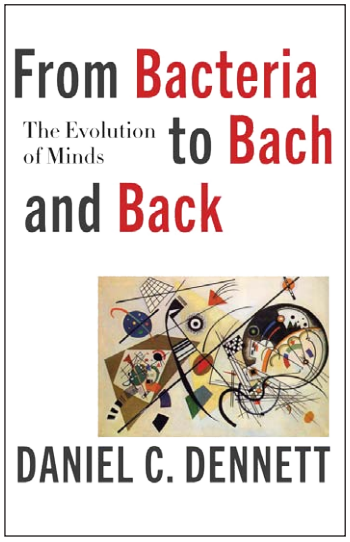
Explaining Top-Down Minds from the Bottom Up

Sven Delarivière • Vrije Universiteit Brussel, Belgium • sven.delariviere/at/vub.be

> Upshot • The main topic of Dennett's book is intelligent design and the design of intelligence, trying to make intuitive the processes of both, be it the top-down process of comprehension that designs with foresight and reasons or the bottom-up process of evolution that has, through blind trial and error, captured free-floating rationales and ultimately, through co-evolution (between memes and genes), achieved top-down intelligence, flipping its original design process upside down.



Review of *From Bacteria to Bach and Back: The Evolution of Minds*
by Daniel C. Dennett, W. W. Norton & Company, London, 2017.
ISBN: 978-0-393-24207-2
496 pages

From Descartes to Darwin and Dennett

"How come there are minds?" (3) This is the intriguing question with which Daniel Dennett starts his book and the 15 subsequent chapters are an attempt to answer that question with a scientific, materialist and Darwinian theory of how one could get from mindless and mindlessly designed competences of the likes of bacteria to the genius design-processes and mysterious consciousness of the likes of Bach. You will not actually find much about bacteria or Bach in this tome, but the subtitle *The Evolution of Minds* is a good description of what Dennett sets out to elucidate.

To start, he sets up the antagonist of the piece: Descartes. This is not to discredit or vilify the 17th century philosopher and scientist, but to label alluringly intuitive philosophical ideas that can be associated with Descartes, but which Dennett considers to be outdated and in need of revision. The ideas in question are: Cartesian dualism (i.e., the idea that the mind is so special it must be distinct from matter and free of the laws of physics), the Cartesian theater (i.e., the impression that consciousness is a theater-production set up for a comprehending homunculus) and the presumed primacy of the first-person point of view over the scientific third-person point of view in exploring consciousness. While it is rare to see anyone defend any of these Cartesian ideas outright, Dennett attempts to show (and then combat) the ways in which these intuitions nonetheless sway one's thinking. So strong are they that he describes the pull towards them as "Cartesian gravity" (17).

If Descartes is the antagonist, then who is the protagonist? Undoubtedly, Darwin. For it is by Darwinian evolution that Dennett crafts a story that accounts for the hows (and whys) of the design of bacteria and Bach, gazelles and Gaudi, termites and Turing, as well as termites' castles and Turing's computers. The word "design" is crucial (and not metaphorical) here, because Dennett explores what goes into various *design processes of* and *from* these organisms. Each of them boasts a beauty and/or competences that invites us to consider the *reasons* for their design. Exploring the various research and development (finding and exploiting good designs) processes involved in achieving the competent and/or comprehending designs is what Dennett sets out to do here, ultimately culminating in a bottom-up explanation of top-down intelligence.

Competence without comprehension

We humans have created artifacts, knowing beforehand what they were set to achieve and using our creative intelligence in implementing their competence. From this perspective, there is a powerful urge to think that where there is a design, there must be an intelligent designer and that where there is a reason, there must be a reasoner. Dennett tells us, however, to resist both urges and tries to make intuitive the notion that competence can come without comprehension and reasons can come without a reasoner.

