

What Can Be Learned from Brainstorms?

ROBERT CUMMINS

University of Wisconsin-Milwaukee

Brainstorms^{*} is a collection of Dennett's essays on the philosophy of mind and psychology. In spite of having been written over a period of years, it hangs together remarkably well, and exhibits a linear development often absent in "regular" books. In addition to being important and good, it is the most entertaining bit of non-fiction I've read in a long while.

The book is divided into four parts:

- I. Intentional Explanation and Attributions of Mentality
- II. The Nature of Theory in Psychology
- III. Objects of Consciousness and the Nature of Experience
- IV. Free Will and Personhood.

Parts I-III are a defense of what I hereby dub a "neo-functionalist" theory of mind and psychological explanation. Part IV is more or less what the title leads one to expect.

Brainstorms resists brief summary; I shall therefore rely on a catalogue of characteristic and central claims that I hope will help locate the book on the philosophical map and at the same time provide background for the discussion in Section Two. (1) Propositional attitudes are not mental states. In particular, beliefs and desires are not states—especially not representational states—of the systems that have them. Instead, the idiom of propositional attitudes ('intentional idiom' is Dennett's phrase, though he does not discuss non-propositional cases) finds its place in "intentional systems theory," which seeks to explain, predict and interpret behavior by rationalizing it. Thus, a system has propositional attitudes just in case it is possible to explain/predict/interpret its behavior by attributing propositional attitudes to it (Chapters 1, 3, 6, 14). (2) Homuncular explanations are legitimate provided (a) the homunculi appealed to do not duplicate the capacities they are supposed to explain, and (b) they can be "discharged" by analysis into homunculi whose capacities

^{*} Daniel C. Dennett. *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, Vermont: Bradford Books, 1978.

are not intentionally characterized (Chapters 4, 7). (3) The Law of Effect is logically tied to the concept of intelligence: we will not recognize as intelligent a system whose successes do not render more likely whatever leads to them (Chapter 5). (4) Introspective awareness, including awareness of one's "mental images," is simply fallible knowledge of one's semantic intentions (Chapters 2, 9, 10). (5) Dreaming (and, perhaps, experience generally) is a non-conscious process of construction and storage of material for later recall (Chapter 8). (6) Subpersonal cognitive accounts of pain and consciousness are possible, though incoherencies in our intuitions about consciousness and pain make it difficult to produce anything uniquely or completely satisfying (Chapters 2, 9, 11). About an entity realizing such a theory, Dennett writes:

... There will be at least the illusion that it is like something to be the entity. In fact it will tell us ... just what it is like. But inside it is all darkness, a hoax. Or so it seems. Inside your skull it is also all darkness ... Can it be said that just as there is some other point of view that *you* have, there is some other point of view that *it* has? [pp. 164-65]

(7) Judgments (or opinions), thought of as semantic intentions or dicta we are disposed to produce, should be sharply distinguished from beliefs. The former are representational states, but do not rationalize behavior as do the latter (Chapters 2, 16). (8) Explanations of novel behavior that rely on response generalization are question-begging because they assume an automatically theory-satisfying notion of "similar behavior" and "similar contingencies/circumstances" (Chapter 4). (9) Since the intentional stance toward something—explaining its behavior *via* intentional characterization—is compatible with both mechanism and indeterminism, so is the moral stance, resting as it does on intentional attributions (Chapters 12, 15). (10) The concept of a person is an idealization; hence, illuminating, sufficient conditions for personhood cannot be given, though a series of nested necessary conditions can be (Chapter 14). (11) Gödel's incompleteness theorem is irrelevant to the capacities of persons or computers, since it applies to abstracta, not to their realizations (Chapter 13).

