

Intentionality

(with John Haugeland), in R. L. Gregory, ed., *The Oxford Companion to the Mind*, Oxford University Press 1987; reprinted in *Actes du 3ème Colloque International Cognition et Connaissance: Où va la science cognitive?* Toulouse: CNRS/Université Paul Sabatier 1988; reprinted in K. Lehrer and E. Sosa, eds., *The Opened Curtain: A U.S.-Soviet Philosophy Summit*, Westview Press, 1991, Chapter 3.

Intentionality

Daniel C. Dennett and John Haugeland

Intentionality is *aboutness*. Some things are about other things: a belief can be about icebergs, but an iceberg is not about anything; an idea can be about the number 7, but the number 7 is not about anything; a book or a film can be about Paris, but Paris is not about anything. Philosophers have long been concerned with the analysis of the phenomenon of intentionality, which has seemed to many to be a fundamental feature of mental states and events.

The term was coined by the Scholastics in the Middle Ages, and derives from the Latin verb *intendo*, meaning to point (at) or aim (at) or extend (toward). Phenomena with intentionality point outside themselves, in effect, to something else: whatever they are *of* or *about*. The term was revived in the 19th century by the philosopher and psychologist Franz Brentano, one of the most important predecessors of the school of Phenomenology. Brentano claimed that intentionality is the defining distinction between the mental and the physical; all and only mental phenomena exhibit intentionality. Since intentionality is, he claimed, an irreducible feature of mental phenomena, and since no physical phenomena could exhibit it, mental phenomena could not be a species of physical phenomena. This claim, often called the Brentano Thesis, or Brentano's Irreducibility Thesis, has often been cited to support the view that the mind cannot be the brain, but this is by no means generally accepted today.

There has been a second revival of the term in the last twenty years by English and American philosophers in the analytic tradition. In response to seminal work by R. Chisholm and W. V. O. Quine, a vigorous attempt has been made to develop an account of intentionality in harmony with the canons of modern logic and semantics. Since the Phenomenological tradition, mainly on the Continent, has continued to exploit the concept of intentionality along rather different lines, the problem of intentionality is one of the best points of convergent concern in these two largely separate--and often antagonistic--research traditions.

In spite of the attention currently devoted to the concept, there is a striking lack of received wisdom about the proper analysis of intentionality. What agreement there is concerns the nature of the problems raised. If we make an initial rough catalogue of the things that can be about things, it will include a great variety of *mental* states, and events (ideas, beliefs, desires, thoughts, hopes, fears, perceptions, dreams, hallucinations . . .) but also various *linguistic* items (sentences, questions, poems, headlines, instructions . . .) and perhaps other sorts of *representations* as well (pictures, charts, films, symphonic tone poems, computer programs, . . .). Many have thought that these linguistic and nonlinguistic representations are only derivatively about anything. They depend for their intentionality on their being the creations and tools of creatures with minds, and particular representations derive their specific aboutness from the specific aboutness of the ideas, beliefs, and intentions of their creators. What makes the particular sequence of ink marks on the next line

Napoleon was exiled to Elba.

about Napoleon is its role as a sentence in a language used by people who *know* about Napoleon, or at least have *beliefs* about Napoleon, and *wish* to communicate about Napoleon. What *representations* mean depends on what *people* mean by using them. This suggests that the primary or underived intentionality of

mental states and events is a very special feature indeed--the source of all meaning in the world. On this view sentences, pictures, diagrams, and the like are in effect prosthetic extensions of the minds of those who use them, having no intrinsic meaning but only the meaning they derive from their utilization.

