

for Artificial Intelligence Journal
August 27, 2007

Instead of a Review

Douglas Hofstadter, 2007 *I am a Strange Loop*, New York: Basic Books.
Marvin Minsky, 2006, *The Emotion Machine*, New York: Simon & Schuster

I am much too close, personally and intellectually, to Doug Hofstadter and Marvin Minsky to write a proper academic review of these wonderful books, so this will be an appreciation, an avowedly partisan paean to both the projects and the methods of these two original and bold thinkers. Such a celebration might seem unnecessary, just one more bouquet tossed onto the stage by an ardent fan, if it weren't for the fact that in spite of—or perhaps because of—their fame, many researchers who ought to know better ignore or dismiss them, and this is a golden opportunity for me to try to change their minds. I travel in several quite different academic circles and I find that each gang has its particular way of not taking these thinkers seriously. The neuroscientists deplore the absence of rigorous experiments and the refusal of both Hofstadter and Minsky to canvass the relevant experimental literature thoroughly and explicitly. *Where are the data?* The philosophers of mind, at the other extreme, find few formal arguments and a frustratingly cavalier refusal by the authors to define their terms at the outset. *Where are the proofs?* The cognitive psychologists, in the middle, are offended by the fact that neither Hofstadter nor Minsky sees the need to adjudicate between all the competing models and theories that have been painstakingly developed and defended, and instead offer their own impressionistic and oversimplified sketches. *Where are the models that make testable predictions?* The artificial intelligence crowd wants to see a running demo program. *Where is the code?* It's all just *speculation!* And then there is the style, too clever and playful, apparently written for bright high school students, not professors and graduate students. How can any self-respecting academic researcher admit to having learned anything from these crowd-pleasers? This isn't research, it's *edutainment!*

The books amply support these charges, but they aren't bugs, they're features. Hofstadter and Minsky have both developed models and theories that should satisfy the most hard-bitten empirical scientist, but now, enlightened by those projects, they are operating in a domain that many researchers find unbearably anarchic; nobody knows the rules. But that doesn't mean we should shun this lawless territory. And their super-accessible styles are integral to the task; these authors aren't out to impress anybody; they're out to explain, using whatever might work. (And finally, isn't Dennett just biased because of the heady praise both authors give his own work? That's a distinct possibility, but I will try to overcome that suspicion by explaining just what I think is valuable, indeed necessary, in the contributions both Hofstadter and Minsky are making to the daunting task of explaining how the human brain makes the human mind.)

The difficulty with human minds is that they are much too familiar to us. We already “know all about them” and are effortlessly fluent in describing and explaining how they operate *using the terms of folk psychology that we learned as children*. We “see” and “hear” and “think”; we “make up our minds” and “decide” and “remember” and “intend” and “imagine” and “dream” and “wonder” and “hope”; we are beset by “pains” and “urges” and “anxiety” and

“boredom.” But when we ask ourselves how we do all these things, we usually go around in circles. For instance, as Minsky notes:

. . . what is a goal, and how can we *have* one? If you try to answer such questions in everyday words like “*a goal is a thing that one wants to achieve*,” you will find yourself in circles because, then, you must ask what *wanting* is—and then you find that you’re trying to describe this in terms of other words like *motive, desire, purpose, aim, hope, aspire, yearn, and crave*.

More generally, you get caught in this trap whenever you try to describe a state of mind in terms of other psychology words because these never lead to talking about the underlying machinery. (p187)

