Book Review
Intuition Pumps and Other Tools for Thinking

Ian O’Loughlin
Pacific University

Intuition Pumps and Other Tools for Thinking
ISBN 978-0-393-34878-1
Daniel Dennett is among the most preeminent and influential living philosophers, and his book, *Intuition Pumps and Other Tools for Thinking*, encapsulates several of his major contributions to the field, his approach to philosophy and other domains of inquiry, and his account of the role of philosophy in human intellectual endeavors. The book is written for a curious and attentive lay audience; Dennett takes himself to be instructing the reader in the use of tools for thinking that he has found particularly helpful, using examples—case studies in thinking about difficult problems—from philosophy and science to demonstrate the operation of these thinking tools. He is largely successful in this aim, introducing and deftly wielding a number of conceptual tools that are both broadly applicable and important, while at the same time giving concise presentations of several loci of discussion and inquiry in the philosophy of mind and science.

In Dennett’s presentation of his project, he cites the autobiographical work of Richard Feynman—*Surely You’re Joking, Mr. Feynman*, and *What Do You Care What Other People Think?*—as a primary influence (Feynman & Leighton, 2010, 2011). Dennett writes, “I decided to try my own hand at a similar project, less autobiographical and with the ambitious goal of persuading you to think about these topics my way” (3). While it is true that autobiographical anecdotes and sketches figure only minimally throughout, the philosophical ground covered and the approaches offered are markedly characteristic of the author; the reader familiar with Dennett’s work will recognize themes and arguments in nearly every chapter. The book thus serves nicely as an introduction to his thought and work, while at the same time introducing a number of topics of philosophic and scientific investigation, and also providing tools, notions, and patterns of reasoning—good and bad—that Dennett takes to be widely applicable in human life and inquiry.

This is an ambitious collection of goals to attain in 450 pages, even given Dennett’s admirable gift for incisive yet conversational analysis. Dennett writes, “I have always figured that if I can’t explain something I’m doing to a group of bright undergraduates, I don’t really understand it myself, and that challenge has shaped everything I have written” (12). There is something importantly right about this; anyone who teaches has, at some point, recognized deep interconnections between understanding and explanation. The book usually reads as though this bright undergraduate, interested in philosophy and in Dennett’s positions but without particularly extensive background in either, is the intended reader. However, there are also numerous discussions throughout that non-undergraduate readers—be they researchers in other disciplines, philosophers in other areas, or specialists in philosophy of mind or cognitive science—will find illuminating.

The book can be usefully divided into five main parts. The first is an introduction to several thinking tools and practices that are widely applicable, across varied contexts.
After these are introduced, Dennett goes on to discuss four particular cases—difficult topics and collections of discussions that provide the opportunity to introduce more tools, talking through intuition pumps, and wielding those thinking tools already introduced. These topics are, in order, Meaning, Evolution, Consciousness, and Free Will. Each of these discussions is rich with insights about the matter at hand, about how to do philosophy, and about how to think, and each deserves consideration.

Dennett devotes the first section of the book to general thinking tools that he has found helpful in a number of contexts. Some recent compendia of philosophical tools may come to the reader’s mind by way of comparison. To name two, Baggini and Fosl’s *Philosopher’s Toolkit* (Baggini & Fosl, 2011) or Papineau’s *Philosophical Devices* (Papineau, 2012). Although these primarily aim at philosophers in particular, each also includes helpful tools for any thinker. Sustained and accessible analyses of broadly applicable concepts—such as Frankfurt’s *On Bullshit* (Frankfurt, 2009)—can and do also contribute to the cause of philosophers offering every thinker some of the clarifying tools that have been developed in philosophy. Initiating the pattern that will then be followed throughout the later sections of the book, some of these principles and fallacies Dennett offers have been long known to philosophers but given a Dennettian gloss. Others are those that he has borrowed and adapted from other thinkers, and a few are more or less original inventions. This work—offering patterns of reasoning that may help almost anyone think more clearly and effectively about difficult topics—is work that philosophers are too often guilty of having unduly neglected. The world includes altogether too much muddled thinking, and philosophers, at least in theory, are among those best poised to help clear away these muddles.