II

Dennett doesn't raise the mind-body problem in *Brainstorms*; he attempts to solve it (but see his "Current Issues in the Philosophy of Mind," in *The American Philosophical Quarterly*, October, 1978). I

think it helps, though, to see his contribution as a response to a particular version of the problem. We seem to have two distinct languages (or sets of general terms) for describing persons—a language of physical characteristics (PL) and a language of mental characteristics (ML).¹ Both work fine. The problem is their almost complete independence of each other. Can both be true of the same thing? Dualists explain the independence by denying unity of subject matter: Mental and physical descriptions are not true of the same things, appearances to the contrary notwithstanding. Their problem is explaining mind–body interaction. Monists avoid this problem by insisting on the unity of subject matter. Their problem is to explain how ML and PL descriptions can be true of the same thing. Monists divide over what the “same thing” is, idealists opting for the mental, materialists for the material, and a few for “something else.”

Until recently, the favored monist strategy was to explain how ML and PL could be about the same thing by attempting to define one in terms of the other. Idealists attempted to define PL expressions in ML, and materialists attempted to define ML expressions in PL. (The “something elses” tended to duck the issue, which is, or ought to be, why their view never catches on.) This is what is generally meant by *reductionism* in the philosophy of mind. Reductionism can be tolerant or intolerant depending on whether failure of translation is taken to be a knock on the program or a knock on the recalcitrant idiom. A rather remarkable tendency toward tolerance, in the profession if not in the individual monists themselves, together with massive translation failure, produced a crisis in both monist camps: if persons are really physical (mental), how could there be truths about them not formulatable in PL (ML)? What else could it mean to insist that persons are physical (mental) if not that they are completely describable in PL (ML)? Reductionism appeared to many to be the only alternative to dualism, and it still appears that way to many.

But not to functionalists. Functionalism in the philosophy of mind is exciting precisely because it proposes to save physicalistic monism from an impossible reductionism. Functionalism in the philosophy of mind is, in the first instance, a strategy for reconciling ontological physicalism with anti–reductionism. For example: adding machines are as physically heterogeneous a class of things as one can imagine. The abacus and the electronic calculator have no physics in common that they don’t also share with rocks, and hence translating ‘x is an adding machine’ into PL is hopeless. What

adders—physical adders—share is a function, not a physics. We might make a list, relying only on PL specification, but this wouldn't be translation; it would be a demonstration of ontological physicalism for adders.

Until very recently, functionalism was the *only* strategy available for reconciling anti-reductionism with ontological physicalism, but Dennett has introduced another, which I may as well call *intentionalism*. Dennett is a functionalist about mental states, but he thinks that surprisingly little of ML traffics in such things. In particular, he thinks that the idiom of belief, desire, purpose—the idiom of propositional attitudes, in short—does not traffic in mental states.² Intentional characterization is legitimate, when it is, in virtue of its explanatory and interpretive success.³ This much is true of the attribution of mental states, as well, of course. What makes Dennett's view distinctive is its instrumentalist cast: there are no intentional states on his view, only true intentional characterizations.

Lingering doubts about whether the chess-playing computer *really* has beliefs and desires are misplaced; for the definition of intentional systems I have given does not say that intentional systems *really* have beliefs and desires, but that one can explain and predict their behavior by *ascribing* beliefs and desires to them.
[p. 7]

Though Dennett is a bit cagey about it, he is deeply committed to the view that lingering doubts about people are misplaced too: nothing *really* has beliefs and desires. It follows that the viability of one large part of ML—intentional characterization—is neutral with respect to dualism and monism, hence no *problem* for ontological physicalists. If intentionality is the mark of the mental, then the mental (thus far, anyway) is no problem either. Intentionalism thus constitutes a second strategy for defending anti-reductionist monism, and a good part of Dennett's effort is directed to expanding this strategy by assimilating as much of ML as possible to its intentionalist component. (See claims (4), (5) and (6).) There are lots of details to quibble about there, but I don't think it's profitable to pursue them in a review. Instead, I hope to get a kind of fix on Dennett's intentionalism by raising a very general question about it.

