Astride the Two Cultures: A Letter to Richard Powers, Updated

Daniel C. Dennett

Where the hell is the Two Cultures split when you need it? —*Galatea 2.2*

Suppose you want to write a novel about a great poet. The instructor in your creative writing class keeps admonishing you "show, don't just tell" and in this instance the way to follow the advice is to *exhibit* some of the hero's great poems. But that means you'll have to write some great poetry. Alternatively, you can give up and just tell the reader how ravishingly beautiful the poems were, how deep, how elegant, how intricate, and you may support these claims with anecdotes about how the poems made strong men weep, brought jaded critics to their feet, and were an inspiration to all who read them. In short, you can represent the hero's poems *as* wonderful, but without having to come up with any particularly wonderful representations of them yourself. Child's play. There are other such dodges, well-known to us all. Consider, for instance, the standard cartoonist's trick of portraying genius scientists at the blackboard covered with a thicket of equations that we are to understand to be brilliant and deep, when in fact what is written on the blackboard is gobbledygook. That's another cheap way of representing something as wonderful and intricate without having to come up with a wonderful and intricate representation of it.

Richard Powers faces this problem in *Galatea 2.2.* He wants to give his readers a representation of cutting-edge research in artificial intelligence and he doesn't want to settle for the dodge. His representation of AI is wonderful. It is remarkable how this very interested bystander has managed to cantilever his understanding of the field out over the abyss of confusion and even

throw some pioneering beams of light on topics I thought I understood before. I, too, am something of an interested bystander, not a participant; I've never myself written a serious AI program, but I've talked and worked with people in AI for more than thirty years, so I'm not really an outsider. I was introduced to Powers's book by my good friend, Seymour Papert, one of the founding fathers of artificial intelligence, creator of the kids' computer language, Logo, and the co-author, with Marvin Minsky, of the classic book *Perceptrons*, a critique of the first wave of connectionism. Papert and I have been talking about artificial intelligence and related topics for years, and if he thought it was worth recommending, that was already high praise. I read the book and then did something I had never done before. I wrote a fan letter.

That letter captures my sense of the book with more immediacy and passion than any newly composed essay could do, so I am simply reproducing it here, unedited aside from some updating and clarifying comments. What particularly excited me, as you will see, was how Powers had managed to find brilliant ways of conveying hard-to-comprehend details of the field, details people in the field were themselves having trouble getting clear about. Sometimes outsider observers can hit upon felicitous ways of putting things that have escaped the searches of the primary workers, and when this happens, it isn't just teaching. It is a contribution to the research itself, just the sort of contribution I myself have hoped to make to the fields I hang around in. "I wish I'd said that!" was a frequent, and particularly heartfelt, sigh as I read his pages.

The text below includes quotations from the 1995 novel, my original 1997 comments, and then new 2003 comments on those comments, and since some of the latter include italicized titles and the like, I've settled on the following scheme to mark the different dates and authors: quotations from the novel are in italics; my original comments are plain roman, and my additions are in boldface between curly brackets.

◆

June 6, 1997
Dear Mr. Powers,

Last summer Seymour Papert lent me his copy of *Galatea 2.2* with a high recommendation. I devoured it with delight, unable to restrain myself from

highlighting favorite passages in his copy, and bedecking the pages with Post-its. I offered him a choice between his copy back or a fresh copy. He chose the former, so I went out and got my own copy and lovingly reproduced the highlights and Post-its. Now, the Commencement behind me, I can turn to personal mail at last and write you a fan letter—an unusual project for me.

Herewith, then, a partial collection of those passages, with occasional comments.

Page 6. *"At the vertex of several intersecting rays—artificial intelligence, cognitive science, visualization and signal processing, neurochemistry—sat the culminating prize of consciousness's long adventure: an owner's manual for the brain."* Love it. I expect I'll use it soon as an epigraph for something. **{I haven't used it yet, but the quest for the owner's manual continues apace, with ever more aspirants.}**

9. Excellent paragraph about the web as stock exchange.

18. *"How do you deal with that syntax? And even by native speakers not until the ultimate grammatical arrival capable of being unraveled word order that one's brain in ever more excruciatingly elaborate cortical knots trivially can tie."* A gem. **{Chomsky and his many followers have been trying for decades to specify the Universal Grammar that they claim is innate in the human brain, and part of the problem facing them is that trivialization looms unless the UG has some engineering quirks, some features that *could have been otherwise* without significant loss of efficiency. Without such telltale "frozen accidents" of evolutionary R and D, the common features of all human languages may be due not to our genes but to our reinvention, again and again, of the optimal tricks. Languages can be remarkably different in word order and yet still have underlying similarities of abstract structure that enable the brain to compose and parse. Can you describe a language that the brain *can't* learn with ease—and that isn't just horribly inefficient (like MCXXXIV compared to 1134)? The difference between German and English word order shows how elusive this goal is. The languages would have to be *more different than that* and yet the unlearnable one would have to be approximately as efficient (by "engineering" standards) as these.}**

41. *Rhesus pieces.* Oh my. Oh well. I guess I wouldn't be able to resist, either.