The very first offering is a sort of panegyric of mistakes. Dennett seeks to demonstrate how important it is to err, in life, in nature, and in philosophy. Introducing a theme that is to recur throughout the book, Dennett writes, “The chief trick to making good mistakes is not to hide them—especially not from yourself. Instead of turning away in denial when you make a mistake, you should become a connoisseur of your own mistakes, turning them over in your mind as if they were works of art, which in a way they are” (22). Dennett convincingly connects this importance of error with both the study of the history of philosophy (so that we can learn from “very smart people making very tempting mistakes”) and the success of trial-and-error in natural selection and science. He is right that the importance of understanding our mistakes as such, and being willing to move forward knowing both that we have erred and that we will err again, can hardly be overstated. Also concerned with what philosophy has to offer the thinking person, Russell articulated this sentiment, writing, “To teach how to live without certainty, and yet without being paralyzed by hesitation, is perhaps the chief thing that philosophy, in
our age, can still do for those who study it” (Russell, 2013, p. 14). Dennett’s meditation on mistakes may help us to go forward in Russell’s spirit, balancing somewhere between certainty and hesitation, but his exhortations sometimes sound a more positivist timbre than is in keeping with so modest a tone. For example, Dennett commends Wolfgang Pauli’s contempt for a colleague who was “not even wrong”, apparently countenancing only errors of the false variety. Another philosopher concerned with the application of philosophy to human life, Iris Murdoch, wrote a dialogue in which her character Socrates says: “Remember, you are doing philosophy—and sometimes when you’ve been trying really hard to get a glimpse of an idea you can only talk about it in a kind of nonsense. So stop trying to be clear and just talk honest nonsense” (Murdoch, 1999, p. 472). It is at times unclear whether the honest nonsense that Murdoch’s Socrates propounds (and to my mind, rightly) counts among the varieties of error Dennett condones.

One familiar general principle that Dennett does an excellent job of presenting in this first section is what sometimes goes by the name ‘charitable interpretation’. Rather than tackling this under one heading, he articulates three specific phenomena to be mindful of when interpreting the work or arguments of others. First, he instructs the reader how to aim to use *reductio ad absurdum* without making a parody of the target. Second, he offers a list of rules for presenting your opponent’s argument. Third, he urges the reader to remember that there is much mediocre work to be had, and it is up to each of us to address (and to try to refute) only the best. Dennett’s presentation of *reductios* and avoiding parody is perfectly satisfactory, but it is his rules for interpretation that really shine. Dennett adapts these rules from game theorist Anatole Rapaport, and hence calls them ‘Rapaport’s Rules’ (33-34):

1. You should attempt to re-express your target’s position so clearly, vividly, and fairly that your target says, “Thanks, I wish I’d thought of putting it that way.”
2. You should list any points of agreement (especially if they are not matters of general or widespread agreement).
3. You should mention anything you have learned from your target.
4. Only then are you permitted to say so much as a word of rebuttal or criticism.

It is hard to imagine the extent of the transformative effects these rules could have on our discourses—as philosophers and as human beings—if they were widely followed. Straw person fallacies abound, both inside and outside academic circles, and they are most common when the stakes are highest. In decades of paying attention, I have literally never heard any politician, from any party or nation, characterize their opponents’ views
without committing a straw person. I have seldom heard this even from laypersons, characterizing those political views with which they disagree. If we are interested in using reason to make the world and our understanding of it better—and we ought to be so interested—then a good first step is the strict application of Rapaport’s Rules.