Identifying intentionality with aboutness nicely locates the concept, but hardly clarifies it, for the ordinary word "about" is perplexing on its own. A belief can be about Paris, but a belief can also apparently be about phlogiston--and there is no phlogiston for it to be about. This curious fact, the possible non-existence of the *object* of an intentional item, may seem to be an idle puzzle, but in fact it has proven extraordinarily resistant to either solution or dismissal. Brentano called this the *intentional inexistence* of the intentional objects of mental states and events, and it has many manifestations. I cannot want without wanting something, but what I want need not exist for me to want it. It can be true that I now want a two-ton diamond even if there is not now and never has been or will be such a thing. People have believed in Poseidon, and children often believe in Santa Claus, and while in one sense we can say these believers all believe in nothing, their beliefs are quite different states of mind--they have different *objects*--and both are to be distinguished from the state of mind of the skeptic or agnostic who can be said in quite a different sense to believe in nothing.

It might seem that there is a simple and straightforward solution to these problems: while Poseidon does not exist, an idea of Poseidon surely exists in each of the believers' minds, and this idea is the object of their belief. This will not do, however, for it is one thing to believe in the existence of the *idea* of Poseidon--we can all believe in the existence of that mental item--and quite another to believe in Poseidon. Similarly, what I might want could hardly be the idea in my mind of a twoton diamond, for that is something I already have. Moreover, when, as normally happens, the object of an intentional state does exist--e.g., when I believe that London is crowded-- that very object, London, the city itself and not my idea of it in my mind, is what my belief is about.

The relation, then, between a state of mind--or for that matter a sentence or picture--and its intentional object or objects is a peculiar relation in three ways.

First, for ordinary relations, like *x is sitting on y* or *x is employed by y*, *x* and *y* are identifiable entities quite apart from whether they happen to be thus related to each other. The thing which is *x* would be the same *x* whether or not it were sitting on *y* or employed by *y*. But the same is not true of intentional "relations". One and the same belief cannot at one moment be about a frog (that it is green, say) and at another moment be about a house (that it is green). The latter is a different belief. What a belief is supposed to be about is crucial to which belief it is.

Second, for ordinary relations, each of the things related must exist (or have existed), but as we have seen, intentional "relations" can be to non-existents.

Third, ordinary relations obtain between things regardless of how they might be specified. If I am sitting next to Jones and Jones is the Mad Strangler, then it follows that I am sitting next to the Mad Strangler, whatever anybody may think. But if I believe that Jones is harmless, or hope that he will marry my sister, it does not at all follow that I believe that the Mad Strangler is harmless or hope that the Mad Strangler will marry my sister. Even if one is tempted to object that in this case in one sense I do hope the Mad Strangler will marry her, there is clearly another sense in which I might hope this and don't.

For these reasons the normal logic of relations cannot accommodate the presumed relation between an intentional state and its intentional object or objects, but it has also not proven comfortable for theorists to deny on these grounds that there are such things as intentional relations--to hold that mental states, for instance, are only apparently relational. This, then, is the unsolved problem of intentionality.

Faced with this problem, the Anglo-American tradition, characteristically, has tended to favor a tactical retreat, to a logical analysis of the language we use to talk about intentional states, events and other items. This move, from the direct analysis of the phenomena to the analysis of our ways of talking about the phenomena, has been aptly called "semantic ascent" by Quine, and its immediate advantages are twofold.

First, we set aside epistemological and metaphysical distractions such as: "How can we ever know another person's mental state anyway?" and "Are mental states a variety of physical state, or are they somehow immaterial or spiritual?". The things people say about mental states are in any event out in the public world where we can get at them and study them directly. Second, switching to language puts at our disposal a number of sophisticated techniques and theories developed by philosophers, logicians, and linguists. Semantic ascent is not guaranteed to solve any problems, of course, but it may permit them to be reformulated in ways more accessible to ultimate solution.

In its new guise, the problem of intentionality concerns the semantics of the so-called intentional idioms-- ". . . believes that p", ". . . desires that q", ". . . dreams that r", etc. (where "p", "q", and "r" are replaced by clauses, such as "frogs are green" or "Reagan will soon retire"). Linguistically and logically, intentional idioms are a subset of those that Quine calls "referentially opaque". What this means is that many normally valid logical moves are not valid for the clauses "within the scope" of intentional idioms. For instance, normally, if two words happen to be words for the same thing, then one can freely substitute one for the other without affecting the truth of the whole sentence (although one may change its meaning, or effectiveness, or style). Thus, since "Cicero" and "Tully" are names for the same man, from the truth of "Cicero was an orator" we can infer the truth of "Tully was an orator". By contrast, however, the same substitution is not always allowable in "Tom believes that Cicero denounced Catiline", for Tom may believe that Cicero denounced Catiline but not believe that Tully did. So ". . . believes that p" is an opaque idiom.

