Essay review

As close to the definitive Dennett as we're going to get

Tadeusz Wieslaw Zawidzki

Department of Philosophy, George Washington University, 519 Phillips Hall, 801 22nd St. NW, Washington DC 20052, USA


1. Introduction

Daniel Dennett's Intuition pumps and other tools for thinking is his most exciting, original, and rigorous work since Darwin's dangerous idea (1995), published almost twenty years ago. It is also Dennett's most sustained and explicit discussion, in print, of his distinctive philosophical method. Dennett is notorious for eschewing the standard methodology of his discipline: analytic philosophy's fetishization of formal, conceptual analysis. In this book, he defends his maverick philosophical style, largely through illustration. The book is a compendium of thinking tools, or "intuition pumps", mostly devised by him in other writings, meant to show their utility (and sometimes lack thereof) at elucidating conceptual puzzles at the heart of the subject matter to which Dennett has devoted his long and distinguished career: the place of the human mind in nature, as understood by contemporary science. However, the book is also more than a mere compendium of thinking tools. In the process of exploring his conceptual toolbox, Dennett paints a compelling picture of the nature of the human mind. This thematic duality, the simultaneous expounding of philosophical method and content, is unavoidable given Dennett's understanding of the human mind, according to which it is distinctive in its competence at inventing and wielding culturally transmitted cognitive tools. Dennett's philosophical method is tailored to the needs of human minds as he conceives them: he aspires to enrich the conceptual toolbox we use to think about our own nature and its place in the universe.

According to Dennett, human minds succeed in some domain only to the extent that they wield cognitive tools that are well crafted for navigating it. When the domain is the human mind itself, and its place in nature, successful navigation is especially fraught. The reason is that we are burdened with a cultural inheritance of poorly crafted tools for thinking about this subject matter. Unfortunately, according to Dennett, rather than discarding such tools and replacing them with better ones, many contemporary philosophers merely elaborate obsolete tools for thinking about the mind, giving them a seductively shiny, modern veneer that makes them all the more difficult to dislodge. The central, animating theme of Dennett's book is a clear characterization of the key difference between useful and counterproductive cognitive tools for thinking about the human mind. The latter all share a key defect. To use Dennett's own words, they mistake "a failure of imagination for an insight into necessity" (1991a, p. 401). The former, on the other hand, enhance our imaginative capacities, exploring what can be the case, rather than wallowing in what can't.

As an example, consider Frank Jackson's famous "Knowledge Argument" (Jackson, 1982). According to Dennett this is a counterproductive cognitive tool (2014, pp. 347–351). It asks us to imagine a color vision scientist named Mary who learns every physical, biological, scientifically expressible fact about human color vision, while unable to experience color directly herself. This is meant to "pump" the following intuition: there must be non-physical facts about human color vision, since Mary knows all the physical facts yet not what it is like to see color. Dennett calls such conceptual tools "boom crutch[es] … thinking tools that backfire, the ones that only seem to aid in understanding but that actually spread darkness and confusion instead of light" (2014, p. 14). The point is that such tools mistake failures of imagination—in this case, what it would be like to know every physical fact about human color vision—for insights into necessity—in this case, that there are facts about human color vision that science will never explain. As an example of a useful thinking tool, on the other hand, consider Dennett's thought experiment about a giant robot designed to keep your cryogenically frozen body intact until the twenty-fifth century, in the face of imperfectly predictable environmental challenges (2014, pp. 166–174). Dennett gradually proposes additions to the capacities such a robot would need to have in
order to succeed at its task, eventually concluding that only an artificially intelligent android could stand a chance. Then Dennett unmasks the point he is making. According to Richard Dawkins’ influential contemporary version of Darwin’s theory of evolution (1976), we humans are strictly analogous to such robots: we have been designed over millions of years to keep intact and promulgate our “selfish genes” in imperfectly predictable environments. Dennett’s point is that there may be less difference between artificial and natural intelligence than many philosophers assume. His thought experiment enhances our imaginations in order to make intuitive something that at first seems counter-intuitive.

Dennett’s basic insight is that there are under-explored possibilities implicit in contemporary scientific ideas about human nature that are, for various well understood reasons, difficult for brains like ours to grasp. However, there is a familiar remedy for this situation: as our species has done throughout its history when restrained by the cognitive limitations of the human brain, the solution is to engineer new cognitive tools that enable us to transcend these limitations. Just as the invention of Arabic numerals made previously inconceivable mathematical feats routine for creatures with brains like ours, Dennett hopes that the cognitive tools he has crafted for thinking about human minds, over the course of his long career, will have comparably liberating effects on our thought about our own minds and their place in nature.