What worries me about the intentionalist strategy is that it seems to reduplicate a version of the mind-body problem. Functionalism gives us an account (actually a variety of accounts) of how a certain part of ML—the part that traffics in psychological states—relates to

PL. Call this part of ML *functional state language* (FSL). (In Dennett's terms, functionalism is an account of how the design stance relates to the physical stance.) What's left when we subtract FLS from ML? On Dennett's view, we have (at least) intentional language—IL. Now how does IL relate to FSL? We *seem* to have two largely independent chunks of descriptive machinery that *seem* to apply to the same thing. According to Dennett, they *do* apply to the same thing. Metaphysical dualism isn't allowed as an option here: since IL is supposed to be metaphysically neutral, its relative independence cannot be metaphysically explained. So it *seems* we aren't back to square one (though I shall question that assessment below). But neither are we home free, for we still lack an account of the relation between IL and FSL. Compare functionalism: functionalism "locates" us in PL, as it were, and leads us to see from that perspective how an FSL specification could constitute a characterization of the PL territory "around us." Analogously, intentionalism must locate us in FSL and lead us to see from that perspective how an IL specification could constitute a characterization of our FSL environment. This would allow us to understand the intentional from a physicalist perspective without reducing it, since we already know how to understand (without reduction) FSL from a physicalist perspective.

In Dennett's terms, this amounts to saying how the intentional stance is related to the design stance, but that way of putting it doesn't focus on the problem I want to raise. For while Dennett does tell us something about the relation of the two stances, he doesn't tell us how an inhabitant of FSL-land is to assimilate IL characterizations, and I think only this, or something comparable, would make us understand the intentional.

The characterization in terms of stances tempts us to substitute questions of justification for questions of truth: we ask, "what would justify adoption of the intentional stance?" rather than "in virtue of what sorts of FSL-characterizable facts does (would) an IL characterization truly apply?" Dennett has a good deal to say about the former question, but little if anything to say about the latter.⁴ A full-fledged intentionalism must show how intentional idioms can fill the logical space Dennett has created for them. Functionalism went through a comparable stage. No one wants to espouse metaphysical dualism for adders; hence a demonstration of the non-reducibility of 'x is an adder' to the idioms of PL leaves us free to use functional concepts without reducing them or facing charges of dualism. This legitimizes functional characterization, but doesn't

explicate it. The second step, still very much in progress, is to see how PL-characterizable facts could make an FSL description true. Once again, calculators come to the rescue: when we design an adder—tell someone how to build one—what we do is give a PL description. Slogging through a design problem of this sort is therefore showing how FSL can truly describe PL-land. Since FSL isn't reducible to PL, it's the sort of thing that has to be done case-by-case. *C'est la vie*.

Well, then, what *would* it take to explicate an IL characterization from an FSL (or PL) perspective? I confess to having a kind of theory (or hunch) about this, which I want to follow for a bit because reflecting on it has convinced me that an instrumentalist treatment of propositional attitudes will ultimately undermine their explanatory value. (Actually, this should come as no surprise: if there are not any, how can they explain anything?)

My hunch is best introduced via some simple reflections on functional interpretation. Suppose we ask what makes a circuit an AND-gate? It is this: there are types of events identified as inputs, and types identified as outputs, and a rule for interpreting these as representations of truth-values, such that the output represents 'true' iff each of the inputs represents 'true'.⁵ This is enough to make something an AND-gate, but it isn't enough to make it important to treat it as such. Assuming it is relative dc levels that are interpretable as representations of truth-values, why look at relative dc levels rather than temperature, or mass, or shape? Well, if we hook a lot of these and kindred circuits together in fancy ways, we will have a system whose most striking capacities are explicable only by analyzing them into simple capacities such as the capacity to compute a truth function. Thus, it is the need to explain the capacities of a containing system that makes the truth-functional characterization useful—gives it a role to play. Indeed, only systematic context can make it right to say that it is an AND-gate and not a NAND-gate—or a resistor—and we might put this point by saying that only in an appropriately ritzy neighborhood is it *actually* an AND-gate; otherwise it is only potentially an AND-gate, a would-be AND-gate, the sort of thing that would be an AND-gate in a better neighborhood.