Clearly this is just another (and more precise) way of putting the point made earlier that intentional relations depend on how their objects are specified. The other points have analogues too. Thus, normally, a relational statement is false if one of the alleged relata doesn't exist; not so within the scope of opaque idioms. And, of course, the fact that the identity of a particular belief depends on the object or objects it is supposed to be about emerges on this treatment as the fact that the ascription of a particular belief depends crucially on the words used in the clause expressing it. We can see now that this condition is essentially the same as the first, and that the second condition (possible non-existence of the object) is also just a special case of opacity: believing that Santa Claus is generous and believing that Poseidon is generous are different beliefs, in spite of the fact that "Santa Claus" and "Poseidon" refer to "the same thing", i.e., to nothing.

Seeing this unity in the various conditions of intentionality is one of the benefits of semantic ascent. Another is that it thus provides us with a relatively formal and uncontroversial test for the intentionality of idioms, and hence a test for appeals to intentionality in a theory. This is an interesting test, for theories relying on intentional idioms-- such as classical "rational agent" economics and cognitive psychology--cannot be formulated in any noncontroversial way within standard logic, while it seems that other theories, pre- eminently in the physical sciences, can be so formulated. The logical oddity of intentional idioms, and their resistance to regimentation, has led Quine and several other theorists to declare the bankruptcy of all intentional theories, on grounds of logical incoherence. The only sound alternatives within the social sciences, then, would have to be theories making no appeal to meaning or intentionality at all: purely behavioristic or purely physiological theories. This claim strikes a familiar note: many psychologists and brain scientists have expressed great skepticism about the utility or permissibility of "mentalistic" formulations in their fields--while others of course have held them to be indispensable. The philosophical analysis of intentionality yields a clear logical characterization of this fundamental theoretical division in the social sciences and biology: "mentalistic" theories are all and only those making ineliminable use of intentional idioms, and hence inheriting the logical problems of construing those idioms coherently.

Dispensing with intentional theories is not an attractive option, however, for the abstemious behaviorisms and physiological theories so far proposed have signally failed to exhibit the predictive and explanatory power needed to account for the intelligent activities of human beings and other animals. A close examination of the most powerful of these theories reveals intentional idioms inexorably creeping in--for instance in the definition of the stimulus as the "perceived" stimulus and the response as the "intended"

effect, or in the reliance on the imputation of "information-bearing" properties to physiological constructs. Moreover the apparent soundness of information-processing theories, and their utility in practice, has strengthened the conviction that somehow we must be able to make sense of the ineliminable intentional formulations they contain without compromising thoroughgoing materialism.

So we are left with the problem of providing an account of the intentionality of mental states or events which deals with the peculiarities of this feature in a naturalistic way. The most promising approach, found in various versions throughout artificial intelligence and among certain schools of thought in Anglo-American philosophy, is to interpret mental states and processes as representations (perhaps even in a "language of thought"), and to deny the powerful pretheoretical intuition that the intentionality of such mental representations is somehow original or intrinsic--unlike the derivative intentionality of public representations such as sentences, pictures and maps. The familiar and intuitive distinction to be denied is discussed by Haugeland. Our artifacts

Consider an encyclopedia. It has derived intentionality. It contains information about thousands of things in the world, but only insofar as it is a device designed and intended for our use. Suppose we "automate" our encyclopedia, putting all its data into a computer, and turning its index into the basis for an elaborate question-answering system. No longer do we have to look up material in the volumes; we simply type in questions and receive answers. It might seem to naive users as if they were communicating with another person, another entity endowed with original intentionality, but we would know better. A question-answering system is still just a tool, and whatever meaning or aboutness we reside in it is just a byproduct of our practices in using the device to serve our own goals. It has no goals of its own, except for the artificial and derived goal of "understanding" and "answering" our questions correctly.