Another aspect of selection and interpretation that Dennett wants to draw our attention to is what he calls “Sturgeon’s Law”, after the science fiction author Ted Sturgeon, from whom he borrows and adapts this dictum. In Dennett’s words, Sturgeon’s law is simply that “Ninety percent of everything is crap” (36). That is, any target domain—from television shows to theology, and certainly including philosophy—includes plenty of material that rates mediocre at best. “Let’s stipulate at the outset,” Dennett continues, “that there is a great deal of deplorable, stupid, second-rate stuff out there, of all sorts.” To characterize the whole of anything by reference to its lesser instances is misguided, and if we ever find ourselves saying “the whole of X is stupid”, we ought to pause. Dennett is right to point out these dangers, and “Sturgeon’s Law” provides an effective rejoinder to any pronouncement of universal stupidity. This goes for the work of philosophers as well, as Dennett points out in a lesson any newcomer (and many non-newcomers) would do well to learn.

The very best theories and analyses of any philosopher, from the greatest, most perceptive sages of ancient Greece to the intellectual heroes of the recent past (Bertrand Russell, Ludwig Wittgenstein, John Dewey, Jean Paul Sartre—to name four very different thinkers), can be made to look like utter idiocy—or tedious nitpicking—with a few deft tweaks. Yuck, yuck. Don’t do it. The only one you’ll discredit is yourself” (37).

In both of these cases, the stated importance of owning mistakes, and the stated importance of interpreting charitably, however, it is not always clear that Dennett follows his own rules. Indeed, he explicitly (and wisely) admits that he often struggles to apply Rapaport’s rules, and in the course of the book there are interpretations that fail to be charitable. For example, when Dennett is characterizing the argument in favor of zombies, in the section that addresses consciousness, he describes the way that we might think that we already know that a person is conscious, from our interactions with them, and then goes on to write:

Some philosophers think that your imagination would be playing a trick on you if you fell for this “merely behavioral evidence of consciousness and jumped to that conclusion. “Don’t fall, don’t jump!” might be their motto (284).
However, Dennett grants elsewhere that even according to the zombie theorist, such a lack of consciousness might be nomologically impossible even if it were logically possible. If this is so, then it must violate Rapaport's first rule to present the view as though anyone who takes philosophical zombies to show something about consciousness also thinks that to ascribe consciousness to others is to fall for a trick. As it happens, I tend to be sympathetic to Dennett's ultimate position with respect to zombies—I think he is right that this is a basically confused, albeit tempting, notion, and I sometimes borrow Dennett's arguments to show this—but I also suspect that the pro-zombie philosophers of mind among us would not respond to the above characterization with, “thanks, I wish I'd thought of putting it that way.” If zombies were logically but not nomologically possible, acceptance of the qualia of others would be no more egregious an inference than any other that is grounded in the nomological character of the world. Uncharitable moments such as this make at least occasional appearances throughout the book, and although, in keeping with his exhortation to do so, Dennett does present some of his own past mistakes, he not infrequently seems keener to present the mistakes of those with whom he has disagreed. Perhaps shortcomings such as these are all but unavoidable.

For the most part, Dennett articulately and convincingly makes the case that these general thinking tools are valuable and graspable. Although in some cases his injunctions fall flat—his two-page crusade against uses of the term ‘surely’ (53-54) can come off as overwrought (surely this term isn’t universally a sign of unreasoned bluster)—in others the warnings he gives about common pitfalls are important and clearly sounded. At one point Dennett coins the term ‘deepity’ to denote a statement that gives the impression of wisdom or profundity, but only by virtue of its ambiguity (56-57). Once the statement is pressed for clarity, the apparent wisdom disappears. From the vague utterances of politicians to the platitudes of self-help books, shallow, ambiguous claims are all too often accepted as true and important, and curing these and similar ills should be the special province of the philosopher.

After the initial section introducing general thinking tools more are introduced, by way of sketches of some of the main questions and positions in four areas of philosophical inquiry. The first of these sections, and the longest, offers tools for thinking about meaning and content. This case study is the most developed, and serves as a nice introduction to one of Dennett’s main offerings—the titular “intuition pumps” themselves—but is also perhaps the least effective as an introduction to philosophical discussions about the topic in question.