This ambiguity explains the organization of the book. Seventy-seven very short chapters, each devoted to a clear and concise discussion of one cognitive tool, are organized into eight sections, flanked by a brief introductory section, and two brief concluding sections, one an optimistic admonition to use cognitive tools to transcend limits on imagination, and the other a brief note on what the book leaves out. Each of the eight meaty sections is devoted to a different aspect of the puzzle of the human mind and its place in nature. Section II discusses a dozen general-purpose thinking tools often employed in thinking about the mind, as well as in other intellectual endeavors. Sections III—V examine twenty-one tools for thinking about meaning and its place in nature, particularly in light of the idea that the human mind is a naturally evolved computer. Section VI discusses nineteen tools for thinking about evolution which, according to Dennett, is the key to understanding how nature managed to grow minds. Section VII turns to twelve tools for thinking about consciousness, and Section VIII to nine tools for thinking about free will. Finally, Section IX discusses four tools for thinking about the distinctive role of philosophy in explaining the human mind and its place in nature. Thus, the book is organized as befits a toolbox for thinking about the human mind and its place in nature: with tools placed in different compartments corresponding to different aspects of this problem.

I cannot hope to do justice to Dennett’s rich, nuanced, wide-ranging discussion in this brief essay. Dennett’s philosophical imagination and expository skill are inimitable, and I strongly encourage readers to indulge in this characteristically enjoyable read, to fully appreciate the mind-bending ideas it contains. In the remainder of this essay, I suggest some friendly amendments to an overall philosophical posture with which I am overwhelmingly sympathetic.

2. In defense of (some) “deepities”

Dennett coins the term “deepity” for apparently profound pronouncements that really say nothing cogent or useful (2014, pp. 56–57). His example is: “Love is just a word” (p. 56). According to Dennett, “deepities” are claims that appear true and profound only thanks to ambiguity: on one reading they are true yet not profound, while on another reading they are manifestly false yet would be profound if true. For example, it is trivially true that the letter string L-O-V-E is just a word. But it is manifestly false that the phenomenon of love is just a word.

Clearly Dennett is right to warn us about such uses of language. They can be and often are used to forestall useful, critical thought, perhaps deliberately, e.g., by religious authorities who feel threatened by such thought. However, I think he is too quick to dismiss such apparently mystical pronouncements tout court. Some “deepities” can draw attention to features of the human predicament that Dennett himself seems to appreciate in places. For example, because so much of our thought about the world is mediated by language, it is often hard to appreciate that language imposes certain non-compulsory structures on our experience. Dennett famously argues that our practice of expressing thoughts in language leads to the illusion that non-linguistic thought has the hard edges and systematic organization of language. For example, it could be that there is nothing determinate that one wants to consume at a restaurant until one reads the menu and is forced by the words it contains to give one’s appetites a greater determinacy than they would ever have on their own (Dennett, 1987, p. 20). Many “deepities” are deliberately constructed to draw attention to such artificial structuring of experience by language; their seemingly paradoxical contents show how some experiences cannot be fully captured using the structures of public language. Consider Wittgenstein’s injunction: “Whereof one cannot speak, thereof one must be silent” (1922). This certainly seems like a statement Dennett would classify as a “deepity”. Yet, it eloquently makes Wittgenstein’s point that some uses of language must be treated as ladders to be thrown away because they show truths about language and experience that they cannot literally, according to their own strictures, state.

Or consider Dennett’s own example—another thinking tool—of the mythical “prime mammal” (2014, p. 240). It is very tempting to assume that there must have been a first mammal. At the same time, this seems impossible, as any mammal must have mammals for parents. Our language seems to force us to come down on one side or the other: either there was a first mammal or there was not. But this categorical stricture of language falsifies the Darwinian reality: mammals evolved from mammal-like precursors that were not quite mammals. It is very hard to capture the messy, seamless processes that constitute the history of life in terms of the binary categories of language. Many classic “deepities” are designed to highlight the inadequacy of language at capturing realities that fail to parse as neatly as sentences. Consider Nagarjuna, the Second Century (CE) Indian Buddhist philosopher who arguably did for classical Indian philosophy what Wittgenstein did for modern Western philosophy. Nagarjuna was a prodigious generator of “deepities”, such as this gem: “It is empty’ is not to be said, nor that something could be non-empty, nor both, nor neither” (cited in Siderits, 2007, p. 204). According to Siderits, the point of such seemingly paradoxical claims is to make manifest the inapplicability of linguistic categories to so-called “ultimate reality”. As Wittgenstein appreciated, some things cannot be said in language, but they can be shown through strange uses of language. Or, to put it in terms of Dennett’s example, we might say that it is not the case that there either was or was not a first mammal. This “deepity” draws attention to the fact that the linguistic phrase “first mammal” cannot be applied to, i.e., either affirmed or denied of, relevant components of the evolutionary process. Indeed, in his commentary on Nagarjuna, Stephen Batchelor uses the seamless process of evolution to make precisely this point about Nagarjuna’s struggles to use language in order to explore its own limits (2000, p. 53). For both Dennett and Nagarjuna, language tends to artificially essentialize a reality that is devoid of essences, and some “deepities” are very effective at drawing attention to this fact.
Consider again the “deeply” that Dennett dismisses: “Love is just a word”. This could be used to draw attention to the fact that the complex set of emotions and interpersonal relations, to which the word “love” is typically applied is not neatly captured with a linguistic label. Conceptualizing the experience of love in terms of the label “love” encourages assumptions like: “either I love this person or I do not”. But such assumptions are arguably as misleading about the phenomenon of love as the assumption, that there either was a first mammal or not, is about the phenomenon of mammals.

