PHIL 5983: Rationality Seminar

University of Arkansas, Fall 2004

Topic: Rationality and Meaning Restrictions on Belief
Readings: Dennett's "Intentional Systems" and "Brain Writing and Mind Reading", and Davidson's "Thought and Talk"

*Here are some links to class notes from a seminar I taught last fall on mental causation. These links are specific to Daniel Dennett's theory of intentionality/rationality, and the Dennett material they cover is more recent (from his book *The Intentional Stance*) than what we are reading in this course (which is from his book *Brainstorms*).

http://comp.uark.edu/~efunkho/notes2.html; http://comp.uark.edu/~efunkho/notes3.html

For general information on belief, check out:
http://www.faculty.ucr.edu/~eschwitz/SchwitzPapers/BeliefEntry030927.html

"Intentional Systems"
Intentional system: "a system whose behavior can be—at least sometimes—explained and predicted by relying on ascriptions to the system of beliefs an desires (and hopes, fears, intentions, hunches ….)" (3)

Intentionality: those contexts in which "substitution of codesignative terms do not preserve truth" or in which "the "objects" of the idiom are not capturable in the usual way by quantifiers." (3)

*Here are some notes I have online, again from previous courses, on *opaque contexts*, which is what Dennett's definition of 'intentionality' refers to:
http://comp.uark.edu/~efunkho/lecture11.html
http://comp.uark.edu/~efunkho/lang4.html
http://comp.uark.edu/~efunkho/lang5.html

I.
*3 stances, and the chess-playing computer

--Design stance: "The essential feature of the design stance is that we make predictions solely from knowledge or assumptions about the system's functional design, irrespective of the physical constitution or condition of the innards of the particular object." (4)

--Physical stance: "From this stance our predictions are based on the actual physical state of the particular object, and are worked out by applying whatever knowledge we have of the laws of nature." (4)

--Malfunctions can be predicted at this level, but not from the design level.

--Intentional stance: Predicting the behavior of a system by assuming it is *rational*, and has certain goals and beliefs.

> --"First, rationality here so far means nothing more than optimal design relative to a goal or optimally weighted hierarchy of goals (checkmate, winning pieces, defense, etc., in the case of chess) and a set of constraints (the rules and starting position). (5-6)
> "It is a small step to calling the information possessed the computer's *beliefs*, its goals and subgoals its *desires*." (6)

*_This_ version (1971) of Dennett does not endorse a Realism about belief:

> "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." (7)

*Dennett stresses the *pragmatic* benefits of adopting the intentional stance. (7) It works best: "Whenever we have reason to suppose the assumption of optimal design is warranted, and doubt the practicality of prediction from the design or physical stance." (8)

II.

*More on the assumption of rationality:

> "The presumption of rationality is so strongly entrenched in our inference habits that when our predictions prove false, we at first cast about for adjustments in the information-procession conditions (he must not have heard, he must not know English, he must not have seen *x*, been aware that *y*, etc.) or goal weightings, before questioning the rationality of the system as a whole." (9-10)

> --Importantly, once we abandon the assumption of rationality, we abandon the intentional stance:
> "In extreme cases personalities may prove to be so unpredictable from the intentional stance that we abandon it, and if we have accumulated a lot of evidence in the meanwhile about the nature of response patterns in the individual, we may find that a species of design stance can be effectively adopted. This is the fundamentally different attitude we occasionally adopt toward the insane. To watch an asylum attendant manipulate an obsessively countersuggestive patient, for instance, is to watch something radically unlike normal interpersonal relations." (10)

> --See pp. 10-11 for Dennett's discussion of lower animals and their believing (or at least following) the truths of logic. In particular, what should we make of the following passage?

"Surely our mouse follows or believes in *modus ponens*, for we ascribed to it the beliefs: (a) *there is a cat to the left*, and (b) *if there is a cat to the left, I had better not go left*, and our prediction relied on the mouse's ability to get to the conclusion." (11)

--Also, note the discussion on p. 11 about the trade-off between inferences rules and beliefs in logical truths.

*Pp. 13-16 is a critique of Behaviorism.

III.
*Claim: "A species might 'experiment' by mutation in any number of inefficacious systems, but none of these systems would deserve to be called belief systems precisely because of their defects, their nonrationality, and hence a false belief system is a conceptual impossibility." (17)

*Compare/contrast this with Davidson: "The faculty of communication would not gain ground in evolution unless it was by and large the faculty of transmitting true beliefs, which means only: the faculty of altering other members of the species in the direction of more optimal design." (18)

*Note Dennett's two preconditions for belief ascription (bottom p. 18), and his corresponding norms for belief and belief avowals.

--The essential normative component of belief:
"We get around the 'privacy' of beliefs and desires by recognizing that in general anyone's beliefs and desires must be those he 'ought to have' given the circumstances." (19)

--The potential conflicts between the above two norms (19-21). We can see how Dennett will react to cases of apparent delusional belief:
"What better source could there be of a system's beliefs than its avowals? Conflict arises, however, whenever a person falls short of perfect rationality, and avows beliefs that either are strongly disconfirmed by the available empirical evidence or are self-contradictory or contradict other avowals he has made. If we lean on the myth that a man is perfectly rational, we must find his avowals less than authoritative: "You *can't* mean—understand—what you're saying!"; if we lean on his 'right' as a speaking intentional system to have his word accepted, we grant him an irrational set of beliefs. Neither position provides a stable resting place; for, as we saw earlier, intentional explanation and prediction cannot be accommodated either to breakdown or to less than optimal design, so there is no coherent intentional description of such an impasse." (20)

"Brain Writing and Mind Reading"
Q: "Is it *in principle* possible that brain scientists might one day know enough about the workings of our brains to be able to 'crack the cerebral code' and read our minds?" (39)

--The story of Sam the art critic. (39-40)

manifest (or explicit) beliefs: Beliefs that we have thought about (e.g., that 2+2=4).

covert (or implicit) beliefs: Beliefs that we have never thought about (e.g., that a pink thimble could easily fit under a cowboy hat).