Dennett casts the "competence without comprehension" (57) creed in two ways, two strange inversions of reasoning: "Darwin's strange inversion" (4) is that in order to design competently, one need not *know how* to do so and "Turing's strange inversion" (55) is that in order to be competent, one need not *know how* to be (this due to Turing's conception of machines that can compute competently without knowing any arithmetic). Nature can design a competent animal without comprehending the specs or implementation required to do so. What design process, what process of research and development, can account for it then? The answer is, of course, evolution by natural selection. Nature has no foresight or goals to

solve problems with, but it has a lot of time (and no shame) to exploit things that almost never happen in a trial-and-error process that sticks with what works and discards the vast array of things that do not. This makes it a most excellent designer. And a design, be it an animal (designed through natural selection) or a Turing machine (designed through Turing's genius) can *be* competent without comprehending how or why it is competent. It does not need to comprehend its own competences any more than an elevator needs to know how or why it works. Nonetheless, there is a reason for it to work.

It is a safe bet that all parts of an organism are good *for* something, that there is a *reason* why it got that way. Figuring out what the design is good for (reverse engineering) is not about figuring out the foresight that a comprehending Mother Nature had but appreciating which free-floating rationales the process of natural selection has uncovered (by a successful mutation) and exploited (by natural selection). Hence the reasons are not those of Nature or those of the designee. There is a reason why a gazelle stots, but that reason is appreciated neither by Nature nor the gazelle. One of the marks of human intelligence is that we *can* appreciate reasons, meaning we can recognize their value and be persuaded by them and so design with goals, cost-considerations and some foresight, or, in short, with comprehension.

On the meme-way to comprehension

Dennett warns us against the general tendency to over-credit human genius or comprehension "as if all of human cultural evolution took place near the ceiling of comprehension, and all involved objects *valued for good reasons*" (308). As an example, he cites the Polynesian canoe. Each canoe is copied from previous ones, though not perfectly, and the canoes that are least likely to be taken by the sea, the "fittest" canoes, will be most subject to copying. Here it is clear that the R&D process responsible for the boat's culminating design is not human foresight but natural selection. While boatbuilders may bring theoretical knowledge to the table, these need not have played a role in the success of design. Similarly: how many times have you made a move in chess of which you could only appreciate the brilliance afterwards? Nonetheless, the ability to appreciate a brilliant reason and be persuaded by it for future action is the essence of comprehension and many human achievements are due to comprehension. The design of Gaudi's *Sagrada Família* was carefully planned before erecting, its research and development did not stretch through years of bottom-up design via natural selection,[1] but was due to the top-down intelligence of Gaudi.

Armed with his creed of "competence without comprehension," Dennett sets out to continue his Darwinian explanation of minds such as those of Gaudi (and our own). First, it must be noted that our faculties of comprehension must ultimately be rooted in our physical architecture, the brain being the most noteworthy part of that. It will not do to explain comprehension by delegating all of the understanding to another comprehending homunculus leading a well-oiled bureaucracy or even a comprehending area of the brain. This merely shifts the problem. From Dennett's physicalist perspective, if top-down comprehension is to be accomplished by brains, it must be rooted in billions of neurons without any top-down control system. Top-down comprehending skills are to be decentralized, distributed control by semi-autonomous neurons fending for themselves. The answer to how this could be accomplished can once again be found from the perspective of Darwinian evolution. But not by genes alone. To account for the comprehending skills of the human mind, genetic evolution is neither powerful nor fast enough to suffice. For that, we must consider a new replicator: the meme. A meme is, in essence, "a kind of *way of behaving* (roughly)" (206) not passed on via genes, but nonetheless passed on and indeed striving to be passed on. Memetic evolution, Dennett argues, shares enough similarities with genetic evolution for the Darwinian perspective to be illuminating and he goes to great lengths (devoting a separate chapter to it) to bring this point home, exploiting the similarities to (and differences from) genetic evolution.

1| Though it may have involved some aspects that did evolve gradually. Gaudi did not invent architecture from scratch, after all.

