice-Hall, 1980)). For, according to Dennett, we can attribute beliefs to a person only if we view him as an intentional system. And intentional systems believe what they ought to believe; they do not commit inferential errors.

None of this is news to Dennett. Nor do I think that these paradoxical consequences of his view constitute an argument against it. At best they are symptoms that all is not well. I think these symptoms can be elaborated into a full-blown argument, but that is a project for another essay which I have in preparation ("Dennett on Intentional Systems"). Note, however, that even if my qualms about the notion of intentional systems can be quieted, pain and other qualia are still a problem for Dennett. He holds that neither beliefs nor pains are good theoretical entities for a scientific psychology. If we buy his story about intentional systems, then we can still truly attribute beliefs to our fellows, since we can view them as intentional systems. But what about pains? They are not "legitimized" by the intentional-systems stance. So it would appear that eliminative-materialist-Dennett has no choice but to deny that we can truly attribute pain or other states of consciousness to people. However, Dennett also holds that we are persons and that to be a person requires that states of consciousness can be attributed to us. Can the contradiction be resolved?

It is the singular virtue of reviews in this journal that the author gets to answer such questions himself.

UNIVERSITY OF MARYLAND

STEPHEN P. STICH

#### REPLY TO PROFESSOR STICH

I have no quarrel with Stich's summary of my views. Indeed, I am very grateful for the light it sheds on them, for in several instances he draws the connections I see between the issues better than I had drawn them myself. In particular, he accurately diagnoses and describes the strategic role I envisage for the concept of an intentional system, permitting the claim that human beings are genuine believers and desirers to survive almost any imaginable discoveries in cognitive and physiological psychology, thus making our status as moral agents well-nigh invulnerable to scientific disconfirmation. Not "in principle" invulnerable, for in a science-fiction mood we can imagine startling discoveries (e.g., some "people" are organic puppets remotely controlled by Martians) that would upset any particular home truths about believers and moral agenthood you like, and - more importantly - a partial erosion of our self-image as rational, self-controlled agents due to discoveries about our cognitive imperfections is not ruled out.

This attempt of mine to ground ascriptions of belief and desire not directly in the imagined details of people's "machine architecture" but in their predictability via a familiar stance is well described by Stich, but does not find favour with him. Ah me, to be so well understood, and yet not believed! Stich is "profoundly sceptical" about my reliance on a norm of rationality for any intentional system, for two reasons. First, he notes that I ground the concept of system-rationality in evolutionary considerations and suspects this appeal to evolution is "simply confused", and second, he finds the norm of rationality, however grounded, too demanding; its requirement of "full rationality" would disqualify all natural aspirants to the role of believer and desirer. He does not develop the first charge of confusion, no doubt for a good tactical reason: I have left my claim about the relation between rationality and evolutionary considerations so open-ended that it is hard to argue against efficiently. It is one of those disconcerting bits of soft-peddalling of which Stich rightly accuses me, but since I view the demand for premature precision to be one of philosophy's crippling mores, I unblushingly admit the charge. I have rendered my views on evolution somewhat more explicit - though not explicit enough for Stich - in "Three Kinds of Intentional Psychology", forthcoming in a Thyssen Philosophy Group volume edited by R. A. Healey. Here I will just indicate how minimal my reliance is. I grant the possibility of malignant beliefs and desires, of suicidal rationality, and of rationally irreproachable patterns of belief and desire that are only very remotely, tenuously and speculatively related to any clearly biological needs. All I maintain is that such cancer-like growths of cognitive mass, to

be expected in creatures as sophisticated and socialized as human beings, can only be understood as belief and desire structures on the assumption that they are the natural and designed extensions of control systems selected by nature because they are benign in the long run. Stich's view must be that by the time I have retreated to a position that is mild enough to avoid obvious counterexamples, it will be too mild to play the foundational role I require. This is not a position to defend in a short review, but it is certainly worth pursuing, and I look forward to Stich's detailed account.

Stich's second objection, that the norm of full rationality sets too high a standard for belief, runs into a similar cautionary proviso from me. Of course no one is "fully rational", but the occasions on which we are constrained to convict a person of irrationality are notoriously difficult to describe in terms of the putatively irrational beliefs (and desires) maintained. Why? Because, on my view, there is no defensible stable version, precisely because of the rationality requirement on belief and desire attribution. On a rival view, there would have to be some considerations (of internal structure, for instance) that permitted characterizing an irrational believer as one who, for instance, fully understood and believed that  $p$ , but also fully understood and believed that  $q$ , where  $p$  implies not- $q$ . But to echo a theme I have long cherished in Quine's work, all the evidence - behavioural and internal - we acquire for the correctness of one of these ascriptions is not only evidence against the other, but the best sort of evidence. Nothing could count more heavily against the interpretation of one bit of internal machinery as (subserving) the belief that  $p$  than the evidence in favour of interpreting another bit of internal machinery as the belief that  $q$ . So unless the psychologists who claim to study human foibles in inference couch their belief ascriptions very cautiously and circumspectly, they will indeed be confused - not "simply", but subtly. Of course they can study human irrationality, an all too ubiquitous phenomenon, but what they discover when they find good cases of it will not be unproblematic cases of people believing contradictions, believing two propositions that contradict each other, failing to believe the implications of some of their beliefs, or embracing invalid rules of inference. In normal affairs we tolerate the instability encountered: does Jones really believe that  $p$  (since it seems that he also believes that  $q$ )? We shrug; he sort of believes that  $p$ , he believes that  $p$  for almost all practical purposes. We tolerate the slack, but not because it never lands us in substantial difficulties. On the contrary, it often does, especially in cases where questions of moral responsibility arise. The fully accountable villain, for instance, whose genuine comprehension of right and wrong is matched by his clear view of his deed, is not a phenomenon to be simply discovered in nature, given our actual canons of belief ascription (given what we all recognize to have a bearing on ascriptions), but still we routinely find cases close enough to convict. The possibility of developing a norm-free, naturalistic theory of belief which would independently support (some of) these convictions is not ruled out, but any such theory would propose a