‘Intuition pump’ is a term coined by Dennett in his 1980 response to Searle (D. Dennett, 1980). These are thought experiments that are designed to illuminate and amplify a par-
ticular intuition about difficult cases. Dennett’s original example of an intuition pump is Searle’s Chinese Room thought experiment, which he again deftly addresses later, in the section on consciousness. In the section on meaning, it is Dennett’s overarching aim to motivate a broadly Quinean, roughly deflationary, interpretationist view of intentionality, convincing the reader that “meaning isn’t going to turn out to be a simple property that maps easily onto brains, and we’re not going to find “deeper” facts anywhere that just settle the question of what a sentence, or a thought, or a belief really means” (197). It is not that Dennett’s arguments in this section fail—on the contrary, I find myself mostly convinced—but rather that his audience blurs. Philosophy of language is a notoriously difficult subject for an introductory audience to approach, and unless Dennett’s “bright undergraduates” are already fairly well versed and motivated in this area in particular, even the more than one hundred pages Dennett spends on meaning will probably fail to bring them up to speed. Conversely, much of the argumentation and considerations herein are too familiar to be persuasive to any professional philosopher who is not already basically in agreement. As a treatise on meaning and the right way to frame it, this section has something of a difficult time finding its audience.

This being said, the section on meaning still includes, along the way, a number of illuminating discussions and helpful tools, which means that it is a success by most of Dennett’s own lights. His brief expositions of a Quinian holism about belief (65-68) are impressively perspicacious and accessible; these make good use of what he introduces as the ‘sorta’ operator, as he insists that the only answer to whether a very young child really believes that her daddy is a doctor is that she sorta believes this—and theoretical perspective that forces us to draw a bright line that divides whole beliefs from non-beliefs must be flawed. His concise introductions of the distinction between the manifest and scientific images (69-72), and of folk psychology (73-76), are as effective as any of the lengthier versions of these to be found in introductory philosophy of science and philosophy of mind texts. His use and presentation of a crossword with two equally good solutions (176) is a brief, brilliant, efficacious introduction to the concept (and some of the consequences) of indeterminacy.

In the middle of this long section on meaning, Dennett strays into a forty-page interlude on computers and computing (107-150). Although it may try the patience of some among his audience, this interlude could stand on its own as an important reading in a theoretical computer science course, or a course on artificial intelligence, and any reader willing to patiently work through his use of register machines (which lead up to an explanation of Turing machines), his presentations of conditional branching, virtual machines, and algorithms will find themselves in possession of a more whole and sharp grasp of the basics of computing. These do contribute to his later discussions of meaning
and consciousness—especially through the graded mindlessness he claims is available on these bases—but the sudden shift to theoretical computer science does not do anything to help the confusion about the intended reader that this part of the book already bears.

The next three sections, on evolution, consciousness, and free will, are shorter, only summing in total to about the same number of pages as the previous section on meaning. While evolution is the least orthodox of these as a subject of philosophical inquiry, Dennett does nice work using discussions of evolution to highlight several philosophical tools and ideas. He also, in this section, treats evolution itself philosophically, examining how thought experiments and intuition pumps can accelerate or impede our acceptance of evolutionary explanations.

One of the impediments to understanding evolutionary explanations is that evolution deals in such alien scales. This is true of many things; easily conceivable scales of time and space occupy only a narrow margin among the possibilities. In order to lend credibility, comprehensibility, to these vast and vanishing probabilities, durations, and spaces, Dennett borrows the Library of Babel from Jorge Luis Borges (Borges, 1964). Borges’ beautiful descriptions of the library of all possible books—which are already philosophical in tone—are borrowed by Dennett as a metaphor for every possible genetic combination (205-213). This conceit is effective in its aim, but more importantly the imagery is used to make otherwise incomprehensible quantities, scales, and probabilities comprehensible.