Now it is possible to spend one’s whole academic career exploring the epicycles of these circles. Many philosophers of mind have done just that, filling hundreds of books with carefully analyzed accounts of the implications and assumptions of these everyday “psychology words” and never even trying to ask what kind of underlying machinery there might be and how we manage to catalogue its products and actions so deftly without having a clue what we’re talking about. Some of them haven’t asked this question because they firmly believe that there *is no* machinery; it’s all done by some kind of utterly non-mechanical mind-stuff. Or at least: the mental realm and the physical or causal realm are “incommensurable,” or there is some sort of “radical emergence” that makes it utterly impossible to go back and forth between mind talk and neuron talk. Hofstadter has a typically clever (and tongue-in-cheek) way of exposing and resisting this defeatism. He coins the terms *thinkodynamics* and *statistical mentalics*. The idea is that like thermodynamics, thinkodynamics is conducted in terms of large-scale structures and patterns, while like statistical mechanics, statistical mentalics shows how this all “reduces” to the kazillion little steps at the lower level. (p34) There *is* a sort of “emergence” here, but it is not the sort that mysticians like to imagine; it is like the emergence of traffic jams and Conway Life patterns and, well, the temperature of a gas: don’t try to define the macro-level phenomena as *specific* complexes of micro-phenomena (the way we can define an atom of some element as a satisfyingly specific organization of sub-atomic particles). The macro-level phenomena are statistically robust patterns that permit vast variation in their micro-elements. And, one has to do a little squinting to see these patterns. They are only approximately tracked by the micro-machinery. All this is music to my ears and it is, really, philosophy.¹ As it stands it isn’t an empirical theory of anything, and it doesn’t prove anything, but it does *open up the idea of possibilities that otherwise might be ignored or underestimated*. It is a bit of conceptual calisthenics, exercising our imaginations to rethink the issues and cast off some of the assumptions that have been hobbling us.

Some may think that they don’t need any such calisthenics. They already have a firm, clear grasp on the nature of the phenomena they are trying to explain. Perhaps they are right, but then where are their theories or models? We’re drowning in data, but nobody seems to know how to refine it into evidence for any comprehensive theory. So people work on truncated issues, leaving till later the daunting task of merging their local visions into a single non-contradictory model. It’s a defensible research strategy (but see Allen Newell’s classic article “You can’t play

¹See my “Real Patterns,” 1991, for some different paths to the same happy territory.

Twenty Questions with Nature and win” 1973). People shouldn’t try to do what they are no good at doing, and it is also true that this genre of free-wheeling “informed speculation” can be an utter waste of time in the hands of somebody who isn’t well-informed and hasn’t got a disciplined imagination. Very few people are equipped to do this kind of work well, but both Hofstadter and Minsky have had years of practice using the best available thinking tools. They haven’t just marinated their minds in Lisp and other computer languages; they have built and dismantled computer models of a wide variety of phenomena, and sympathetically explored the efforts of others. Among the treasures of Minsky’s book are the morals he draws from the pioneering work in GOFAI (Good Old Fashioned AI, Haugeland, 1985). Although those avenues are now widely branded as having been disproven and discredited, Minsky extracts the *conceptual* fruits they yielded in spite of their now almost comical simplicity. Yes, they were naive, and sometimes offensively hubristic, but they also illuminated otherwise murky corners of the mind, and Minsky does a fine job of saving these insights for another generation of theory-mongers. Minsky’s evolving model-sketch of the mind now is cast in term of a gaggle of “Critics” that are specialists on, well, *particular ways of dealing with things that happen*, and his recommendation to anyone who wants to join him in the sort of exploration he is engaged in is simple: you have to turn off some of your Critics. This policy is anathema to some kinds of thinkers—especially philosophers, whose training largely consists in having their noses rubbed in their uncritical leaps and lapses. It is undignified, and dangerous, and unprofessional to cavort around the way Hofstadter and Minsky do! So I know that many of my philosopher colleagues around the world will be simply unable to read these books with any sympathy; all they will see are *non sequiturs* and opinions supported by impressionistic reflections. But in fact if they look closely, they will find plenty of very tough reasoning indeed, pinching off otherwise tempting paths.

