When does the intentional stance work?

Daniel C. Dennett

Philosophy Department, Tufts University, Medford, Mass. 02155

Ben-Zeev raises the issue of referential opacity – the failure of substitution of codesignative terms *salva veritate* – and claims that the application of the intentional stance to thermostats (an example of mine) doesn't exhibit it. This shows, he thinks, that the use of the intentional stance in application to such things as thermostats is fundamentally different from its use in application to things – people and at least some creatures, presumably – that have genuine intentional states. This would leave the concept of an intentional state in need of some other account than mine, and Ben-Zeev proposes that "the reference to the intentional stance . . . is meaningful only in regard to agents with some cognitive capacities. The greater the complexity of those capacities, the more relevant and suitable are the considerations concerning the intentional stance."

But this is not an alternative account; it is my account obscured and weakened by its reliance on an unanalyzed notion of an agent with a cognitive capacity. **Ben-Zeev's** point about substitution failure is important, but slightly misses the mark. Finding a case where substitution works is not the issue; after all, in many cases of attributing intentional states to human beings, one may substitute codesignative terms *salva veritate*. The proper question is: Does explanation of the behavior of thermostats (or bees or birds) *ever* require attributions that resist substitution? And here the answer is interestingly indirect. Thermostats are *virtually* "oblivious" to substitution, but not quite, and we can easily imagine more

sensitive (wily) thermostats, attributions to which were clearly sensitive to substitution. But thermostat designers have not felt the need for such sophisticated devices for the most part. Similarly, consider the bee, which surely does not need to recognize or distinguish her oleic-acid-exuding dead sister qua health hazard, or even qua corpse. The bee has a very minimal "cognitive capacity" – to use Ben-Zeev's term. But when we go to explain why this phenomenon exists in nature, why bees should have this proclivity built in, our explanation will single out the dead bee under the marked description; it was qua *health hazard*, and not qua anything else, that dead bees were "recognized" by the evolutionary process itself (Mother Nature). The rationale of the behavior (if not the individual bee's rationale, then a free-floating rationale) is nevertheless expressible only in the referentially opaque language of intentional explanation. So whatever a "cognitive capacity" is, if its presence is marked by an appeal to referentially opaque explication, then natural selection is itself an "agent with cognitive capacities."

As **Bogdan** says, "the adaptive development we try to understand is, in some sense, sensitive to, and has a raison d'être in, aspects that transcend the internal territory of hardware and design."

Ben-Zeev's assertions that my view is "close to" Wittgenstein, and "similar, in some respects," to Aristotle, must be true, of course. With a sufficiently relaxed standard of similarity, affinities can be found between the views of almost any two philosophers on any subject – Hegel and Aristotle, say, or Sartre and Quine. (I have long yearned to write the rather obvious paper entitled "How Sartre's 'transparency' is just Quine's 'opacity.'") In this instance I do not see anything particularly striking or useful or worth quarreling about in the comparisons, so I will resist the temptation to "compare and contrast" as they say in final exam essay questions.

Bogdan makes the point that I think gets obscured by **Ben-Zeev's** proposal to define intentionality in terms of cognitive states (and not vice versa). What makes something a (central) cognitive capacity or contentful state is that it is "unexplainable, . . . because underdetermined, by hardware and design laws." That is not to say that it is not *in principle* fully predictable by, say, a Laplacean ominiscient scientist working with nothing but "hardware and design laws," but that any such nonintentional (mechanistic, atomistic, local) explanation would miss something important: that peculiar relatedness to remote conditions, real or implied, that is most familiarly recognized as *aboutness* – what philosophers call intentionality. It almost looks like "action at a distance," but of course it is not. The indirect bearing of the Eiffel Tower on my thought about the Eiffel Tower, like the indirect bearing of the toxicity of those ancestral bee corpses that weren't removed from their hives on the current behavior vis-à-vis corpses of today's bees, is not the sort of relation that can be illuminated by a mechanistic, nonintentional account, however voluminous.

Bogdan sees as an implication of this that "the intentional stance cannot discharge its assigned heuristic task precisely when most needed, that is, when there is no obvious (I mean: internal) way from form to content." That is, for the best truly "central" cases of content,

where the intentional stance is our only grip on the phenomenon, we cannot expect the intentional stance to point out the path to its own elimination in favor of design stance accounts. I guess that is true, for both psychology and evolutionary theory, and it amounts to a very mild sort of "irreducibility" claim. Not that mind is irreducible to brain, or that intentionality is inexplicable in mechanical terms, or that adaptation cannot be the result of (nothing but) evolution by natural selection (and genetic drift and other clearly mechanical processes), but just that the only sense we will ever be able to make of the play and interaction of "central" intentional states will be the explanations we make from the intentional stance. Other accounts may be true, and predictive, but won't explain everything that needs explaining.