Another nicely crafted use of a philosophical distinction (and once more, a distinction which is characteristic of Dennett's work) is Dennett's emphasis, in the course of explaining evolutionary explanations, on the difference between competence and comprehension. Citing the notorious prime-number year cycles of cicadas, Dennett makes a good case that the primeness of these numbers figure into reasons for the cicadas' behaviors, even though the cicadas remain starkly unaware of said reasons (236-237). Even in the absence of any reasoners, these reasons would remain, argues Dennett. Darwin, for Dennett, gave us grounds for teleological explanations, rather than eliminating the need for these, and Dennett's use of uncomprehending but competent behaviors, like those of the cicadas, make a good case for this.

One last discussion in this section worthy of mention is Dennett's presentation of “Mitochondrial Eve” as a helpful way to think about both speciation and retrospective ascription (247-251). In a sort of Anscombian, Davidsonian twist, whether a new species has begun is a fact that can only be established in retrospect. After all, any one of us, under the right circumstances, could be the forbear to what comes to be understood as a new
species, just as any female human who is an ancestor of any living human (and who is also not earlier than the present mitochondrial eve) could, by virtue of just the right happenstance deaths in the present, become the mitochondrial eve herself—that is, the most recent woman who is a direct ancestor to all living humans. Dennett astutely talks the reader through these retrospective ascriptions—indeed, the reader is left wondering whether there is not more to say here, whether the kind of mid-range ontological robustness these ascriptions are purported to confer could not be used fruitfully in action theory, or in questions of moral responsibility. Whether or not someone has been dishonest, for example, may in some cases be defeasible in time in the same way that someone’s status as mitochondrial eve is defeasible in time, if intentions, like ancestry, can only be properly read diachronically.

In the section on consciousness, Dennett employs several familiar arguments. As stated above he argues against Searle by “turning the knobs” on the intuition pump that is the Chinese room, trying to establish the truth of something like the systems reply, and he also argues against the “zombic hunch” (which he admits that he can feel, but tries to suppress) by stretching and investigating the notion that anything at all could really be behaviorally indistinguishable from us but lack qualia. These discussions provide clear introductions to Dennett’s thought, and on matters that have been central to his interest and influence as a philosopher, for any reader seeking these. They also provide conversational introductions to the topic of qualia itself, and to some of the main points in the surrounding literature. For those readers already familiar with Dennett’s work, much, but not all, of the consciousness section will be readily recognizable, but even so, several discussions present Dennett’s mature summaries of these ideas in clear and convincing terms.

One such section is Dennett’s discussion of the self (333-340). Dennett conceives of the self as “the center of narrative gravity.” That is, the way that a center of mass is a theorist’s fiction, but also a theorist’s fiction which genuinely helps to explain the phenomena of the world, is just the way that the self, as an organizing center point of our experience, is at the same time a genuine part of the world and also a fiction. This is a position that Dennett has taken before, and that he has carefully revised and crafted over several decades (D. C. Dennett, 1989, 1992). This latest rendition, however, is markedly clear and persuasive. Dennett starts from a Humean absence of self, and then asks the reader to consider the ways that when it comes to fictions, some usually straightforward questions may simply be misguided (whether, for example, Sherlock Holmes is taller or shorter than the conductor on the train to Aldershot). In this same sense, he argues, we do not ask what a center of mass is made of, and we ought not ask where the self is located. The self, for Dennett, is a theorist’s fiction, “posited in order to unify and make sense of an
otherwise bafflingly complex collection of actions, utterances, fidgets, complaints, promises, and so forth, that make up a person” (336). Dennett is clear that this posit plays an important role in our understanding and explanations of the world:

> It may be a theorist’s fiction, but it is a very valuable fiction from which a lot of true predictions can be generated. Can such an abstract entity, having no material existence, actually cause anything? Not directly, but explanations that cite a center of gravity compete with explanations that are clearly causal. Why didn’t the coffee mug tip over when the sailboat heeled so drastically? “Because it has an unusually low center of gravity” competes with “Because it is glued to the deck” (335).