44. *"Where the hell is the Two Cultures split when you need it?"* Indeed. A great line. I'll try to remember to cite you when I use it. {**John Brockman, the literary agent and proprietor of the website edge.org, has been promoting the idea of the "third culture [which] consists of those scientists and other thinkers in the empirical world who, through their work and expository writing, are taking the place of the traditional intellectual in rendering visible the deeper meanings of our lives, redefining who and what we are."** ***The Third Culture: Beyond the Scientific Revolution*****, 1995. When this third culture is good, it is very very good, and when it is bad, it is horrid, which is why I like this line.}**

88–9. Excellent discussion/description of the issues. Ditto, pp. 110–11, pp. 120–21. {**In the first passage, Powers sees clearly why consciousness can be a red herring in the quest for the owner's manual. The second has some vivid observations on "scripts" and "inheritance of classification qualities," and the third gets to the heart of the matter: recursion can permit a system to engage in second-order (and third- and ever higher-order) reflections: ideas about its ideas about . . .**}

114. *"It could not move ideas around. All it could move around were things. And the things had to be visible at all times."* This gets close to the heart of current work of mine and Andy Clark's. (I enclose a copy of my recent review of his new book {***Being There***} in *TLS*, and a few forthcoming papers of mine.)

121–2. Brilliant. Absolutely brilliant stunt. Diana's responses are right on the wrong side of the cutting edge, not quite believable. I loved it. {***"Many things are so bright without being happy."*** **Only an AI program would say something like that!**} (Do you know about the time Doug Hofstadter was fooled in just the same way, by grad students in Kansas? Perhaps that trick was the source of your inspiration here. Was it? I'm curious.) {**It was, Powers told me.**}

126. Symbolic grounding—yup, or as I once put it, "eyes, hands and history" (at the end of "Can Machines Think?"). {**The term *symbolic grounding* has become a buzzword in cognitive science. I think Powers's use is a case of convergent evolution, not the source coinage. Stevan Harnad has long been writing about symbol grounding, and at the Media Lab at MIT, Deb Roy's cognitive robotics group is currently doing by far the most interesting and impressive work on the topic. See http://www.media.mit.edu/cogmac/projects.html.**}

153. I like the reflexive "*epistemological parfait.*" Great image.

154. "*Meaning was not a pitch but an interval.*" I've already quoted this, several times, but not yet in print, I guess. Duly referenced, you can be sure. I also liked: "*Every neuron formed a middle term in a continuous, elaborate, brain-wide pun.*" {**For a magnificent exploration of similar ideas, see Dan Lloyd's book, *Radiant Cool: A Novel Theory of Consciousness*, a novel that carries the reader's imagination even farther into the world of multi-dimensional scaling of brain activity.**}

156. Here you get close to the heart, again, of the current thinking I take most seriously about hybrid systems. "*Giving in to a limited, rule-based control structure freed Lentz to recurve G's layers, turning them back in on themselves . . .* " Now do you have one or two prize informants in the field—if so, who are they, because I want to be in touch with them—or are you figuring all this out for yourself? {**From his response I gathered that Powers's imagination was his primary informant. Aside from some stage-setting and a few anecdotes, he was not relying on informants to clarify the quest.**}

160. "*I meant to reverse-engineer experience.*" Music to my ears (see DDI {***Darwin's Dangerous Idea***}), but then two lines down, I catch one of the very few minor blemishes in the book: you mean *retinotopic* not "retinoptic" map. (It occurs several times.) I'm assuming it's a typo, not a thinko. {**For the distinction see my "From Typo to Thinko: When Evolution Graduated to Semantic Norms," forthcoming, but available in draft at http://ase.tufts.edu/cogstud/papers/THINKO.htm**}

179. "*H had learned something. Whatever stuck in the throat, indigestible, could be made less acute by slipping it into a question.*" Perfect. Another epigraph for sure.