Dahlbom offers six numbered points, to which I will respond in turn.

1. His account of the current trend away from Enlightenment values toward a Romantic vision of science provides a novel, valuable and, I believe, largely correct perspective on contemporary controversies. He is right that I am unfashionably bucking the Romantic trend, but then I have long been branded a "verificationist," "reductionist" opponent of Chomsky and Fodor, for instance, so my sympathies should not surprise Dahlbom. But why, he asks, do I willfully place Gould and Lewontin, arch-Romantics, with Skinner, the embodiment of the Enlightenment creed? Because I realized that their arguments – not just the style but the substance – were the same. The joint theme is that both the intentional stance and adaptionism make a "question-begging" appeal to optimality when the proper way for science to proceed here is to do unadorned mechanical history of actual selection. Just as Quine and Skinner abjure borrowing intelligence (intentionality), Gould and Lewontin abjure borrowing optimality of design. And since rationality *is* optimality of cognitive design, one can look at the intentional stance as just a special case of adaptationist thinking. I lump Gould and Lewontin with Skinner because in spite of their ideological differences elsewhere, here they are saying the same thing about the same issue, and they have all overstated their case.

Dahlbom's point about paying lip service is important. The truth about *which* ideas to emphasize *when* in science would probably lie in the boring middle ground, were it not psychologically important to researchers and theorists to have a somewhat radical and intolerant conviction about how everything fits (will jolly well be made to fit) into one elegant vision. So we can expect the adaptationists to pay mere lip service to "constraints," just as their opponents pay mere lip service to adaptation, and with any luck, like Jack and Mrs. Sprat they will lick the platter clean – which is probably more than we could hope for from a herd of mealy-mouthed pluralists.

2. Yes, what **Dahlbom** calls the problem-solving approach and intentional system theory are one and the same thing under different names. I am not claiming to have a whole new way of doing science to offer to ethologists; I am just pointing out, as philosophers are wont to do, the conceptual obligations and privileges of a way of doing business that is already familiar, if often undertaken under false or strained pretenses. And surely Dahlbom is right that the higher-order intentional

attributions will find their main utility among the higher animals if anywhere, but one must remember that the higher-order attributions are also useful in analyzing relationships that are *not* "appreciated" by the individual organisms – the low-nesting birds, or the bees, for instance – but only by the evolutionary process that created and preserves the regularity in those relationships.

3. The intentional stance will fail, **Dahlbom** thinks, when common ground is lacking – when the target creatures are not enough like (human) *persons*. I am unconvinced (cf. Stich 1981, and my reply to Stich, Dennett 1981, and Stich 1983). It is no doubt a more difficult exercise of the imagination to "think like a Martian" – or a beaver or a coyote – but so what? And I do not see that it is made more difficult to think of superpersonal organizations when one thinks of individuals as intentional systems. A bureaucrat *is* an intentional system; to think of someone as a bureaucrat is not to demote him but to identify a type of intentional system to which the individual belongs. Bureaucrats – when they act true to form – do exactly what it is rational to do if one is in the bureaucrat's particular (and ubiquitous) predicament. If you or I were stuck being bureaucrats (if the options of rebellion or obstreperousness were particularly unattractive – thanks to our having to see three children through college, say) we would be stuck believing as bureaucrats do, and behaving as bureaucrats do, for under those grubby circumstances that is what it is rational to do. Some of the distinctive features of Weber's ideal types may be merely habitual or traditional – manifesting a sort of drag that creates a gap between actual practice and ideally rational practice – but the core of every such system is a rational practical reasoner. So I don't see the problem Dahlbom poses as looming large at all.

4. **Dahlbom** nicely describes the way "the Romantic theory [of "prejudiced" trial and error] corrects the gradualism and adaptationism of the Enlightenment by introducing saltational and structural elements." A good thing, but haven't we already seen that this is a matter of emphasis – like the grain problem that bedevils (and vitiates) the debate over punctuated equilibrium? What looks gradual from a bird's-eye view looks like fits and starts midst stasis from close up. Of course there has to be some biasing structure to constrain the trial and error process. Dahlbom suggests that a "move whereby such 'constraints' take on a central role" would "call into question the basis" of my theory. That is true; any theory of learning or development or evolution that gives the *central* role to what I call the constraints will be strongly opposed to mine, and to adaptationism, but such a theory will have the huge task of explaining (and not merely paying lip service to) the excellence of design and aptness of thought so normal in our world – without ever appealing to adaptationist trends. I am less convinced than Dahlbom that Gould and Lewontin (or Kimura or anyone else in biology) are pointing the way to such a theory. Simply describing – let alone explaining – these phenomena has traditionally depended on intentional language, with its assumptions of rationality or optimality built in. Skinner learned, to his discomfort, that he simply couldn't describe the domain of his field without lapsing into the suspect "mentalistic" vernacular with its

tacit appeals to rationality. As the antiadaptationists are learning, it is equally quixotic to set oneself the purist goal of an account of evolution that is shriven of all Panglossian formulae.