The ability for memes to invade at all must have been due to a genetic adjustment such as the instinct to imitate one's role models. Memes could be anything from wearing your baseball cap backwards or a catchy tune to the utterings of language. It will not surprise the reader that it is words that are the most powerful among them. The ancestors of today's well-designed languages were probably inefficient, hard-to-learn behavioral patterns that were parasitic habits, infectious and hard to shake, without any reward beyond replication. Like viruses, it is the most infectious "ways of behaving" that are the fittest, be they harmful, neutral or beneficial to their host. But *if* beneficial to the host – for instance as signalers of affordances the host was already equipped to discern – then the genetic mutations that favor a symbiotic relationship with memes would have been selected for, bootstrapping the meme's influence in a process of coevolution that accommodates the brain to its memes and vice versa (much like software has co-evolved with hardware).

Comprehension and consciousness

Early memes did not require any comprehension in shaping them or even employing them (they could be used even without recognising their benefit, as the example of copying boats also made clear). But gradually, memes evolved into thinking tools that power minds that do not only benefit from using valuable memes, but also notice their own meme use (self-monitoring) and can appreciate and question their valuable talents in meta-memes, which are once again ready for reflection. Our reasons are now no longer free-floating, but part of us, ready to use and think about, enabling us to plan ahead, thus effectively flipping the bottom-up R&D process of design into a top-down one. Today, all those memes that gradually evolved "have to compete with memetically engineered inventions, crafted with foresight and purpose by professional memesmiths" (331).

It is not only our words and plans that are subject to monitoring and reflection, but also ourselves. Unfortunately, however, our cognitive processes come from neural activity into which we have no more privi-

leged access than we do to our own digestive processes and trying to understand our own minds and comprehension via bottom-up neuroscience is as difficult as explaining the competences of your smartphone's app by deciphering its hardware circuit design and the bit strings. There is a solution, however, that works for both apps and brains: a user-illusion interface. In smartphones, this was the product of intelligent designs over many years, but in our brains it is a meme-complex, a functional shorthand of what is going on, designed through genetic and cultural evolution for (and through) interacting with oneself and others. Where, then, is the user in the user-illusion? It is easy to fall into the trap of conceiving of our subjective experiences as private presentations rendered in a Cartesian theater for an audience member, an understanding homunculus or a dedicated portion of neural circuitry, to process the presentation and make it conscious. Dennett explains that the user is itself part of the illusion, to be broken up and distributed into the lesser agencies and competences of the brain and accomplished by a complex of memes riding on them (in much the same way that apps run on hardware).

So where do the subjective experiences come in, the *something it is like to be you* (Nagel 1985)? What, in short, about the hard problem of consciousness (Chalmers 1995)? There is no such problem, replies Dennett. It arises only because we mistakenly presume that our ability to describe what it is like to be oneself is *due* to there being an inner subjective and private presentation to be described and explained. This is a mistaken assumption. If our subjective experience were to cause our judgements, then we would regress the problem to an internal theater that is being described (which leads to an infinite regress of theaters). Instead, Dennett argues that the opposite is the case: the judgements and meta-judgements we make constitute our subjective experiences and not the other way around. This belongs to the last strange inversion of reasoning, one attributed to David Hume, who noted that our impression of causation comes from inside (as a habit of expectation), not outside (as a direct experience). "It is how our brains respond that causes 'us' […] to 'project' an illusory property into the (manifest) world" (356). So, sweetness, for instance:

> "is not an internal recreation or model of [glucose's] chemical properties, nor is it a very special property in our non-physical minds that we use to decorate the perceptible things out there in the world. It is no property at all; it is a benign illusion. Our brains have tricked us into having the conviction, making the judgment, that there seems to be an intrinsically wonderful but otherwise undescribable property [...]: sweetness." (356)

And in that very line, our judgements about ourselves form our Selves. There is only something it is *like* to be you because you are able to tell what it is like to be you (to describe your judgements, assertions and emotional reactions).