But suppose we endow our computer with somewhat more autonomous, somewhat less slavish goals. For instance, a chess playing computer has the (artificial, derived) goal of defeating its human opponent, of concealing what it "knows" from us, of tricking us perhaps. But still, surely, it is only our tool or toy, and although many of its internal states have a sort of aboutness or intentionality--e.g., there are states that represent (and hence are about) the current board positions, and processes that investigate (and hence are about) various possible continuations of the game--this is just derived intentionality, not original intentionality.

This persuasive theme (it is not really an argument) has convinced more than a few thinkers that no artifact could have the sort of intentionality we have. Any computer program, any robot we might design and build, no matter how strong the illusion we may create that it has become a genuine agent, could never be a truly autonomous thinker with the same sort of original intentionality we enjoy. The issue can be dramatized with an example.

Consider a standard soft-drink vending machine, designed and built in the United States, and equipped with a transducer device for accepting and rejecting U.S. quarters. Let's call such a device a two-bitser. Normally, when a quarter is inserted into a two-bitser, the two-bitser goes into a state, call it Q, which "means" (note the scare-quotes) "I perceive/accept a genuine U.S. quarter now." Such two-bitseres are quite clever and sophisticated, but hardly foolproof. They do "make mistakes" (more scare-quotes). That is, unmetaphorically, sometimes they go into state Q when a slug or other foreign object is inserted in them, and sometimes they reject perfectly legal quarters--they fail to go into state Q when they are supposed to. No doubt there are detectable patterns in the cases of "misperception". No doubt at least some of the cases of "misidentification" could be predicted by someone with enough knowledge of the relevant laws of physics and design parameters of the two-bitser's transducing machinery, so that it would be just as much a matter of physical law that objects of kind K would put the device into state Q as that quarters would. Objects of kind K would be good "slugs"--reliably "fooling" the transducer.

If objects of kind K became more common in the two-bitser's normal environment, we could expect the owners and designers of two-bitseres to develop more advanced and sensitive transducers that would reliably discriminate between genuine U.S. quarters and slugs of kind K. Of course trickier counterfeits might then make their appearance, requiring further advances in the detecting transducers, and at some

point such escalation of engineering would reach diminishing returns, for there is no such thing as a foolproof mechanism. In the meantime, the engineers and users are wise to make do with standard, rudimentary two-bitser, since it is not cost-effective to protect oneself against negligible abuses.

The only thing that makes the device a quarter-detector rather than a slug-detector or a quarter-or-slug-detector is the shared intention of the device's designers, builders, owners, users. It is only in the environment or context of those users and their intentions that we can single out some of the occasions of state Q as "veridical" and others as "mistaken". It is only relative to that context of intentions that we could justify calling the device a two-bitser in the first place.

This is, apparently, a textbook case of derived intentionality, laid bare. And so of course it embarrasses no one to admit that a particular two-bitser, straight from the American factory and with "Model A Two-Bitser" stamped right on it, might be installed on a Panamanian soft-drink machine, where it proceeded to earn its keep as an acceptor and rejecter of quarter-balboas, legal tender in Panama, and easily distinguished from U.S. quarters by the design and writing stamped on them, but not by their weight, thickness, diameter or material composition.

(I'm not making this up. I have it on excellent authority-- Albert Erlar of the Flying Eagle Shoppe, Rare Coins--that Panamanian quarter-balboas minted between 1966 and 1984 are indistinguishable from U.S. quarters by standard vending machines. Small wonder, since they are struck from U. S. quarter stock in American mints. And--to satisfy the curious, although it is strictly irrelevant to the example--the current official exchange rate for the quarter-balboa is indeed \$.25!)