3. How the computer metaphor can limit the scientific imagination

One of the highlights of Dennett’s book is his extended discussion, in Chapter 24, of the “register machine … an idealized, imaginary computer” (2014, p. 111). This is by far the longest chapter of the book, and Dennett clearly intends it as one of its most important. In fact, it contains exercises in programing register machines that Dennett strongly admonishes the reader to complete, and for which the appendix contains solutions. In a footnote, Dennett writes: “exercises are compulsory! I mean it. If you want to take advantage of this thinking tool, you have to practice, practice, practice until you become fluent” (2014, p. 121, fn.). According to Dennett, the idea of using the register machine as a pedagogical tool came from an introductory course on computers that he co-taught with his Tufts University colleague, George Smith, in the mid-1980s (2014, p. 111, fn.). As far as I know, this is his first discussion of it in print.

It is not surprising that this is the most prominent component of the book. Dennett’s discussion of the register machine encapsulates his central contribution to philosophy of mind: taking the computer metaphor employed by cognitive science to study psychological phenomena extremely seriously, in order to see how far it can take us toward resolving traditional philosophical problems about the mind. Dennett’s entire corpus is animated by the idea that once we truly appreciate the extraordinary power of the theories of information and computation, most of the traditional barriers to understanding how the mind can be a natural phenomenon will melt away. Much of his polemic against mainstream philosophical treatments of the mind consists in accusing mainstream philosophers of underestimating the power of the computer metaphor due to insufficient engagement with the actual practice of programming computers. So taking his readers through an intensive mini-course on computer programming, as he does in Chapter 24, is central to Dennett’s whole agenda in this book, and indeed his entire career. For him, the computer is the key cognitive tool necessary to unravel the mystery of the mind. Without a full appreciation of its power, nothing else Dennett says can be fully appreciated.

Dennett is surely right that the theory of computation is one of the most significant accomplishments in human intellectual history, and of enormous relevance to understanding how the mind can be a part of nature. The pedagogical skill with which he leads readers through the technicalities of programming the register machine is dazzling. One cannot help but be impressed by the amazing intellectual feats of which blind, mechanical procedures, when appropriately organized, are capable. Dennett’s discussion of the register machine brings this point home as clearly and compellingly as any I’ve ever seen. For this reason, it deserves its prominence in the book. It is exactly the kind of thinking tool that Dennett lauds: a way of enhancing our limited imaginations to see the seemingly limitless potential of a key, contemporary, scientific concept. However, it is important to keep in mind that even such powerful cognitive tools can also serve to stifle the imagination.

Thinking of the mind as a computer commits us to certain non-compulsory assumptions, e.g., that it makes sense to divide the mind into an input layer, an output layer, and an information processing layer that mediates between them. But a number of philosophers and cognitive scientists have questioned this assumption. One of the earliest was J. J. Gibson, who argued that perception required no information processing, involving instead the direct detection of environmental affordances (1986). Discussing James Watt’s steam engine governor as a metaphor for human cognition, Van Gelder (1995) argues that conceiving of this bit of intelligent machinery in terms of input, information processing, and output obscures what makes it so effective: the seamless, dynamical coupling between the mechanism and its domain. According to Van Gelder, the governor works so effectively precisely because, rather than encoding, in a first step, information about steam flow, and then, in a second step, calculating an appropriate response, triggered in a third and final step, all components of its response vary continuously with the steam flow, due to a kind of “coupling” that can be understood only using dynamical equations, that are not expressible in terms of algorithms defined over discrete, symbolic states. A number of theorists have developed such suggestions into thoroughgoing critiques of the computer metaphor, urging, in its stead, a reconceptualization of the mind as essentially embedded, embodied, and enactive (Clark, 1997, 2003; Varela, Thompson, & Rosch, 1991). This is still a dissident though vociferous minority in cognitive science, and whether it succeeds in supplanting the computer metaphor depends on as yet unarticulated, empirically constrained arguments. The point I wish to make here is simply that even the best, most imagination enhancing thinking tools can also limit the theoretical imagination in some ways. This is a limitation of the computer metaphor that must be acknowledged.