Now my hunch is that what makes intenders, believers, and pretenders is something rather analogous. To understand the intentional is to see that and how something satisfying an FSL description would receive an IL description in the course of explaining the capacities of a containing system—e.g., a community

of Gricean communicators. Thus, psychological states are at best would-be intentional states—the sort of thing that would actually be (or count as) an intentional state in a sufficiently ritzy neighborhood. What makes intentional characterization apt, in other words, is something similar to what makes it apt to treat dc levels as representations of truth-values. If this is on the right track, then to understand the intentional we need to know what sort of system would have capacities *requiring* intentional characterization, i.e., how a particular FSL-characterized structure could have such capacities. To see this would be to see how FSL facts could *make* an intentional characterization true.

A paradigm case would be a system consisting (in part) of two components capable of being the *parties* to a Lewis-type convention.⁶ (A pair of people in a natural environment is such a system, but perhaps there are others.) The capacities (or some of them) of such a system would require characterization in intentional terms, and a function-analytic explanation of those capacities would entail intentional characterization of the components. Hook a couple of fancy computational engines together in the right way, and you get a system whose capacities are explicable, indeed describable, *only* by construing them as such capacities as the capacity to believe, desire and intend. Thus, 'believer' turns out to be a functional term applying to things with the capacity to believe. Like 'AND-gate,' it will have a (containing) system-relative use, applying to actual believers, and a categorical use, applying to would-be believers-systems that would be believers in a ritzy enough neighborhood. Adapting a term of Stich's, I call capacities that would require intentional characterization in an appropriate context 'sub-doxastic' capacities.⁷ A sub-doxastic capacity is what you get when you take out of context a capacity used to analyze an intentionally characterized capacity such as conventional communication or reference.

Notice that this gives us believers, but not beliefs. My theory is neutral over whether beliefs are mental states or not: it all depends on how believers are to be analyzed (or perhaps instantiated), and I've said nothing about that. Still, although I've taken my cue from Dennett's idea that intentional idiom explains and interprets environmental interactions, I don't think I can sustain the rather radical instrumentalism about intentional states that Dennett recommends. After all, when a believer does its stuff, it believes, and this, it seems, must be a state or process of (or in) the believer, just as adding is a state or process in the adder. For surely to say that

a believer is believing is, on my account at any rate, to give an intentional characterization of a process or state in the believer. Still, it would be a great mistake to unhook a believer from its containing system, open it up, and look for believings (let alone beliefs). This would be like yanking an AND-gate out of a calculator, and searching for truth-functional computations (or truth-values). Yet this is just what we expect if we demand that the intentional properties of a psychological state or process be derivable from its "psychological" characterization, for psychology (at least as pictured by philosophers) takes as its unit a system whose typical capacities are, at the chosen level of analysis, only sub-doxastic,⁸ not actually intentional, but only potentially so.⁹

Let us call this a *contextualist* account of intentionality, though it is really a functionalist account of a property of supra-personal systems. Perhaps some capacities of less-than-whole organisms are actually intentional, but it seems unlikely. Dennett sees no harm in such characterization, but gives no compelling reason for thinking it necessary. My hunch is that, being an instrumentalist about the intentional, he is prepared to use intentional idiom whenever it is useful, whereas I am inclined to use it only when cornered by lack of explanatory or descriptive alternatives. Dennett tends to think the design stance (or even the physical stance) is always an alternative. In a way, this is right: we can choose to operate exclusively in PL, say. But if we do, there are a lot of capacities we won't even be able to describe, let alone explain—that's the anti-reductionist lesson.

For the most part, Dennett resists a functionalist account of propositional attitudes: when we can explain behavior by rationalizing it, the line goes, intentional idiom is appropriate, and that is all there is to say about how intentional idiom relates to the world. I've been claiming, on the other hand, that there is more to be said; that we won't understand propositional attitudes until we know how design facts can force the intentional upon us. It must be admitted, I think, that there is something of this in Dennett as well. On Dennett's view, when we characterize intentionally, we introduce a homunculus, but homunculi need to be discharged eventually.

