# Where is Daniel Dennett?

Now that I’ve won my suit under the Freedom of Information Act, I am at liberty to reveal for the first time a curious episode in my life that may be of interest not only to those engaged in research in the philosophy of mind, artificial intelligence, and neuroscience but also to the general public. (Daniel Dennett, “Where am I?”)

In “Where am I?” (from *Brainstorms,* 1978) the philosopher **Daniel Dennett** offers one of the stranger thought experiments in contemporary philosophy. While it can be a bit difficult to follow on your first reading, I think that you’ll find the philosophical payoff is worth it. What follows is a rough guide to characters and events, as well as some thoughts on what exactly Dennett is trying to do here.

**Who is Daniel Dennett? Why is He Writing This?** Daniel Dennett (1942-) is a one the most important and well-known living American philosophers. He has written widely on topics such as evolution, neuroscience, and religion. He is a committed physicalist and functionalist, and is probably best known for his defenses of Darwinian evolutionary biology, functionalist theories of mind, and for his critiques of religion. He’s been called one of the “Four Horsemen” of New Atheism (along with Richard Dawkins, Sam Harris, and Cristopher Hitchens). In “Where am I?” Dennett tells a (obviously fictional) story about a strange series of events. Among other things, he is trying to undermine the conclusions of the “Chinese Room” or “What is it Like to Be a Bat?” thought experiments, both of which he finds unconvincing. Dennett wants to show that many of the intuitions we have about “ourselves” or “our experiences” are actually much less stable than we might expect, and that we should be *very careful* if we try to use these sorts of intuitions to overrule the results of our best scientific theories of the brain and its evolution. Dennett’s story will suggest that our own subjective experience is actually a pretty poor guide to these sorts of questions (so, for example, it turns out on his story that there’s nothing in our experience which would allow us to tell whether “I” am actually just a computer program).

## Our cast of Characters

“Yorick,” I said aloud to my brain, “you are my brain. The rest of my body, seated in this chair, I dub ‘Hamlet.’” So here we all are: Yorick’s my brain, Hamlet’s my body, and I am Dennett. Now, where am I? And when I think “where am I?”, where’s that thought tokened? Is it tokened in my brain, lounging about in the vat, or right here between my ears where it seems to be tokened? Or nowhere? Its temporal coordinates give me no trouble; must it not have spatial coordinates as well? I began making a list of the alternatives…(Dennett, “Where am I?”)

**Who are the character(s) is the story?** In some sense, there is only ONE main character in the story—a fictional version of Dennett himself, whose brain is removed from his body. However, in another sense there are LOTS of characters in the story, as “Dennett” gets split and copied in various ways:

* **Hamlet** is Dennett’s original body, but without a brain. Instead of a brain, the skull cavity contains radio transmitters that can allow Yorik (see below) to receive feedback from Hamlet’s five senses, and to control Hamlet’s movements. Hamlet eventually meets an unpleasant demise deep below the Earth.
* **Yorik** is Dennett’s original brain, which sits in a vat in Houston, TX. Originally, it is hooked up to Hamlet using radio transmitters. At Hamlet’s death, its sits in a vat, listening to music, thinking about philosophy, and sleeping. After the construction of Fortinbas (see next), it regularly receives input from the outside world. It sometimes can control Fortinbas, when Hubert is NOT doing so.
* **Fortinbas** is a *new* brainless body that the scientists provide for Yorik to use, after Hamlet dies. Like Hamlet, it is controlled using radio transmitters. It is similar to Hamlet, but not identical to it.
* **Hubert** is a digital computer that has been programmed to behave *exactly* like Yorik would, in every possible situation. Hubert is sometimes in control of Fortinbas, when Yorik isn’t. Like Yorik, Hubert is continuously receiving input from Fortinbas’s five senses. Since Hubert always makes the same choices as Fortinbas (at least until the end of the story), there is no way for outside observers to tell them part. One of Dennett’s worries: it’s not clear that “Dennett” knows whether Hubert or Yorik is in charge either.

## Some strange Happenings: THe Plot

An interesting way to view the story is to think what the events look like from the first-person point of view—from the viewpoint of the narrator of the story. This is because the main “action” of the story concerns the narrator’s efforts to figure out who exactly he is, or where his “mind” is located. As it turns out, the story suggests this is something of meaningless question, at least if we demand a definitive answer. So, here’s what happens:

1. **In the beginning.** Dennett’s brain (Yorik) and body are attached (Hamlet), just like the rest of ours. From Dennett’s point of view, he “is” wherever his body happens to be. This is the sort of scenario that most philosophers like to start with when trying to figure out what “minds” are. Dennett wants to complicate things, however.
2. **After the initial experiment.** Yorik stays in a vat in TX, while Hamlet heads off to OK. From the narrator’s point of view, however, it “seems” like he is traveling with Hamlet, as opposed to living inside of a vat. In fact, with a few exceptions (hitting a baseball is a bit tougher, since the connection between Yorik and Hamlet isn’t quite as fast as it used to be), there doesn’t seem to be much difference at all. On reflection, the narrator concludes that Yorik’s location doesn’t matter much to him. For example, it wouldn’t bother him if Yorik were put in jail, so long as Hamlet could walk around, and radio back the experiences to Yorik. At this point, flipping the switch will stop Yorik from controlling Hamlet (though not from seeing/hearing/feeling using Hamlet’s senses).
3. **The disconnection.** As Hamlet descends into the earth, the connection with Yorik gets disrupted. This leads to Hamlet dying, as tends to happens to brain-less bodies, and to the narrator suddenly finding himself “disembodied.” At this point, there is only Yorik (a disembodied brain with no body), and so the narrator begins to think that maybe “he” really is just a brain, after all. The narrator’s only experiences here are internal thoughts, or are the results of manipulation of the scientists (when they pipe music directly to the brain sans ears). The narrator goes to sleep…
4. **The new body.** Eventually, the scientists find a new body (Fortinibas) to hook up to Yorik (who is still housed in a vat). At this point, the narrator “wakes up” to discover himself back in the “real world,” with a working body. As before, he intuitively thinks he is wherever Fortinibas is. However, this is tempered by the knowledge that this could change at any moment.
5. **A surprise: A second brain.** The twist in the story occurs when the narrator decides to try out the old switch that prevents Yorik from communicating with Fortinbas. Instead of finding himself in a body that doesn’t respond to his commands (as happened when he previously did this), he finds that nothing at all changes! So what was the trick? As it turns out, the switch now toggles control of Fortinbas from Yorik (Dennett’s original, biological brain) to Hubert (the new digital brain with identical functional behavior). However, the narrator can’t tell (at any given moment) which brain is controlling the body, since the two brains make all the same decisions (and so, it always *seems* to the narrator that Fortinbas is behaving according to the commands he gives him). Since it makes no difference, the narrator loses track of which brain happens to be in control at any given moment.
6. **And another surprise: The brains part ways.** At the end of the story, the switch is flipped one more time. We now discover that the two brains—Yorik and Hubert—have diverged, and that they have begun noticing when the “other” brain is in control of the body. (So, for example, the brain thinks “raise my hand” and nothing happens). The narrator concludes that there are (at this point) two genuinely different people, and that he’ll need to find an additional body. So, it turns out there is NO answer to the question “Who am I?” Both brain-body combinations have an equal right to be called Daniel Dennett.

## Lessons from the Story

There are several lessons one might take from Dennett’s story:

**Our “privileged access” to our minds isn’t what it is cracked up to be.** Thought experiments like Searle’s and Nagel’s lean heavily on the intuition that we “know” things about ourselves that other people can’t. We know what we understand (and don’t); we know what we experience, and so on. In Dennett’s case, however, it looks we don’t actually have this sort of unproblematic access to who and what we are. After all, by the end of the story, the narrator has *literally no idea* whether his “thinking” is being done by neurons or by silicon processors. Moreover, it turns out not to matter! The only reason for caring about “who Dennett is” is because of *practical* issues—having two Dennetts would cause problem for Dennett’s wife, his employer, and so on. At a metaphysical level, there is simply no “right” answer to the question of who Dennett “really is” or where he is located.

**Functional equivalence is all that matters.** On Dennett’s telling, the fact that one brain is made of neurons and one is made of silicon doesn’t much matter, so long as the brains give the same “output” for any given “input” (i.e., so long as they both have all the desires, intentions, and beliefs). However, the brains clearly DO become different when this functional equivalence disappears at the end of the story—when the brains start wanting their shared body to do different things here. Again, though, Dennett suggests that there is no reason to think that the biological brain’s desires are any more “real” than are the silicon brain’s desires.

**Stop saying machines can’t think!** In Dennett’s story, the narrator is unsure whether or not his “brain” is a computer or not. However, this turns out to make no practical difference to him. After all, from his “subjective” point of view, it doesn’t make much difference: in either case, his body behaves the same way, he feels the same things, and so on. The only time it *does* make a difference is when Hubert and Yorik start having diverging personalities, and thus begin to notice things like “my body—Fortinbas—isn’t doing what I want it to do!” Here, again, the problem is a practical problem (“We better find another body!”) rather than a metaphysical one (“Where is the *real* Dennett?”). Dennett thinks the former problem actually makes sense; the latter problem just reflects a demand for an answer that doesn’t actually exist. For Dennett, there’s simply no reason to suppose that Yorik (brain)-body is any more of a mind than Hubert (computer)-body.

**Thought experiments should be treated with caution.** Authors like Nagel and Searle offer thought experiments that seem to show the *impossibility* of imagining certain sorts of things—what it is like to be a bat, or a computer program that genuinely understands Chinese. Dennett’s story can be seen as an argument that these thought experiments don’t actually show of anything of the sort. Instead, they simply show that humans aren’t always great at predicting what is (or what isn’t) imaginable or possible. Dennett loves thought experiments, but he thinks that we should always be aware of how the thought experiment “works”—e.g., what would happen if we changed various aspects of the story, etc.

## REview Questions

1. Where is Dennett at the end of the story? Defend your answer using what you’ve learned in class so far.
2. Much of Dennett’s career has been spent trying to reconcile our prescientific beliefs about the world (about consciousness, religion, the origin of life, etc..) with the results of modern science. In general, Dennett’s work is based on the idea that once we understand *why* our (all-too-human) brains can be misled by mystical ideas, we will get over our discomfort with physicalism and reductionism. To what extent do you think this is a good/plausible concept of philosophy?