Both of these books, for all their differences, provide us with examples of what might be called super-duper pseudo-code. It is quite a few hard steps away from actual code (in any computer language now known) but it is also quite a few hard steps closer than the folk psychology with which we all start. As Minsky says of one of his models it is “intended to be vague” (p47), and he deliberately avoids going into much detail about how to construct the various Critics whose efforts he describes in unabashedly psychological language. That would be ominously question-begging if it weren’t for the fact (and it is a fact, though some thinkers refuse to acknowledge it) that among the fruits of AI are a bounty of *simple* cognitive agents with proven competences that license theorists to describe their talents in psychological terms (“from the intentional stance” in my terminology). AI has made enough simple-minded, myopic, slavish agents to set our minds at ease when we speak, metaphorically, about subsystems in the brain that can search, recognize, remember, judge, infer and so forth.² We may not know exactly what they are made of yet, but they do not have to be made of “wonder tissue”. We have plenty of non-miraculous examples of machinery that can do the tricks postulated.

Hofstadter situates his sketchy super-duper pseudo-code model a bit lower, postulating “symbols” (“simmballs” dancing and bumping around in the “careenium”) that can be “dormant”

² For a book-length diatribe against this practice, see Bennett and Hacker, 2003 and my rebuttal in a newly published “authors meet critics” session at the American Philosophical Association, Bennett, Dennett, Hacker and Searle, 2007.

until awakened by the arrival of their triggering condition at their rudimentary internal sense organs. “In this book, then, symbols in a brain are the neurological entities that correspond to concepts, just as genes are the chemical entities that correspond to hereditary traits.” (p76) And just as population genetics made great progress for decades before Crick and Watson came up with a molecular structure that could do the work genes were known to do, so population symbolics (my coinage, not Hofstadter’s) can make progress by working at a deliberately abstract and idealized level while we await help from neuroscientists on how symbols are realized. After all, Crick and Watson would not have known what questions to ask, what properties to hunt for in DNA, if it weren’t for the quite clearly defined competences that genes were postulated to have. Among the competences of these simmballs/symbols is that they stay “*systematically* in phase with the things going on in the world” (p195). Well yes, there does have to be something in the brain that has this wonderful property, and Hofstadter is willing to take it more or less for granted while Minsky is much more engaged than Hofstadter in saying how to make that property real.

How are these symbols (or Minsky’s counterpart “micronemes”) an improvement over Hume’s *impressions* and *ideas* or Fodor’s words in a *language of thought*? By being active, semi-autonomous, purpose-driven little bots, not inert tokens or mere addresses in a vast memory. It is this fact that saves Hofstadter from what at first blush may look like a particularly forlorn misuse of analogy. Over the years he has had a childlike fascination with curious feedback loops, epitomized by the effect of pointing a camcorder at the television screen it is plugged into, creating fractal-like whorls and geometric embeddings that zoom off to a vanishing point. *Something like this*, he proposes, is the key to the puzzles of human consciousness. Really? Isn’t he just confusing vision with television? No, and that is the point of the book’s title: the camcorder and television set form a magnificent loop, but not a *strange* loop. A strange loop involves “the selective triggering of a small subset of a large repertoire of dormant symbols” (p75), and a television feedback system doesn’t have these symbols at all. One might say that in the television system, there is plenty of feeding, but no digesting. It is the digesting, the transforming at many different levels, that makes it possible for a loop to generate phenomena we can recognize as central to our understanding of what makes our own conscious minds so open-ended and yet so susceptible to traps, bad habits, brainstorming and doldrums. The reflections our minds are capable of are a far cry from the reflections in a hall of mere mirrors. And because of this incessant sharing, copying, modeling, and so forth, our minds become inseparable blends of the minds of our colleagues, friends, lovers, and *imprimers* (Minsky’s term, which nicely captures the parallel with the way goslings imprint on their “mothers”—or any other large, influential, moving thing). Who *I* am distinct from *us* begins to lose both its definiteness and its importance, when we appreciate the “the blurry glow of human identity” (the title of Hofstadter’s chapter 18). In the same way that a mosquito’s symbols just barely deserve the label because, thanks to the very limited company they keep, they cannot do very much, so what *we* “as individuals” can do depends very heavily on what we as a community of communicators can do. Of course this means that both Doug and Marvin had a big hand in writing this essay, even before I let them see the penultimate draft! That’s what I meant at the outset when I said I couldn’t write a proper review.³