Discussion

Recounted here is merely the narrative thread of Dennett's story, which in the book he strings together using both tools and empirical data from various fields such as computer science, neuroscience, linguistics, biology, and chemistry. It is always a joy to read philosophers who have not only found a way to express their intuitions, but have also done their homework. As, of course, one should have, if one wants to tackle the grand theme of the book: the evolution of minds. What is especially interesting is that the book manages to both laud the human mind as a cognitive marvel and rein in overstated claims of the human mind as a cognitive miracle by exposing the R&D processes that are responsible for our mind and our mind's achievements. There are, however, definitely some passages that might rub the reader up the wrong way. Consider this quote:

> "some would say […] it is a mistake to think of the 'things' we interact with and manipulate, and fall in love with, as reality. That's a defensible position, I suppose. In fact, it's a *version* of what I have said about the manifest image of each species: a user-illusion brilliantly designed by evolution to fit the needs of its users. My version differs only in being willing and eager to endorse these ontologies as ways of carving up *reality*, not *mere* fictions but different versions of what actually exists […] The more shocking way of saying it – 'we live in a fictional world, in a dream world, in unreality' – tends to cause confusion because it suggests we are somehow victims, duped by some evil force." (222)

Clearly, Dennett's adherence to "realities" does not fit a constructivist narrative. And yet it does not seem to me that he differs on crucial claims. Let me point out that Dennett is not here arguing why we should adhere to a mind-independent reality, but expressing his eagerness and willingness to use the word "reality" over alternatives (such as "fictions") that could easily be associated with deception. So instead, he just reins back the ontological weight of the term, making it system-relative and pragmatically useful, albeit at the cost of being misunderstood by constructivists. Dennett may not couch his terms in constructivist language, but his ideas seem compatible. Allow me to unpack that by addressing his rejection of (a) a passive role for knowledge, (b) a reflection correspondence approach to representation and (c) its relation to "reality."

First of all, mental structures are not construed as passive. Dennett's evolutionary perspective makes him eager to consider every aspect of an organism as an element of design that is good for survival. He mentions his long-standing criticism of traditional AI creations as too "disembodied, 'bedridden'" (158) to be autonomous in their competences. And if mental structures are to be good for survival, they had better play an active role. Dennett considers an organism's "ontology" and defines this as the "set of 'things' that an animal can recognize and behave appropriately with regard to" (60). Compare that with the constructivist conception of representations as system-relative "mechanisms which allow the generation of adequate behavior" (Riegler 2001: 4). Both quotes implicitly rely on "knowledge" being knowledge only if the organism or system is in some sense active. And in many cases, even an active representation would be superfluous design, which can be accomplished without it – much like Brooks's (1987) "Creatures." One could equally call part of Dennett's story competence without representation.

Dennett does, however, consider humans to have explicit representations but, crucially, they are not construed in any literal or reflection-correspondence sense of the word. Something can be said to be representational only if it triggers an appropriately complex process and its implemented competences, not because any neuron clus-

ter correlates with external events or because there is an inner show put on that is literally like or in some sense homomorphic to the world. In this sense, reading Dennett suggests a manner in which Humberto Maturana's (1978) metaphor of the pilot could be further improved. Maturana compares the situation of the mind to a pilot managing to fly a plane, not by looking "outside" (it could even be windowless), but by responding appropriately to the activity of the instrument panel. While Maturana's metaphor does a good job of making intuitive that the outer world cannot and need not be directly and neutrally perceived from the cockpit, it also has two related limitations. First, it's a bit unfortunate that actual pilots leave their cockpits after their shift is done, entailing that they have an acquaintance with the world outside the cockpit that is not due to their instruments and which may furthermore direct their interpretations of what their instruments are purported to indicate. This may be a trivial thing to point out, but Dennett (2013) has a thought experiment where he uses a closely related metaphor of being trapped inside a control room of a giant robot with no windows, unlabeled lights flickering and the difficult task of keeping the robot alive.