The perspicacity and succinctness with which Dennett delivers this intersection of complex notions—prediction, understanding, causation, explanation—is characteristic of many of the better moments in his book. Moreover, the apparent simplicity of his explanations is clearly driven by serious and honed philosophical consideration. Some of the brief considerations he offers, seemingly in passing, sketch deep-rooted arguments. Elsewhere in the same section, Dennett asks the reader to consider the difference between a question like “Have you ever been to Paris?” and a question like “have you ever broken a white coffee mug?” We know the answer to the former immediately, and it seems almost unfathomable that anyone would hesitate in this. The latter, however, may even be hopelessly unanswerable (and interestingly, it once more seems somewhat alien to imagine someone answering this immediately). Dennett argues that this difference lies in the narrative nature of self: we shed those experiences that fail to be important to our narrative, and retain those that succeed.

Indeed, it is difficult to see how any theory of the self that is not grounded in narrative could account for these differences. This is especially so since we can imagine world that supported the opposite reactions to these questions. Even in this world, perhaps, we can imagine someone growing up on the outskirts of Paris, near its municipal boundary, but being raised by parents who engendered a particular reverence for white coffee mugs. In such a case, someone might turn out to hesitate at the first question but answer the second immediately—even in the negative. If this is so, then it is not any intrinsic properties of these experiences that informs their status as available elements of autobiography, but rather the role they have played in the stories that we tell ourselves and others.

Dennett’s last case study is a relatively short, but effective and illuminating discussion of free will. This is perhaps the most successful of the four case study discussions, given his aims regarding talking the bright undergraduate reader through a difficult topic while at
the same time demonstrating the use of careful, philosophical thinking. He builds a case for compatibilism, utilizing several of the distinctions he has already introduced: between causes and reasons, between personal and subpersonal explanations, and between intentional and design stances. Although such a discussion could easily descend into being too technical for an introductory reader or too simplified to be of interest to the veteran philosopher, Dennett's treatment avoids both. He argues adroitly, using analogies from chess-playing programs, that there is an important sense in which statements about what could have happened, even where strictly speaking the system is deterministic, can be true or false.

To motivate this argument, Dennett spends several pages discussing Conway's Game of Life (359-369), and the way that a simple and fully determined system can lend itself to various and real levels of description, as well as to genuinely unexpected discoveries and inventions. Despite the fact that any given Life world already contains its entire future in its initial position and deterministic (and very simple) laws, there are levels of description at which "marvelous and unanticipated" things happen. There are high-level, robust patterns that can bear the weight of counterfactual claims. This allows us to say of one sophisticated chess program that it may have castled instead—because there are similar situations in which that very program would have castled—but to deny that of some less sophisticated program, which in no similar cases would have done so. Together, these provide a basic defense of compatibilism.

In general, Dennett's book is a success. There are discussions and tools that will be helpful to the bright undergraduate who is interested in philosophical topics like consciousness and free will. There are passages and presentations that will be helpful to the philosopher who is seeking to be better acquainted with Dennett's arguments and methods. There are explanations and sketches that will be helpful to the non-philosopher specialist who is interested to better understand how philosophical approaches are brought to bear on these interdisciplinary topics. Perhaps miraculously, there are sometimes sections that manage to be all three of these things at once. Most importantly, perhaps, these discussions set a precedent that is important for philosophers to bear in mind: much of what we do, as philosophers, is supposed to concern thinking clearly about difficult topics. We take our field, and the skills it cultivates, to be broadly applicable in the world. Thinking clearly about difficult topics is something that is not always done well, and Dennett is right to want to articulate the tools that he has used, and continues to use, in the hopes that readers might apply them to a wide array of tasks, problems, and discussions, inside and outside philosophy. Philosophers are well positioned for this work, and Dennett's book is a welcome example. Anyone who has spent a lifetime learning to think through difficult topics should have something to share with those of us who are
still trying to think through difficult topics—which is everybody—and perhaps the best way to share this, like Feynman (or for that matter Proust, or Montaigne), is to lead the reader through those very thickets that formed the writer’s own path. Dennett should be commended for doing just this, and for doing it with characteristic spirit and clarity.

References


