LSU PHIL 4941 / Spring 2016 / John Protevi

http://www.protevi.com/john/PhilMind
Classroom use only.

Daniel C Dennett, "True Believers: The Intentional Strategy and Why It Works," in Chalmers 2002, 556-568. Original: 1981.

THESIS

Dennett is usually called an "instrumentalist" to distinguish him from realists and eliminativists with regard to folk psychology. For him, using belief-desire psychology (linked chains of propositional attitudes) and other elements of intentionality is a strategy ("the intentional stance" or ISt) for explaining and predicting behavior of "systems."

STRUCTURE

Death Speaks
- Topic: belief attribution
- Possible positions: realism, interpretationism, intentional stance
- Thesis: an intentional system is that whose behavior is well predicted by the "intentional stance"

The Intentional Strategy and How It Works
- Three Stances: physical, design, intentional
- Belief ascription process: interest / desire-relative truths plus rationality

True Believers as Intentional Systems
- Coarse granularity of ISt: allows picking out real patterns
- Internal complexity and mirroring vs representing

Why Does the Intentional Strategy Work?
- Evolution
- Language of Thought

ARGUMENT

DEATH SPEAKS

1. Possible positions about belief ascription
    a. Realism: belief ascriptions can be verified by (in principle) access to an objective fact about brain about which we now can make educated guesses
    b. Interpretationism: black-boxing the internal states, we make belief ascriptions by interpreting behavior
    c. Intentional stance: a strategy of treating the system as having beliefs, desires, and other elements of intentionality
2. Thesis: an intentional system is that whose behavior is reliably and voluminously predicted by the "intentional stance"

THE INTENTIONAL STRATEGY AND HOW IT WORKS

1. Three stances: physical, design, intentional
    a. Physical:
        i. Predict based on knowing the details and the rules
        ii. Laplace

  - iii. Quantum indeterminacy can ignored
  - iv. JP: see also chaotic systems: measurement errors can multiply
- b. Design:
  - i. Can be more effective to ignore physical stance and predict based on knowing what it was designed to do
  - ii. Multiple abstraction levels possible
    1. E.g., you can know, if you think it relevant, that a clock has gears
    2. Without bothering to know materials of the gears
- c. Intentional:
  - i. When design stance is practically inaccessible
  - ii. Four steps
    1. Decide to treat system as a rational agent
    2. Figure out beliefs it ought to have
       - a. Given its place in the world
       - b. And its purpose
    3. Figure out desires it ought to have
       - a. Given its place in the world
       - b. And its purpose
    4. Predict behavior
       - a. As furthering goals in light of its beliefs
       - b. That is, apply some practical [means – ends] reasoning
  - iii. Truisms about belief acquisition
    1. Exposure is normally sufficient for acquisition of knowledge qua beliefs about relevant truths
    2. So, attribute as beliefs those truths relevant to interests / desires a system has been exposed to
    3. Attribution of false belief requires a special genealogy
       - a. There will be an origin of the falsehood
       - b. In a system of largely true beliefs
  - iv. Fundamental rule: attribute instrumental rationality
    1. Attribute beliefs and desires a system *ought to* have
       - a. Basic desires: survival, absence of pain …
       - b. Other desires as means to those ends (and others)
    2. Side note: Verbal behavior
       - a. Allows specification of desires
       - b. Forces hyper-precision to beliefs they don't really have
       - c. This tempts us to think beliefs and desires as sentences stored in head
       - d. But these are special cases and not models for whole domain
    3. Attribution of rationality
       - a. Start with perfect rationality
         - i. Believe all implications
         - ii. And don't belief contradictory pairs
       - b. Revise downward: you only need enough rationality for predictability
    4. Ubiquity of intentional stance explained by its success

TRUE BELIEVERS AS INTENTIONAL SYSTEMS

1. Coarse granularity of ISt:
   - a. Advantage in picking out real behavior patterns
   - b. Compared to physical stance (Martian example)
   - c. Martians would have to treat themselves as intentional systems

2. Intentional stance is not perfect
    a. “Cognitive pathology” (e.g., contradictory beliefs)
        i. Hard realists say there are beliefs / desires the ISt can’t access
        ii. Dennett’s “mild realism”
            1. No facts about actual beliefs and desires
            2. But there are facts about success of ISt with different attributions
    b. Sheer relativism:
        i. Radical indeterminacy of translation due to radical incommensurability of cultures
        ii. That is, equal success of prediction from radically different ascriptions in ISt
            1. Is theoretically important
            2. But practically negligible when dealing with humans
3. Complexity of linkage:
    a. The thermostat is very simple: let’s say it has 6 beliefs and desires
    b. But these aren’t semantically rich:
        i. Even if it’s currently attached to one, it doesn’t have a concept of “boiler”
        ii. It simply believes X and desires Y when X obtains
        iii. So attach it to a refrigerator and it would still work w/o any changes in its beliefs
    c. Suppose you enrich its “modes of attachment”:
        i. Give it multiple, different inputs and outputs
        ii. More of what it can believe and what it can desire
            1. This would enrich the semantics of its beliefs and desires
            2. And make it less portable
                a. It would be a room thermostat
                b. Not fit to work as a refrigerator thermostat
    d. Thus, "a two-way constraint of growing specificity between the device and the environment."
    e. This is the difference between mirroring and representing
        i. A mirror is a semantically poor and hence “portable” state: it can regulate behavior in different environments without changing its internal configuration
        ii. A representation on the other hand is a semantically rich internal state that should be (i.e., rationally) sensitive to changes in environment
            1. It only works with a narrow range of fit with its target
            2. That is, it would be a disaster to try to interact with a tiger by using the representation of a kitty cat
            3. Rather, when confronted by a tiger when you had believed it was a kitty cat behind the door
            4. You need to exercise “rational revision of beliefs”
    f. So when we find something for which the ISt works we interpret some internal states as representations of world
4. Continuity between thermostats and us: “no magic moment” of transition
    a. There is only a difference of degree, but it’s a big difference
    b. Simple systems are portable due to minimal semantic content of its beliefs
    c. Complex systems produce a change in internal states in new environments
        i. That is, (relevant) changes in the world will change your representations
        ii. But the relevance criterion can be exploited in cases of subliminal changes:
            1. Because you don’t have normal access to chemical analysis, you don’t notice a difference on Twin Earth between H2O and XYZ
            2. You only see wet stuff which both you and your TE companions call “water”
            3. So you are like the thermostat that doesn’t know it’s connected to a boiler or a refrigerator:
                a. A change in the world has not changed your beliefs
                b. That is, in this case, your beliefs are mirrors, not representations

4. Thus you have different semantic content of your beliefs from others on Twin Earth even with identical internal states
    a. You both have the sense / intension “water”
    b. But different referents / extensions: H_2O vs XYZ
5. BUT this is a thought experiment not relevant to everyday life

WHY DOES THE INTENTIONAL STRATEGY WORK?

1. Evolution has “designed” humans to be rational
2. Though we don’t know how our rational machinery works
    a. Explanations
        i. Behaviorism: beliefs and desires are shorthand for complex S-R histories
        ii. LOT
            1. Beliefs, desires, inferences are mirrored in physical causes in brain
            2. Logical structure of PAs copied in structural form of states
    b. LOT may be true, but it’s not obviously true
        i. ISt does pick out real behavior patterns in the world
        ii. But it’s not obvious those real behavior patterns are produced by an isomorphic real pattern in the brain
        iii. However, you can argue for LOT:
            1. To avoid combinatorial explosion, language is the only solution we know
            2. So we should explore LOT as a promising hypothesis, not a necessary truth