Also good on the same page: "*Time passed for it. Its hidden layers could watch their own rate of change.*" I am guessing that you may have picked up this important point from Paul Churchland's rhapsodies about connectionist nets and time perception in his recent book {***The Engine of Reason, the Seat of the Soul*, MIT Press, 1995**}, but his badmouthing of me on the topic of time needs a little response. So I include a forthcoming irenic reply, which also expands a bit on some of the points above. {**Shannon Densmore and Daniel Dennett, "The Virtues of Virtual Machines," in *Philosophy and Phenomenological Research*, September, 1999, Vol. LIX, No.3, pp. 747–67. Lloyd's *Radiant Cool* has some excellent extensions of these points into ever more rich phenomenology.**}

191. Another typo: solipsism.

194. "*My pulse doubled, cutting my intelligence in half. My skin went conductive. In the time it took me to drop another step, a bouillabaisse of chemical semaphores seeped up through my pores and spilled out to wet the air.*" Fine. Even better, perhaps is: "*Not only could I no longer write fiction. I could no longer live fact.*"

(You will have figured out by now, perhaps, that by typing these favorite bits into a file on my computer, I'll have them ready-to-hand [as Heidegger would say] for use in the future, thanks to the glories of word-processing. Complete with reference, so I won't have to go to the bookshelf: 1995, New York: Farrar, Straus and Giroux. Most of the good books I read these days I write a file on—this letter is the file on this book.) {**In spite of this, I somehow lost the file for this letter and had to beg a copy from Richard Powers in order to write this piece. But typing it over from the hard copy, as in the good old typewriting days of yore, was an experience I recommend to all authors; I'd forgotten how many reflections, how many editorial nudges, *don't* happen when you don't have to retype your drafts. We've lost some-**

thing important in our bargain with word-processing, but I discover that I'm not prepared to go back. Still, I'm toying with the idea of retyping the text of my next book before sending it to the publisher. A not entirely ceremonial undertaking, I bet.}

195. "*Speech baffled my machine. Helen made all well-formed sentences. But they were hollow and stuffed—linguistic training bras. She sorted nouns from verbs, but, disembodied, she did not know the difference between thing and process, except as they functioned in clauses. Her predications were all shotgun weddings. Her ideas were as decorative as half-timber beams that bore no building load.*" I'm going to send this paragraph to Doug Hofstadter, whose thinking about the role of metaphor and analogy in language is tops. Not only is this passage an apt description of the problem most often found in AI language programs, but it would serve as a wonderful test-sentence in the Turing Test. "Explain this passage, please!"

196. "*I can no more remember its otherness than I can recall the curve of a dream before its red-penciling by the Self.*" Fine.

199. "*Only by hearing it out loud, in her own voice, could Helen probe the thing, test it against itself.*" Ditto. **{This has been a theme in my recent work on consciousness. I would guess that anybody who writes lectures knows that silent reading of your own text can miss meanings (and especially problems in meanings) that blare at you as soon as you read the text aloud. Why should this be? A good theory of language production, and of consciousness, should explain such phenomena.}**

205. "*God only knows the look and feel of a sense of time without a sense of space.*" This could be the motto of "embodied cognition"—but it's too funny.

217. Good on algorithms—except that it fosters the misunderstanding (which it seems to be part of my lifework to undo) that "constellations of neurodes . . ." are **{not}** *also* executing an algorithm—a Turing-computable process, after all. **{I doubt that my typo/thinko, the missing "not," misled Powers when he read my letter. My point was that even sophisticated thinkers some-**

times forget that neural nets and other "massively parallel processors" are not doing anything that is beyond the reach of algorithms. Indeed, all the heralded examples of such neural nets are implemented on classical von Neumann machines running algorithms! Here's how I put this rant most recently in *Freedom Evolves*:

> Correctly disparaging Roger Penrose's attempt to enlist quantum physics against the dread algorithms of AI, [Paul] Churchland writes:
>
> > *One need not look so far afield as the quantum realm to find a rich domain of nonalgorithmic processes.* The processes taking place within a *hardware* [italics mine—DCD] neural network are typically nonalgorithmic, and they constitute the bulk of the computational activity going on inside our heads. They are nonalgorithmic in the blunt sense that they do not consist in a series of discrete physical states serially traversed under the instructions of a stored set of symbol-manipulating rules. (pp. 247–48)
>
> Notice the insertion of the word "hardware" here. Without it, what Churchland says would be false. In fact all the results he discusses (NETTalk, Elman's grammar-learning networks, Cottrell and Metcalfe's EMPATH, and others) were produced not by "hardware neural networks" but by virtual neural networks simulated on standard computers. And so, at a low level, every one of these demonstrations *did* "consist in a series of discrete physical states serially traversed under the instructions of a stored set of symbol-manipulating rules." This is not the level at which to explain their power of course but it is an algorithmic level. Nothing these programs do transcends the limits of Turing computability. (pp. 106–7)}

218. "*She'd cribbed the first-person pronoun as a hollow placeholder sometime earlier. 'I don't understand.' 'I want some more.'*" Perfect.