4. Cognitive tools vs. interaction tools

The notion of a cognitive or thinking tool is an extremely useful thinking tool in its own right. Dennett’s insight that much of our species’ cognitive prowess depends on inventing tools that help compensate for the deficits of our naked brains is surely on the right track. Everything from language, first oral and then written, to numerals, to various calculating technologies, to certain ritualistic practices followed by collaborating teams, like the crew of a navy aircraft carrier (Hutchins, 1995), clearly function to enhance our cognitive capacities. As Andy Clark has argued persuasively (1997, 2003), such cognitive technologies often function by transforming problems that are very difficult for biological brains like ours to process, into sequences of pattern completion tasks at which our biological brains excel. For example, algorithms for doing long division transform seemingly impossible arithmetical problems into sequences of simple tasks involving the transformation of small sets of written numerals. However, I think the focus on such cognitive tools obscures a different kind of transformative tool that our species has developed, which is not happily assimilated to the cognitive tool template. The focus on cognitive tools puts too much emphasis on our status as cognizers. But, when it comes to arguably the most important domain that human beings must master, i.e., the social domain, we are not just the cognizers; we are also the cognized. The tools we have developed to navigate the social domain are better conceived of as interaction tools than cognitive tools. They transform not just our abilities to cognize the social domain; they also transform the social domain itself, turning us into social objects that are easier to cognize and interact with.

Consider such tools as norms and the social roles they sanction. These certainly help with social cognition: we can use assumptions about how, e.g., parents, or professors, or doctors, or clerks, or police officers, etc., are supposed to act to help predict behavior (Andrews,
However, these tools work only to the extent that they transform the domain to which they are applied in social cognition. We are socialized to conform to social roles defined by the ambient norms of our cultures. Thus, such tools transform the domain of social cognition, not just social cognition. This dynamic is significantly different from the dynamic governing the sorts of cognitive tools on which Dennett focuses. For example, calculus is a cognitive tool that transformed our capacities for navigating a domain, i.e., dynamic variables, but it did not transform the domain. However, many of the tools we use to help in social navigation transform both our thinking about the domain and the domain itself, i.e., ourselves. Such interaction tools pervade human social life. Many of our self-conceptions transform not just how we think of ourselves, and others, but also how we behave. For example, thinking of oneself as a unified self, trafficking in verbally expressible reasons, arranged in practical inferences that lead to public behavior for which one is responsible is likely to transform oneself into the kind of agent that can be easily tracked using these assumptions.

The notion of an interaction tool is a very friendly supplement to Dennett’s cognitive toolbox, as it is an idea that Dennett himself entertains in various places throughout his corpus. Consider his depiction of the effects of public language on our thought: “Linguage infects and inflects our thought at every level. The words in our vocabularies are catalysts that can precipitate fixations of content … The structures of grammar enforce a discipline on our habits of thought … The structures of the stories we learn … prompt … the questions that are most likely to be relevant” (1991a, p. 301). Clearly, for Dennett language is more than merely a tool for thinking about the social domain: it is a tool for transforming the social domain in ways that make it easier to think about by language users. His views on moral responsibility have a similar flavor. The basic idea is that by holding each other responsible, whether or not we can really do otherwise, we turn each other into more responsible agents (1984, pp. 163–165 & 168; 2003, pp. 297–305). On Dennett’s view, our concepts of free will and moral responsibility are interaction tools: they help us navigate the social domain only to the extent that they also transform the persons that constitute that domain.

5. Intuition pumps are only a start

Finally, I would like to point out a methodological limitation to the use of intuition pumps or thinking tools. Dennett makes an extremely persuasive case that such tools are invaluable for enhancing our imaginations in ways that shed valuable light on traditional puzzles about the mind. However, thinking tools can be only a suggestive first step. The intuitions they pump then need to be analyzed, and the conclusions they suggest need to be supported by good arguments. As an example, consider Dennett’s “Two Black Boxes” intuition pump (2014, pp. 184–196).

