Phenom Cogn Sci
DOI 10.1007/s11097-014-9359-x

Artifactual selves: a response to Lynne Rudder Baker

Daniel C. Dennett

Abstract Baker's critique of my view of the self as a fiction captures some of its points well but misses the possibility of a theorist's fiction, like the Equator or a center of gravity, which is not an illusion, but rather an abstraction, like dollars, poems, and software—made of no material but dependent on material vehicles. It is an artifact of our everyday effort to make sense of our own (and others') complex activities by postulating a single central source of meaning, intention, and understanding. This is revealed in an example of the heterophenomenological method in action.

Keywords Narrative · Self · Fiction · Author · Abstraction · Heterophenomenology

Lynne Rudder Baker's essay begins well, with an accurate and lively account of my view that human bodies—more particularly, human brains—"spin selves" by generating a narrative starring the self as a fictional character. But it somewhat goes off the track when she fails to appreciate that it is a *theorist's fiction*, not some idle fantasy; it must pay for itself by being supremely useful to the various human beings in the neighborhood who are trying to make sense of, and live peacefully with, the human body whose brain is largely responsible for the spinning. I confess to not having made this sufficiently clear, thinking that it was a clear implication of the role I give to the intentional stance in predicting and explaining the complex behavior people are capable of. Baker says that on my view "the self is a totally fictional character that emerges from the story produced by component modules in the brain. The self—or the subject of experience—is an illusion." Not an illusion, really, any more than a center of gravity is a "totally fictional entity" or an illusion. If you think of a center of gravity as more concrete, more like an atom or subatomic particle than it is, then you are creating an illusion for yourself, I guess. Similarly, if your conception of the Equator leads you to be surprised and disappointed when no bright or dark line appears in the ocean or on land where the Equator lies, you have fallen for a self-generated illusion of sorts. I suppose I am in fact saying that those who think a self is like an organ or brain module, *made of something*—if not matter, then ectoplasm or some such stuff—are succumbing

D. C. Dennett (✉)
Center for Cognitive Studies, Tufts University, Medford, MA 02155, USA
e-mail: Daniel.dennett@tufts.edu

Published online: 18 March 2014

to an avoidable illusion, but setting such overly concrete, overly "realistic" views aside, I see nothing illusory about selves as centers of narrative gravity. They are largely well-behaved abstractions, no more illusory than dollars or home runs or other socially constructed things. Perhaps a better comparison would be to software. Is software real, or an illusion? To my amazement, I've discovered a serious philosopher who claims to doubt the very existence of software! He doesn't think Bill Gates should go to prison for perpetrating a gigantic fraud; he recognizes that Windows, and Acrobat, and Safari . . . are worth buying, worth using, etc., but they aren't made of anything material, and in that sense they are not real. I chide him: neither, then, are Shakespeare's sonnets or calculus. A self that is a center of narrative gravity is an entity with quite demanding identity conditions and can have quite a robust biography. (Is Windows 8 really Windows?) Baker also says this:

> The brain itself cannot produce any character—fictional or not—unless the brain is interpreted as an author (from the intentional stance). With no constraints on interpretation of the brain as author, the narrative could just as well be interpreted as having a real protagonist as a fictional one. If so, how would any genuine project of interpretation get off the ground?

I have trouble with this passage. I had not meant to suggest that the brain (in whole or part) was an *author* when I said it generated a narrative. What else, you may wonder, could generate a narrative? Well, how about a tipsy gang of people playing some party game in which players throw dice, then draw cards with words on them, and then somehow negotiate a consensus narrative according to the rules of the game. Who's the author of that narrative? No single author, obviously (unless the game has been rigged by somebody). Could a brain do something like that without being an author? Yes! There could be a pandemonium architecture of brain circuits all "trying" to get certain words, certain phrases, into the story, winning a few and losing a few, the whole political process being constrained by the needs of the human body, its current circumstances, etc., etc. You may doubt that much of a narrative could be generated by such an authorless (but information-sensitive) process, but then you have a problem that is rather the opposite of the infinite regress of interpreters Baker tries to lay at my door: what is an author—a genuine intelligent, intentional author (like Updike or Nabokov)—made of? What kind of parts? There is no way you could adopt the intentional stance to a full-fledged person-level author unless the body in question was controlled by sub-personal processes that bore a striking resemblance to the party-game players: semi-understanding, myopic, competitive agents more or less unwittingly generating a narrative without intending to do so. It is Baker's insistence that nothing but an author can create a narrative that generates the regress. (I don't understand the second and third sentence in the quotation from Baker above—I may have woefully misinterpreted her.)