Dahlbom's points 5 and 6 develop themes that require more thought from me, but so far as I can see I can agree with him. His reminder of the fate of general systems theory sends a salutary chill down my spine. Forewarned is forearmed.

My recommending a postbehaviorist vision to ethologists has provoked **Gray,** who sees me "using an outdated cast of characters," focusing on Skinner and ignoring such latter-day behaviorists as Hebb, Hull, Tolman, Bindra, and Kamin. Is Kamin really a behaviorist? Are Carew, Walters, and Kandel? I think Gray might take his own advice: "Perhaps it is time to banish 'behaviorist' and related terms" – if we are intent on referring to work on behavior that is only indirectly linked to the tradition of Watson, Thorndike, and Skinner (and Hull, Tolman, etc.).

What is **Gray**'s point? I guess it is that *non*cognitivist, (neo-?)behaviorist psychologists have something to offer ethologists that I have overlooked. I don't agree. It seems to me that Hebb, Bindra, and others managed at best to demonstrate how difficult and barren such approaches were, even when pursued with energy and brilliance. (Gray says I "might also have taken Bindra more seriously." I have, in Dennett 1978b, a commentary on Bindra's article in *BBS* which I am content to let serve as my summary of what was wrong not only with Bindra's approach, but with the other late behaviorists' attempts at theory-construction.) Nor do I think those who really are making progress on the "nature of the mechanisms" (Carew, Walters, and Kandel are a fine example) have much to say to ethologists yet, since the transfer from *Aplysia* or insects to birds and mammals is such a long journey that most of the good baggage must be abandoned along the way. [See also Hoyle: "The Scope of Neuroethology" *BBS* 7(3) 1984.]

Gray says that "we have found out that the nervous system is not made of components available in Radio Shack." This is presumably a Bronx cheer directed at AI, but it misses its mark and strikes some caricature inhabiting Gray's imagination. I cannot think of a single proponent of AI, no matter how fanatic or radical, whose views are challenged by this remark. And when Gray observes that "Dennett is dejected that all romance will be taken out of life" he convicts himself of a rather heroic misreading of my playfully labeled scale from romantic to killjoy. True-blue Behaviorists are only first-order intentional systems; they have beliefs and desires (we all do – behaviorism is false), but they don't believe that they or anyone else does. And so one telling symptom of behaviorism, not surprisingly, is obliviousness to humor.

References

Bateson, P. P. G. (1983a) Rules for changing the rules. In: *Evolution from molecules to men*, ed. D. S. Bendall. Cambridge University Press. [BD]
(1983b) Uncritical periods and insensitive sociobiology. *Behavioral and Brain Sciences* 6:102–3. [BD]
Bennett, J. (1983) Cognitive ethology: Theory or poetry? *Behavioral and Brain Sciences* 6:356–58. [BD]

Ben-Zeev, A. (1983) Toward a different approach to perception. *International Philosophical Quarterly* 23:45–64. [AB-Z]

Bindra, D. (1976) *A theory of intelligent behavior.* Wiley. [TG]

(1978) How adaptive behavior is produced: A perceptual-motivational alternative to response-reinforcement. *Behavioral and Brain Sciences* 1:41–91 [TG]

Bogdan, R. J. (1983) Fodor's representations. *Cognition and Brain Theory* 6:237–49. [RJB]

Carew, T., Walters, E. & Kandel, E. (1981) Associative learning in Aplysia: Cellular correlates supporting a conditioned fear hypothesis. *Science* 211:501–4. [TG]

Chomsky, N. (1959) A review of Skinner's *Verbal Behavior. Language* 35:26–58. [BD]

Churchland, P. S. (1983) Dennett's instrumentalism: A frog at the bottom of the mug. *Behavioral and Brain Sciences* 6:358–59. [BD]

Danto, A. C. (1983) Science as an intentional system. *Behavioral and Brain Sciences* 6:359–60. [BD]

Dennett, D.C. (1969) *Content and consciousness.* Routledge & Kegan Paul. [BD]