³In 1983 Martin Gardner wrote a review of his own book, *The Whys of a Philosophical Scrivener*, under the pseudonym George Groth, in *The New York Review of Books*. It was scathing. It begins: “This is one of the strangest books of philosophical game playing to come along in many a moon. The author seems well acquainted with modern philosophy--indeed, he studied under

Another similarity between Hofstadter's and Minsky's books is that both were written as sequels born of frustration. As Hofstadter says ruefully in his preface, in spite of all the attention and praise lavished on his masterpiece, *Gödel, Escher, Bach* (1979), its "fundamental message . . . seemed to go largely unnoticed" (p xiii). Minsky's 1986 book, *The Society of Mind*, was also lauded but underestimated and misunderstood. The books under review attempt to tell the tales again, with improvements and embellishments, and I think they both succeed, at least for me. Yes, even I, good friend and theoretical bedfellow of both authors, drastically underappreciated some of their favorite themes until reading these new books. That's really why I'm writing this appreciation: to suggest to others that if they think they already know and understand everything they might learn from them, they are almost surely mistaken.

For instance, I never before really appreciated the depth of the relationship Doug saw between Gödel's Theorem and consciousness, and my eyes were opened by his brilliant re-exposition of Gödel's proof and the way it exploited and transcended Bertrand Russell's notoriously unsatisfying theory of types. Russell thought he knew exactly what the representation system he and Whitehead presented in the *Principia Mathematica* was *about*—and what it wasn't about! He was wrong. Gödel showed how it could also be about something else, unimagined and unintended by Russell. The same *invisibility from below* is a feature of the content level in brains, and Gödel's brilliant invention of a mapping scheme opens up the prospect of a similar, if less mathematically rigorous, mapping scheme of mind-events onto brain-events. To put the point in a way that I hope Doug will approve of, just as Russell didn't have to know what he was doing when he created the target system for Gödel's interpretation, so the brain doesn't have to know what it's doing when it creates the mental activities of a person who *does* know what she's doing. Thanks to Doug's exposition (especially p137) I can also see, for the first time, why Gödel's theorem is not so surprising after all. Indeed, if it were not true, we could have miraculous oracles to help us answer all our questions of number theory.

Much of the substance of these books cannot be usefully summarized since they both consist in large measure of dialogues that gently bring the reader along, articulating objections, exploring them and setting them to rest. One might as well try to summarize a lullaby. But these books are wake-up calls, well worth the serious attention of anybody who wants to have a theory of the mind.

References:

Bennett, Maxwell and Hacker, P.M.S., 2003, *Philosophical Foundations of Neuroscience*,

Rudolf Carnap and even edited one of Carnap's books--yet he defends a point of view so anachronistic, so out of step with current fashion, that were it not for a plethora of contemporary quotations and citations, his book could almost have been written at the time of Kant, a thinker the author apparently admires." (*New York Review of Books*, Volume 30, Number 19, December 8, 1983.) It would be fun, I think, to write (and publish) a fiercely critical—but fair!—review of one of my own books, but I can't bring myself to do something like that with Marvin's or Doug's books. It's rather like trying to tickle yourself. As long as you're in control, you can't really be tickled—or hurt, but you wouldn't want to risk discovering that you and your friends had rather non-overlapping senses of what was fair. So there is still plenty of distance between *me* and *us*.

Blackwell, Oxford.

Bennett, Maxwell , Dennett, Daniel, Hacker, P.M.S., and Searle, John, 2007, *Neuroscience and Philosophy*, New York: Columbia Univ. Press.

Dennett, Daniel, 1991, "Real Patterns," *J.Phil.* **88**, pp27-51.

Haugeland, John, 1985, *Artificial Intelligence: the Very Idea*, Cambridge, MA: MIT Press.

Newell, Allen, 1973, "You can't play Twenty Questions with nature and win," **XXXX**