substantial revision of our concepts of belief and responsibility, and, moreover, its acceptability to us as a substitute would depend crucially on its preservation of the bulk of the belief ascriptions we obtain via the intentional stance with its assumption of rationality.

So my "protection" of personhood from the march of science makes a major concession: the conception of ourselves as intentional systems is granted to be an unrealizable idealization, and hence the more demanding concept of a person built upon it is also unavoidably an idealization, but it is this idealization, and no naturalistic but realizable counterpart concept, that figures in our ethical reflections. We need not fear that science might discover that our heads don't have the right arrangement of stuff for us to be believers, but only because we can see that what it is to be a believer is something we already know we only imperfectly achieve.

As Stich notes, however, even if I can save belief and desire via intentional systems theory, there is an untouched parallel problem about such things as pains. If pains turn out not to be good theoretical entities in a mature psychology (as I argue), they must be eliminated from my materialist ontology, and then how will I salvage a set of truths about "people in pain" (now in scare quotes) to provide *inter alia* the factual material for familiar ethical principles about avoiding causing pain and the like? Stich has accurately and sympathetically portrayed my case up to the moment of truth. Now how will I resolve the apparent contradiction between the claim that strictly speaking there are no such things as pains, and yet, of course, people do feel pain on occasion? Consider the options.

(A) Bite the bullet. There is no such thing as pain, and so nobody ever feels pain, and so much the worse for any vision of mankind, for instance a moral vision, that supposes this.

(B) Propose a reform. Although the ordinary concept of pain, as it figures in both folk psychology and ethics, is incoherent, one or another substitute can be found without this deficiency, and a revised vision of mankind built around this new theory of what pains are. This may have the result that some consequences are endorsed that are counter-intuitive (to some folks) - e.g., anaesthetics are gratuitous for fish, dogs, leucotomes,...; or, normal human beings in condition F are in pain in spite of their sincere protestations to the contrary. More complicated novelties are also likely.

(C) Eschew theory and side with common sense. We all know what pains are - they are distressingly familiar items in our manifest image, to use Sellars' term - even if we cannot say, in a reductionist spirit, what pains are, using only the proprietary terms of a mature psychology or neurophysiology. We also cannot say what voices are (as I claimed in *Content and Consciousness*), or what holes are (see D. K. Lewis and S. Lewis, *AJP* 48 (1970), pp. 206-12), or what haircuts are, but only the most doctrinaire reductionist would consider the product of reforming these notions in the language of science to be time well spent.

If the impasse is as I have described it, then there is not going to be any entirely satisfactory theory of pain, so a choice among these options is in

some measure a matter of policy, and depends on the purposes to be served. Alternative (A) is curiously short-sighted, for it utterly fails to alleviate the concerns and dissolve the bewilderments that would typically motivate one to ask what pains are, but still it has its uses; it rudely reminds the questioner that there is something misbegotten about the question. Strictly speaking, there could be no such thing as pain. Alternative (C) has shortcomings from another direction, for not all our perplexity about pain is of the manufactured, merely philosophical sort - as is our perplexity about holes, haircuts and voices (if we can manage to conjure it up) - and we need something like a theory to allay it. But if we take the analogies of alternative (C) to heart, we will let the theory we develop to dissolve the residual mystery find its theoretical terms where it will, with only sidelong glances at the ordinary concept of pain and the intuitions which support it. Still, for some purposes it will be useful to advertise this theory as a theory of pain, as a theory about the real phenomenon that people are actually talking about when they take themselves to be talking about pain. For instance, when technical controversies arise about the effectiveness of various treatments, or the capacity for pain of some species, it is best to ignore the philosophical scruples that in other contexts can make it important to deny that the theory we are relying on is a theory of pain.

I take the philosopher's contribution to the problem of pain to be exhausted once these various avenues and their attractions are made clear: philosophical medicine for the philosophical puzzles and science for the rest. For whose benefit, I wonder, would one then go on to construct a rigorously formulated, counter-example-proof ism about pain? Like attempts to complete the formula "(x)(x) is a living thing iff...)", it might make an interesting exercise, but not very interesting. This leaves me, then, right where Stich finds me: claiming that there are no such things as pains, although of course people do feel pain, and leaving it at that, trusting that the rest of my observations on the subject dissolve the air of paradox.

CENTER FOR ADVANCED STUDY IN THE  
BEHAVIORAL SCIENCES, STANFORD

DANIEL C. DENNETT