(1971) Intentional systems. *Journal of Philosophy* 68:87–106. Reprinted in Dennett *Brainstorms*, 1978a. Bradford/MIT Press. [BD]

(1975) Why the law of effect will not go away. *Journal of the Theory of Social Behavior* 5:169–87. Reprinted in Dennett, *Brainstorms*, 1978a. Bradford/MIT Press. [BD]

(1978a) *Brainstorms.* Bradford/MIT Press. [BD, TG]

(1978b) Requisition for a Pexgo. *Behavioral and Brain Sciences* 1:56–7. [DCD]

(1981a) Three kinds of intentional psychology. In: *Reductionism, time, and reality*, ed. R. Healey. Cambridge University Press. [BD]

(1981b) Making sense of ourselves. *Philosophical Topics* 12: 63–82, reprinted in *Mind, brain, and function*, ed. J. Biro & R. Shahan, 1982. University of Oklahoma Press. [DCD]

(1983) Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences* 6:343–90. [AB-Z, RJB, BD, TG]

(1984) Cognitive wheels: The frame problem of AI. In: *Minds, machines and evolution*, ed. C. Hookway. Cambridge University Press. [BD]

Dretske, F. I. (1981) *Knowledge and the flow of information.* Bradford/MIT Press. [BD]

(1983) Précis of *Knowledge and the flow of information. Behavioral and Brain Sciences* 6:55–90. [BD]

Eldredge, N. & Gould, S. J. (1972) Punctuated equilibria: An alternative to phyletic gradualism. In: *Models in paleobiology*, ed. T. J. M. Schopf. Freeman, Cooper and Co. [BD]

Farley, J. & Alkon, D. (1980) Neural organization predicts stimulus specificity

for a retained associative behavioral change. *Science* 210:1373–74. [TG]

Ginzburg, C. (1980) Morelli, Freud and Sherlock Holmes: Clues and scientific method. *History Workshop* 9:5–36. [BD]

Gould, S. J. & Eldredge, N. (1977) Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* 3:115–51. [BD]

Gould, S. J. & Lewontin, R. C. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London* B205:581–98. [BD, TG]

Hofstadter, D. (1979) *Gödel, Escher and Bach: An eternal golden braid.* Basic Books. [TG]

Kamin, L. (1969) Predictability, surprise, attention, and conditioning. In: *Punishment and aversive behavior,* ed. B. A. Campbell & R. M. Church. Appleton-Century-Crofts. [TG]

Kuhn, T. S. (1962) *The structure of scientific revolutions.* University of Chicago Press. [BD]

Lévi-Strauss, C. (1955) *Tristes tropiques.* Plon. English trans. *A world on the wane,* by John Russell. Hutchinson, 1961. [BD]

(1966) *The savage mind.* Weidenfeld and Nicolson. [BD]

(1969) *The elementary structures of kinship.* Beacon Press. [BD]

Lewontin, R. (1981a) The inferiority complex. *New York Review of Books* 28:12–16. [TG]

(1981b) Sleight of hand. *Sciences* 21:23–26. [TG]

Lloyd, D. (1983) The scope and ingenuity of evolutionary systems. *Behavioral and Brain Sciences* 6:368–69. [AB-Z]

Lumsden, C. J. & Wilson, E. O. (1981) *Genes, mind, and culture.* Harvard University Press. [TG]

Neisser, U. (1967) *Cognitive psychology.* Appleton-Century-Crofts. [BD]

Rosenberg, A. (1983) Content and consciousness versus the intentional stance. *Behavioral and Brain Sciences* 6:375–76. [BD]

Skinner, B. F. (1983) A better way to deal with selection. *Behavioral and Brain Sciences* 6:377–78. [BD]

Stich, S. (1981) Dennett on intentional systems. *Philosophical Topics* 12:39–62, reprinted in *Mind, brain, and function,* ed. J. Biro & R. Shahan, 1982. University of Oklahoma Press. [DCD]

Stich, S. (1983) *From folk psychology to cognitive science.* Bradford/MIT Press. [DCD]

van den Berghe, P. L. (1983) Human inbreeding avoidance: Culture in nature. *Behavioral and Brain Sciences* 6:91–123. [BD]

Watson, J. & Crick, F. (1953) Molecular structure of nucleic acids: A structure for deoxiribose nucleic acid. *Nature* 171:737–38. [TG]

Weber, M. (1964) *The theory of social and economic organization.* Free Press. Reprinted in Weber (1968). [BD]

(1968) *Economy and society.* Bedminster Press. [BD]

Wittgenstein, L. (1958) *Philosophical investigations.* Macmillan. [AB-Z]