Such a two-bitser, whisked off to Panama (the poor man's Twin-Earth), would still normally go into a certain physical state--the state with the physical features by which we used to identify state Q--whenever a U.S. quarter or an object of kind K or a Panamanian quarter-balboa is inserted in it, but now a different set of such occasions count as the mistakes. In the new environment, U.S. quarters count as slugs, as inducers of error, misperception, misrepresentation, just as much as objects of kind K do. After all, back in the U.S. a Panamanian quarter-balboa is a kind of slug.

Once our two-bitser is resident in Panama, should we say that the state we used to call Q still occurs? The physical state in which the device "accepts" coins still occurs, but should we now say that we should identify it as "realizing" a new state, QB, instead? Well, there is considerable freedom--not to say boredom--about what we should say, since after all a two-bitser is just an artifact, and talking about its perceptions and misperceptions, its veridical and non-veridical states--its intentionality, in short--is "just metaphor". The two-bitser's internal state, call it what you like, doesn't really (originally, intrinsically) mean either "U.S. quarter here now" or "Panamanian quarter-balboa here now". It doesn't really mean anything. So the believers in original intentionality (Fodor, Searle, Dretske, Burge, and Kripke, inter alia) would insist.

The two-bitser was originally designed to be a detector of U.S. quarters. That was its "proper function" (Millikan, 1984), and, quite literally, its *raison d'être*. No one would have bothered bringing it into existence had not this purpose occurred to them. And given that this historical fact about its origin licenses a certain way of speaking, such a device may be primarily or originally characterized as a two-bitser, a thing whose function is to detect quarters, so that relative to that function we can identify both its veridical states and its errors.

This would not prevent a two-bitser from being wrested from its home niche and pressed into service with a new purpose-- whatever new purpose the laws of physics certify it would reliably serve--as a K-detector, a quarter-balboa detector, a doorstop, a deadly weapon. In its new role there might be a brief period of confusion or indeterminacy. How long a track record must something accumulate before it is no longer a two-bitser, but rather a quarter-balboa-detector (a q-balber)--or a doorstop, or a deadly weapon? On its very debut as a q-balber, after ten years of faithful service as a two-bitser, is its state already a veridical detection of a quarter-balboa, or might there be a sort of force-of-habit error of nostalgia, a mistaken identification of a quarter-balboa as a U.S. quarter?

As described, the two-bitser differs strikingly from us in that it has no provision for memory of its past experiences--or even "memory" (in scare quotes) for its past "experiences". But the latter, at least, could easily be provided, if it was thought to make a difference. To start with the simplest inroad into this topic, suppose the two-bitser (to refer to it by the name of its original baptism) is equipped with a counter, which after ten years of service stands at 1,435,792. Suppose it is not reset to zero during its flight to Panama, so that on its debut there the counter turns over to 1,436,793. Does this tip the balance in favor of the claim that it has not yet switched to the task of correctly identifying quarter-balboas? Would variations and complications on this theme drive your intuitions in different directions?

We can assure ourselves that nothing intrinsic about the two-bitser considered narrowly all by itself and independently of its prior history would distinguish it from a genuine q-balber, made to order on commission from the Panamanian government. Still, given its ancestry, is there not a problem about its function, its purpose, its meaning, on this first occasion when it goes into the state we are tempted to call Q? Is this a case of going into state Q (meaning "U.S. quarter here now") or state QB (meaning "Panamanian quarter-balboa here now")? I would say, along with Millikan (1984), that whether its Panamanian debut counts as going into state Q or state QB depends on whether, in its new niche, it was selected for its capacity to detect quarter-balboas--literally selected, e.g., by the holder of the Panamanian Pepsi-Cola franchise. If it was so selected, then even though its new proprietors might have forgotten to reset its counter, its first "perceptual" act would count as the correct identification of a q-balber, for that is what it would now be for. (It would have acquired quarter-balboa detection as its proper function.) If, on the other hand, the two-bitser was sent to Panama by mistake, or if it arrived by sheer coincidence, its debut would mean nothing, though its utility might soon--immediately--be recognized and esteemed by the relevant authorities (those who could press it into service in a new role), and thereupon its subsequent states would count as tokens of QB.