This intuition pump is meant to show that semantic properties, like truth, have real causal powers, even in an entirely naturalistic, physical universe. We are asked to imagine two black boxes connected by a conducting wire, the first with two buttons, A & B, and the second with three lights, red, green, and amber. With a tiny number of exceptions, whenever A is pressed on Box 1, the red light flashes on Box 2, and whenever B is pressed on Box 1, the green light flashes on Box 2. Scientists look inside the boxes and discover two extraordinarily complex supercomputers running very intricate programs. Whenever A or B is pressed, Box 1 sends what appears to be a random bit string to Box 2. But the scientists cannot figure out what the A strings have in common such that they almost always trigger the red light, or what the B strings have in common such that they almost always trigger the green light. When they experiment by “short-circuiting” the connection, and sending a slightly altered version of an A or B string, scientists are shocked to find that Box 2’s amber light flashes. It turns out that Box 1 is an expert system that generates any one of an unlimited set of true English sentences, selected at random, when A is pressed, and any one of an unlimited set of false English sentences, selected at random, when B is pressed. Box 2 is an expert system that translates the binary string coding the English sentence it receives into LISP computer code encoding the corresponding Swedish sentence, and then checks its own database of true Swedish sentences to determine whether it is true or false, triggering the red light in the former case and the green light in the latter case. The amber light is triggered by ungrammatical sentences that cannot be translated, such as those formed unknowingly when scientists alter the bit strings that Box 1 sends to Box 2. The point of this intuition pump is to show that semantic properties, like truth, falsity, and meaningfulness, have real causal powers. The property that all the A strings have in common which explains why they trigger the red light is truth. The property that all the B strings have in common which explains why they trigger the green light is falsity. The property that all other strings lack which explains why they trigger the amber light is meaningfulness.

Although this intuition pump is vivid and compelling, it is not enough to rule out epiphenomenalism about semantic properties, i.e., the view that properties like truth, falsity, and meaningfulness do not have real causal powers. The reason is that a similar thought experiment could establish that any regularity noticeable by human beings, no matter how superficial, has real causal powers. In principle, we could rig up two similar black boxes one of which communicated possible and impossible geocentric coordinates of Mars to the other. Would this show that the geocentric coordinates of Mars have real causal powers, and hence that geocentric astronomy tracked real astronomical causes? Analogously, we could rig up two similar black boxes one of which communicated to the other whether or not some substance was emitting or absorbing phlogiston by the standard tests of Eighteenth Century chemistry. Would this show that phlogiston has real causal powers? Nobody doubts that humans can sort true, false, and meaningless sentences, and that this capacity has causal effects. Similarly, humans can sort phlogiston-emitting and phlogiston-absorbing reactions, and this capacity also has causal effects. But these trivial claims do not settle the issue of whether phlogiston and truth have real causal powers. Perhaps these are just spurious properties whose causal powers piggyback on those of real properties, e.g., oxidation in the case of phlogiston, and whatever human brains use to distinguish true from false sentences in the case of semantic properties.

The point here is that intuition pumps, no matter how vivid, imaginative, and compelling are not enough to establish philosophical claims, like the claim that semantic properties have real causal powers. To establish this, we need well-developed theories of causation, and rigorous arguments establishing their truth. In other work, Dennett seems to suggest that properties with real causal powers are nothing more than properties involved in real patterns, roughly, ways of compressing information (1991b). But phlogiston theory tracks real patterns in this sense, as does geocentric astronomy. I do not think this point is devastating for Dennett’s claim. Perhaps the way forward is to acknowledge that such patterns are real, yet not as significant as other patterns. Debates about whether or not certain properties are real or have real causal powers, which have tended to be fruitless and interminable, could then be replaced with debates about the relative significance of different, equally real patterns. Although I think this is a promising direction in which metaphysics and philosophy of science might develop, the point is that clever intuition pumps are not enough. The possibilities they suggest must be further developed via systematic theory and rigorous argumentation.
6. Conclusion

The thinking tools we owe to Daniel Dennett, aides to the imagination aimed at exploring possibilities we find counter-intuitive, rather than failures of imagination mistaken for insights into necessity, embody everything that is admirable about Dennett's intellectual character. They reveal his Yankee optimism and ingenuity, an unparalleled (among philosophers) broad and deep understanding of the key fields comprising cognitive science, an extraordinarily creative imagination, wit, literary showmanship, clarity, generosity of spirit, philosophical insight and depth, and a profound concern with the social and political implications of the philosophy and sciences of mind. Intuition pumps and other tools for thinking is, in my opinion, destined to become a classic. No other work by Dennett or about him captures the Dennettian spirit as effectively. Here, we find his intellectual temperament, his philosophical method, and his theory of the mind expertly woven together in a beautiful, engaging, and provocative tapestry.

References