Homunculi are *bogeymen* only if they duplicate *entire* the talents they are rung in to explain 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. . . . Eventually this nesting of boxes within boxes lands you with homunculi so stupid . . . that they can be, as one says, "replaced by a machine." [123-24]

Dennett rightly insists that the homunculi must be “discharged,” but he doesn’t say *how*. Functionalism tells us how to discharge functions, but not homunculi, unless we assume, contra Dennett, that intentional characterization is functional characterization. Lacking an account of how homunculi (real ones, not sub-doxastic ones) are to be discharged—i.e., lacking an account of how IL relates to FSL—we cannot be at all sure that they will be dischargeable whenever they are useful. Thus, the instrumentalist criterion—usefulness—may be much more liberal than the methodological one (dischargeability). Perhaps Dennett would plump for the methodological criterion if this should turn out to be the case. In any event, Dennett’s instrumentalism about intentional characterization will not be entirely persuasive until it is grounded in an account of how homunculi are discharged. We need something sort of analogous to slogging through calculator design. But only sort of: when we design a calculator, we design something that acutally manipulates numerals, but in designing a homunculus, we had best not design something that has and manipulates beliefs, for, on Dennett’s view, there aren’t any.¹⁰

NOTES

1. These are not mutually exclusive or jointly exhaustive. There are descriptions that seem to fall into both, and some that seem to fall into neither. But there are plenty of clear cases.

2. He seems committed to the view that intentional idiom generally does not traffic in mental states, but he does not discuss the non-propositional cases such as fearing death. Perhaps he holds that these are really disguised propositional cases.

3. Explanatory: “Why did she castle? To prevent the knight from forking her king and rook.” Interpretive: “What is he doing? Trying to attract attention.”

4. I’m not persuaded by what Dennett has to say in response to the question of justification. Dennett argues that natural selection guarantees excellence of design, hence rationality, the precondition of intentional explanation and interpretation. Evolution does no such thing; our cognitive capacities may well kill us. Evolution takes a long time. It’s *much* too soon to conclude that our cognitive design is adaptive. Insects are doing fine without much rationality.

5. Similarly, what makes something an adder is the fact that each output is interpretable as a numeral representing the sum of the numbers represented by the numerals that are the interpretations of the inputs. See John Haugeland’s general characterization of an “intentional black box” in [2].

6. David Lewis, [3]. See also Cummins [1].

7. S. P. Stich [4]. My use of this term may be more like theft and distortion than adaptation.

8. A full-fledged instrumentalism about the intentional would be justified if it turned out that intentional capacities were explicable via intentional instantiations (see Haugeland [2]), for the dimension-shift involved would erase intentional characterization and leave the analyzing of capacities to be carried out in some other idiom.

9. If this is on the right track, then cognitions are not psychological states, though some psychological states may be would-be cognitions. This is what makes AI programs such as Winograd's SHURDLU seem odd. It isn't that, since there aren't any blocks, SHURDLU doesn't know about them. It doesn't know *about* anything. But perhaps it would—perhaps what it does would be knowing about blocks—in a sufficiently ritzy setting. AI abstracts from the setting, philosophers notice the lack of intentionality, and a controversy ensues as to what has been accomplished. Part of this controversy is real enough, for it might be that no surroundings, however fancy, would require intentional characterization of a machine executing SHURDLU. But part of it is just verbal, since standard AI practice is to work with the sub-doxastic intentional idioms.

10. On the "contextualist" alternative, what we need to do is slog through the problem of designing a "community" that has a convention (or something comparable that has capacities requiring intentional characterization). When we do this, I am betting, we will have to design components that function as believers and intenders. It is only in a context like this—one in which intentional analysis is *forced*—that the need for discharging homunculi is *serious*. Dennett's view has it that I have discharged a homunculus (at least one) in designing an adder, since I can treat an adder as an intentional system. It is too easy that way.

REFERENCES

1. Cummins, Robert. "Intention, Meaning and Truth Conditions." *Philosophical Studies* 35 (1979), 345-60.
2. Haugeland, John. "The Nature and Plausibility of Cognitivism." *Behavioral and Brain Sciences* 2 (1978), 215-60.
3. Lewis, David. *Convention*. Cambridge: Harvard University Press (1969).
4. Stich, S. P. "Beliefs and Subdoxastic States." *Philosophy of Science* 45 (1978), 499-518.