*Harman's path to the brain-writing hypothesis: inner representations are required to explain behavior, and these systems of representations are inner languages. (40-41)

*Dennett's 6 requirements for genuine brain-writing (42-44)

1. Generative grammar.
2. Syntax is physically realized.
3. The physical differences amongst tokens must, as a practical matter, be detectable. (Legibility requirement)
4. The language in the brain must match up with most of the manifest beliefs, as well as contain a lot of other truths.
5. The language in the brain must be connected to whatever causes action.
6. The sentences of the language must be consistent (for the most part).

--The consistency requirement is obviously of interest to us. The case of Tom, and the attempt to insert in him the belief "I have an older brother living in Cleveland.", is of particular interest. (44)

*Hypothesis: Not all of our beliefs are explicitly stored. Rather, we have an "extrapolator-deducer mechanism" to generate beliefs from our manifest ones. (45) This, it seems, would be the mechanism behind our dynamic rationality. (See, in particular, the paragraph that starts on the bottom of p. 45.)

*Here's a nice passage to think about, regarding the connection between language and belief: "The effort of retrieval is often an effort to formulate a sentence that is an approximation of a belief, and we are often distressed by the hard edge of determinacy our verbal output substitutes for the fuzziness of our convictions." (48)

*The distinction between belief and judgment—pp. 48-49.

*One major conclusion: The brain represents, but not necessarily in sentences (nor in a system). (49)

"Thought and Talk"

***Check out the Stanford Encyclopedia of Philosophy entry on Donald Davidson: http://plato.stanford.edu/entries/davidson/

*Q: Can there be inexpressible thoughts, or thought without speech? (155)

--Davidson thinks the correct answer to the latter question is "no", and his goal here is to argue for this.

--By 'thought', Davidson primarily (if not exclusively) means the propositional attitudes.

--Belief is the most basic of propositional attitudes. (156-157)

--Davidson holds that *thinkers* must be *interpreters*. "The chief thesis of this paper is that a creature cannot have thoughts unless it is an interpreter of the speech of another." (157)

*What is an interpreter, or what is required of an interpreter?

--"An interpreter knows the conditions under which utterances of sentences are true, and often knows that if certain sentences are true, others must be." (158)

*Davidson introduces folk psychological (belief-desire, teleological) explanations of action on pp. 158-159.

--We see the constitutive role that reason plays in this enterprise, in Davidson's p. 159 passages. Davidson primarily has in mind *practical* rationality, at this point. (E.g., "…the descriptions we provide of desire and belief must, in teleological explanation, exhibit the rationality of the action in the light of the content of the belief and the object of the desire." (159)) But, he quickly extends this norm to belief itself:

"Coherence here includes the idea of rationality both in the sense that the action to be explained must be reasonable in the light of the assigned desires and beliefs, but also in the sense that the assigned desires and beliefs must fit with one another. The methodological presumption of rationality does not make it impossible to attribute irrational thoughts and actions to an agent, but it does impose a burden on such attributions. We weaken the intelligibility of attributions of thoughts of any kind to the extent that we fail to uncover a consistent pattern of beliefs and, finally, of actions, for it only against a background of such a pattern that we can identify thoughts." (159)

--There are many alternative belief-desire explanations for a given piece of behavior.

*What do you think of this claim: "In order to infer from such evidence that a speaker holds a sentence true we need to know much about his desires and beliefs, but we do not have to know what his words mean . .. We can know that a speaker holds a sentence to be true without knowing what he means by it or what belief it expresses for him."? (162)

--This is important, because it generates the following belief-attribution principle: "Abut if we know he holds the sentence true *and* we know how to interpret it, then we can make a correct attribution of belief." (162)

--We can have certain beliefs only if we know a language: "One can believe that Scott is not the author of *Waverley* while not doubting that Scott is Scott; one can want to be the discoverer of a creature with a heart without wanting to be the discoverer of a creature with a kidney." (163)

Interpreting speech allows us to make these more fine-grained belief attributions, which is valuable given that the non-linguistic behavior leaves a wide class of belief-desire explanations open for any set of behavior.

*"Mimic-theory" Hypothesis: "Given some delicate assumptions about the conditions under which such a subjunctive conditional is true, we could conclude that only someone with a language could have a thought, since to have a thought would be to have a disposition to utter certain sentences with appropriate force under given circumstances." (167)

--Davidson rejects this theory, with *its* strong dependence of thought on talk.

*Davidson's Principle of Charity: "But of course it cannot be assumed that speakers never have false beliefs. Error is what gives belief its point. We can, however, take it as given that *most* beliefs are correct. The reason for this is that a belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about. Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter. False beliefs tend to undermine the identification of the subject matter; to undermine, therefore, the validity of a description of the belief as being about that subject." (168)

"What makes interpretation possible, then, is the fact that we can dismiss a priori the chance of massive error. A theory of interpretation cannot be correct that makes a man assent to very many false sentences: it must generally be the case that a sentence is true when a speaker holds it to be." (168-169)

--Example: Did the ancients believe the Earth is flat?

--On p. 169, Davidson makes some passing comments about prioritizing cognitive errors.

*Davidson's big conclusions about thought and talk are in the first two paragraphs of p. 170. Davidson claims that only a creature possessing a language can have the concept of belief, and, further, that a creature cannot believe without a language.