228. "*To remember a feeling without being able to bring it back. This seemed to me as close to a functional definition of higher-order consciousness as I*

would be able to give her." I've just been talking with Endel Tulving {**the eminent psychologist, perhaps best known as the coiner of the term *episodic memory***}, and am happy to say that he has come around strong on episodic memory, agreeing with me that only human beings give strong evidence of having episodic memory (in contrast with one-shot learning, which is ubiquitous in animals in ecologically felicitous circumstances). This passage tantalizes—since "bringing it back" is really, in a way, the key. Tulving now calls the kind of higher-order consciousness that invokes episodic memory *autonoetic* consciousness—really just a made up bit of jargon. In my discussions of Deep Blue with journalists recently {**shortly after Deep Blue beat Gary Kasparov, the first time a World Chess Champion was defeated by a computer**}, when they ask me what Deep Blue doesn't have that people have, I say: consider what must be involved when somebody can say, and mean, "Well, it seemed like a good idea at the time." (It can be added—"in principle"—of course. My point is just that this is what must be added.) {**As I write these further notes, Kasparov has just succeeded in tying X3D Fritz, the current computer chess champion, in a four-game series in New York City. Computer chess is immensely stronger now than it was in 1997, and such victories for human opponents will not come often in the future.**}

230 is wonderful.

248. Good on language; in general the discussion of Helen's "childhood" is excellent.

273. I was amused by your philosopher. Did you have a model in mind? (If so, you probably shouldn't tell me, since I probably couldn't keep it secret.) {**Powers didn't tell me.**}

275 is a good discussion of this issue {**operationalism, whether the structures "under the hood" matter**}.

And 286 is hilarious.

And 291 is fine.

299. Good on prosopagnosia **{inability to recognize faces by people with otherwise normal vision}**. Do you know the latest on it and Capgras delusion **{the delusional belief that one's loved one has been replaced by a duplicate impostor}**? I briefly discuss it in my most recent book, *Kinds of Minds* (pp. 111–12). Sorry. I don't have a copy of it for you as well. Maybe when I get a box of author-copy paperbacks. And of course you are under no obligation to read any of the enclosed things in any case.

Well, that's enough. There are other passages, other long stretches, that I liked as well, but this is highlights only. I've got a copy of *Gold Bug* waiting for summer.

Best wishes,
Daniel C. Dennett

Encl: *Darwin's Dangerous Idea*, review of Clark, "How to Do Other Things With Words," "Making Tools for Thinking," "The Practical Requirements for Making a Conscious Robot," "Can Machines Think," "The Virtues of Virtual Machines" **{these can all be found either in *Brainchildren* or on my website, http://ase.tufts.edu/cogstud/}**

◆

So is a novelist like Richard Powers doing science in a new, informal, "artistic" way, or is he "just" writing fiction? Bad question. Forget about classification and recognize that in the right hands, the novel is an excellent genre for pushing the scientific imagination into new places. Here's the problem: even the most acute pattern finders, the wiliest exploiters of mappings and parallels and similarities, are trapped within their personal styles of thinking, and a style is (roughly) a kit of *partially disabling* thinking-habits. You unthinkingly ignore various options because you've found them unrewarding in the past or just because you're buying some bit of the current wisdom on trust. You couldn't make progress without such ruthless pruning of the myriad branches-to-be-explored in search space, but you pay a price: you erect defenses against certain flavors of idea, censoring them before they can waste your time and energy, and some of these may be just the ideas you

need to take seriously. A novel can break through, catching the theoretician when the sentries are all off duty, during *recreation* time. But you have to be a powerful thinker to pull off the trick. That's why most science fiction doesn't repay the attention of scientists. Perhaps we need another label—scientific fiction?—for the rare novels like *Galatea 2.2* that manage to make a contribution to the scientific imagination.

But has the contribution *taken*? Do articles in the field cite Powers's novel? I don't know of any citations. I know that many people in the field have read *Galatea 2.2* and enjoyed it as much as I did, and since that is so, there is bound to be a subtle contributing effect. Why do I think so? A two-part confession will explain: I often take private delight in seeing a turn of phrase, a way of putting things, a way of *seeing* things, in the literature of AI that I am *pretty sure* stems from something I myself have written. There is no citation, nor should there be. These often aren't the sort of ideas that are big enough to plagiarize, you might say. So I just put a happy checkmark in the margin (and sometimes a cross-reference to the suspected source if I'm feeling needy). And I just as often blush in private when I return to some book or article I'd read years before and discover in my earlier underlinings and marginal comments the source of an idea I was quite sure I had discovered or formulated on my own, and published without citation. Since this has happened so often I now try to go overboard in giving credit, but I probably still fall short. (Don't you despise those footnotes along the lines of "I formulated this idea independently of X, whose earlier work was drawn to my attention when this was in press."?) Since I am not tuned to detect Powers-influence, I can only surmise that it is there in abundance, and most likely in my own work, but others may pick it up more reliably and inform us both.