Another passage from her essay:

> A theorist may provide a narrative interpretation of the neural activity of the subject. But what neural activity? This procedure of beginning with neural activity (from the physical stance) is the opposite of Dennett's heterophenomenology. Suppose that a theorist had a detailed video of the neural

> events going on in a particular brain. How would the theorist (who does not even know when the brain is producing a narrative) be able to identify the neural activity to be interpreted as producing a narrative with a fictional self—not to mention how a theorist would know what story was being produced?

Baker's incredulity is out of place here. Suppose the theorist has just given the subject (the human being under hereophenomenological investigation) a problem to solve:

> Imagine you glued five cubes in a vertical stack and glued it upright to the floor; Then you took two cubes—same size as the others—and glued them to the side of the top cube in the stake, facing to your right; then took a single cube and glued it to the right side of the middle cube in the stack. What would you have made? And now name an animal and a US President that begin with that letter.

The subject will not find these tasks difficult, and, if probed, will say how first she formed the cube assembly image "in her mind's eye," recognized it as a capital "F" and was then able to perform the second task by consulting her memory and coming up with FOX and FORD (or FILLMORE). Nothing contentious or mysterious here—I'm deliberately choosing a dead-simple example. The subject's public, recordable narration purports to describe mental activities very recently conducted, and it is not implausible to suppose that the heterophenomenologist could use such public narratives to map out (at least temporarily—brains are very plastic) cortical activities corresponding handsomely to the F-construction, recognition and memory retrieval. One might be able to tell that the subject neglected to divulge that she first followed the directions incorrectly, putting the top element of the F facing left, and also thought of *fawn* and *FDR*, which she rejected. In such a case, the experimenter could use details of neural events to uncover some of the *private* narrative of the subject. Now these are not the most interesting elements of a person's subjective biography, but we can already identify such simple contentful elements in carefully circumscribed experimental situations.

We shouldn't expect the subject's own convictions about what it was like to respond to these requests to be entirely or straightforwardly born out in neural details. It is very likely that our private narratives are as self-servingly idealized as the redacted portions of them we release to the public. Moreover, our public avowals—on which heterophenomenology depends—will tend to be schematized and laden with metaphors we can't unpack from a privileged vantage point. We have no choice but to rely on imperfectly fitting characterizations of our experiences because we are obliged to couch everything in the *lingua franca* of public language. "It's ineffable!" we say, and we mean it, and we're right. But the thresholds of ineffability are variable, and we can train ourselves to a surprising degree to delve deeper into our experiences, analyzing and fractionating our "holistic", "intrinsic", "atomic" qualia into their otherwise non-introspectible components.

The main reason I talk of the self as the center of narrative gravity is that I view it as an artifact of our (mainly unselfconscious and innocent) efforts to solve the myriad little problems of interpersonal activity we encounter every day, from the moment of our birth. What begins with as-good-as-random babbling and gesturing is swiftly pruned by the selective environment of reactions, of parents and other things, movable and

stationary. We learn eventually how to express our wants, to tell Mother where it hurts, and, as we grow more worldly, to dissemble, deceive, and conceal. Our abilities to *control* all these highly sensitive, nuanced, and demanding activities develop *unconsciously*, competences without explicit comprehension, generating what Erving Goffman aptly called *The Presentation of Self in Everyday Life*, (1959). The only alternative on offer to this vision of an emergent, virtual self is an extremely dubious dualism of real (*really* real, I guess you could say) selves in real bodies, an idea that is about as antithetical to a scientific understanding of mind as *élan vital* was to a scientific understanding of life.

References

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Doubleday.