Presumably Fodor et al. would be content to let me say this, since, after all, the two-bitser is just an artifact. It has no intrinsic, original intentionality, so there is no "deeper" fact of the matter we might try to uncover. This is just a pragmatic matter of how best to talk, when talking metaphorically and anthropomorphically about the states of the device.

But we part company when I claim to apply precisely the same morals, the same pragmatic rules of interpretation, to the human case. In the case of human beings (at least), Fodor and company are sure that such deeper facts do exist--even if we cannot always find them. That is, they suppose that, independently of the power of any observer or interpreter to discover it, there is always a fact of the matter about what a person (or a person's mental state) really means. Now we might call their shared belief a belief in intrinsic intentionality, or perhaps even objective or real intentionality. There are differences among them about how to characterize, and name, this property of human minds, which I will continue to call original intentionality, but they all agree that minds are unlike the two-bitser in this regard.

I part company with these others, because although they might agree with me (and Millikan) about what one should say in the case of the transported two-bitser, they say that we human beings are not just fancier, more sophisticated two-bitsters. When we say that we go into the state of believing that we are perceiving a U.S. quarter (or some genuine water as opposed to XYZ, or a genuine twinge of arthritis) this is no metaphor, no mere manner of speaking. A parallel example will sharpen the disagreement:

Suppose some human being, Jones, looks out the window and thereupon goes into the state of thinking he sees a horse (cf. Fodor 1987). There may or may not be a horse out there for him to see, but the fact that he is in the mental state of thinking he sees a horse is not just a matter of interpretation (these others say). Suppose the planet Twin-Earth were just like Earth, save for having schmorses where we have horses. (Schmorses look for all the world like horses, and are well-nigh indistinguishable from horses by all but trained biologists with special apparatus, but they aren't horses, any more than dolphins are fish.) If we whisk Jones off to Twin-Earth, land of the the schmorses, and confront him in the relevant way with a schmorse, then either he really is, still, provoked into the state of believing he sees a horse (a mistaken, non-veridical belief) or he is provoked by that schmorse into believing, for the first time (and veridically),

that he is seeing a schmorse. (For the sake of the example, let us suppose that Twin-Earthians call schmorses horses (chevaux, Pferde, etc.) so that what Jones or a native Twin-Earthian says to himself--or others--counts for nothing.) However hard it may be to determine exactly which state he is in, he is really in one or the other (or perhaps he really is in neither, so violently have we assaulted his cognitive system). Anyone who finds this intuition irresistible believes in original intentionality, and has some distinguished company: Fodor, Searle, Dretske, Burge, and Kripke, but also Chisholm 1956, 1957, Nagel 1979, 1986 and Popper and Eccles 1977). Anyone who finds this intuition dubious if not downright dismissible can join me, the Churchlands, Davidson, Haugeland, Millikan, Rorty, Stalnaker, and our distinguished predecessors, Quine and Sellars, in the other corner (along with Douglas Hofstadter, Marvin Minsky and almost everyone else in AI).

There, then, is a fairly major disagreement. Who is right? I have argued in my new book, *The Intentional Stance*, (Dennett, 1987) that clinging to the doctrine of original intentionality is the primary source of perplexity in contemporary Anglo-American philosophy of mind, but the arguments for that broad indictment cannot be presented in the short compass of this introductory talk. I can, however, give a summary expression of the main positive point of my theory: the intentionality of our mental states and processes is derived in just the same way as that of our books and maps (and the inner states of our robots). Suppose you have composed a shopping list, on a piece of paper, to guide your shopping behavior. The marks on the piece of paper have derived intentionality, of course, but if you forgo the shopping list and just remember the wanted items in your head, whatever it is that "stores" or "represents" the items to be purchased in your brain has exactly the same status as the trails of ink on the paper. There is no more real, or intrinsic, or original intentionality than that.