> "Your predicament here is none other than your brain's predicament! […] And your brain can't open a window in your skull and look out to see what's happening that might be causing the patterns of signals streaming into your virtual cortex." (Dennett 2013: 103)

Except, Dennett continues, your brain is not in *exactly* the same predicament. There is no homunculus inside our heads that encompasses our (comprehending) self, and positing such a homunculus merely shifts the problem of accounting for what makes up this self. This brings me to the second limitation: the metaphor splits the cognitive process into (a) a presentation (the output of the instruments) for (b) a black box responder (the pilot, taking the instrument-readings as input). To account for the cognition of the black box, it itself needs to be split into a tinier presentation for a tinier black box that must accomplish all of the same work. How does the pilot-homunculus know how to read and respond appropriately to its instruments? Presumably, it too will be operationally closed. Is there a tinier pilot within its head, who has tinier instruments, and no window, to find out what happens in the bigger cockpit? And how does that tiny pilot know how to do it? In the end, it would be pilots all the way down. Dennett might adjust the metaphor by replacing the pilot with a committee of pilots, each responsible and competent in only a minor component of the whole cockpit. This new metaphor further emphasizes that there is no need for a representation of the outside world. The inside of the cockpit is successful at navigation, not because there is a clever homunculus that processes content derived from the outside world, but because of the way in which the cockpit is structured, breaking down the success of the entire plane into smaller competences.

Now, whether the instruments would in any way be displaying anything "true" or corresponding to an outside world is quite simply not part of Dennett's evolutionary story. Evolution is not a process for creating creatures that get things right, but that survive (and multiply). Remember Hume's inversion of reasoning, invoked to make clear that properties are projected from inside judgements, not received from the outside. Additionally, Dennett's account of consciousness, in particular, seems relevant here, as it is not a (re)presentation of the external world, but a construction useful for staying alive, a cognitively generated illusion that is useful, not literally true.

Getting to explain the nature and evolution of minds is a long and winding road, full of technical nuance to expound, conceptual richness to impart and deep intuitions to shake, and Dennett does a fine job of leading the (willing) reader along that difficult road. Many of the stops he makes have been made elsewhere in some form or another, as the book seems to cover all of Dennett's usual themes (evolution, consciousness, qualia, artificial intelligence, memes, etc.) but they have been updated with the support of recent research and rephrased to fit the overall perspective of this book. This wide theme may actually be both the book's strength and its weakness, as I have to admit that the result is weaker than his elucidations on each topic individually. It is not as focused and rich as the brilliant elucidation of consciousness in *Consciousness Explained* (Dennett 1993), or as packaged as his "best-of" compilation *Intuition Pumps and Other Tools for Thinking* (Dennett 2013), nor as lyrical and enticing as *The Mind's I*, where he and Douglas Hofstadter (1985) collected and commented on diverse texts of authors with whom they both agree and disagree. Nonetheless *From Bacteria to Bach and Back* makes lucid each part of the path towards minds, even if it sometimes fails to impart the full map of the journey.

References

Brooks R. (1987) Intelligence without representation. Artificial Intelligence 47: 139–159. ▸ **http://cepa.info/4059**

Chalmers D. J. (1995) Facing up to the problem of consciousness. Journal of Consciousness Studies 2(3): 200–219.

Dennett D. C. (1993) Consciousness explained. Penguin Books, London.

Dennett D. C. (2013) Intuition pumps and other tools for thinking. W. W. Norton, New York.

Dennett D. C. & Hofstadter D. R. (1985) The mind's I: Fantasies and reflections on self and soul. Penguin Books, London.

Maturana H. R. (1978) Biology of language: The epistemology of reality. In: Miller G. & Lenneberg E. (eds.) Psychology and biology of language and thought: Essays in honor of Eric Lenneberg. Academic Press, New York: 27–63. ▸ **http://cepa.info/549**

Nagel T. (1985) What is it like to be a bat? In: D. C. Dennett & D. R. Hofstadter (eds.) The mind's I: Fantasies and reflections on self and soul. Penguin Books, London: 391–414. Originally published in 1974. ▸ **http://cepa.info/2399**

Riegler A. (2001) Towards a radical constructivist understanding of science. Foundations of Science 6(1–3): 1–30. ▸ **http://cepa.info/1860**

Sven Delarivière studied philosophy at the Vrije Universiteit Brussel (VUB) and is currently a member of the Centre for Logic and Philosophy of Science at the VUB, working on a PhD exploring the notion of mathematical understanding and its possible automation.

Received: 18 May 2017
Accepted: 15 June 2017