SCIENCE AT THE EDGE

Conversations with the Leading Scientific Thinkers of Today

edited by

JOHN BROCKMAN

UNION SQUARE PRESS
An imprint of Sterling Publishing Co., Inc.

New York / London
www.sterlingpublishing.com

THE COMPUTATIONAL PERSPECTIVE

DANIEL C. DENNETT

When I go to a workshop or conference and give a talk, I'm actually doing research, because the howls and screeches and frowns that I get from people, the way in which they react to what I suggest, is often diagnostic of how they are picturing the problems in their own minds. And in fact people have very different covert images about what the mind is and how the mind works. The trick is to expose these images, to bring them up into public view and then correct them. That's what I specialize in.

DANIEL C. DENNETT is University Professor, professor of philosophy, and director of the Center for Cognitive Studies at Tufts University. A philosopher by training, he is known as the leading proponent of the computational model of the mind. He is the author of *Content and Consciousness; Brainstorms; Elbow Room; The Intentional Stance; Consciousness Explained; Darwin's Dangerous Idea; Kinds of Minds; Brainchildren; Freedom Evolves;* and *Breaking the Spell.* With Douglas Hofstadter, he coedited *The Mind's I*, and he is the author of over 200 scholarly articles on various aspects of the mind.

If you go back 20 years, or if you go back 200 years, 300 years, you see that there was one family of phenomena that people just had no clue about, and those were mental phenomena—that is, the very idea of thinking, perception, dreaming, sensing. We didn't have any model for how that was done physically at all. Descartes and Leibniz, great scientists in their own right, simply drew a blank when it came to trying to figure these things out. And it's only really with the ideas of computation that we now have some clear and manageable ideas about what could possibly be going on. We don't have the right story yet, but we've got some good ideas. At least you can now see how the job can be done.

Coming to understand our own understanding, and seeing what kinds of parts it can be made of, is one of the great breakthroughs in the history of human understanding. If you compare it with, say, our understanding of life itself or reproduction and growth, those were deep and mysterious processes 100 years ago and forever before that. Now we have a pretty clear idea of how things reproduce, how they grow, repair themselves, fuel themselves. All of those formerly mysterious phenomena are falling into place.

And when you look at such phenomena, you see that at a very fundamental level they're computational—that is, there are algorithms for growth, development, and reproduction. The central binding idea is that you can put together not billions but trillions of moving parts and get entirely novel, emergent, higher-level effects, and the best explanation for what governs those effects is at the level of software, the level of algorithms. If you want to understand how orderly development, growth, and cognition take place, you need to have a high-level understanding of how those billions or trillions of pieces interact with each other.

We never had the tools before to understand what happens when you put a trillion cells together and let them interact. Now we're getting those tools; even the lowly laptop gives us hints, because we see phenomena happening on our desks that would astound Newton or Descartes—or Darwin, for that matter. Phenomena that look like sheer magic. We know they aren't magic. There's not a thing that's magical about a computer. One of the most brilliant things about a computer is that there's nothing up its sleeve. We know to a moral certainty that there are no morphic resonances, psyonic waves, spooky interactions; it's good old push/pull, traditional, material causation. And when you put it together by the trillions with software, you get all this magic that's not really magic.

The idea of computation is a murky one; it's a mistake to think we have a clear, unified, unproblematic concept of what counts as computation. It's less clearly defined than the idea of matter or the ideas of energy or time in physics, for instance. Even computer scientists have only a fuzzy grip on what they actually mean by computation. The question is where to draw the line between what is computation and what isn't. That's not so clear. But that doesn't mean we can't have good theories of computation. Almost any process can be interpreted through the lens of computational ideas, and usually it's a fruitful exercise of reinterpretation. We can see features of the phenomenon through that lens that are essentially invisible through any other.

Human culture is the environment we live in. There's the brute physical environment—the streets, the air we breathe, the water we drink, the cars we travel in—and then there's all the communication going on around us in many different media: everyday conversation, newspapers, books, radio, television, the Internet. Pigeons live in our world, too, but they're oblivious to most of it; they don't care what's written in the newspaper they find their crumbs on. It's immaterial to them what the content, what the information, is. For us it's different; the information is really important.

If we think about the informational world our species lives in, we see that in fact it's got a lot of structure. It's not amorphous. Everything is not connected to everything else. There are lots of barriers. There's an architecture to this world of communication, and that architecture is changing rapidly, in ways we don't understand yet.

Let me give you a simple example of this. You could tune in the Super Bowl a couple of years ago and see these dot-com companies pouring an embarrassingly large amount of their initial capitalization into one ad for the Super Bowl; they were trying to get jump-started with this ad. And this was curious. If this was an Internet company, why didn't it use the Internet? Why do this retrograde thing of advertising on regular broadcast television? And the answer, of course, is that there is a fundamental difference in the conceptual architecture of those two media. When you watch the Super Bowl, you're part of a large simultaneous community—and you know it. You know that you are one of millions, hundreds of millions of people. You're all having the same experience at once, and you know that you are. And it's that second fact—it's that reflexive fact—that's so important. You go to a Web site and there might be 100 million people looking at that Web site but you don't know that. You may have read that somewhere . . . but you're not sure, you don't know. The sense you have when you're communing on the Web is a much more private sense than when you're watching something on network television. And this has huge ramifications regarding credibility. An ad that works well on television falls flat on the Web, because the people who see it, read it, listen to it, don't know what audience they're a part of. They don't know how big a room they're in. Is this a private communication or a public communication? We don't know yet what kind of fragmentation of the world's audiences is going to be occasioned by the Internet. The Internet brings people together, but it also isolates them in a way we haven't begun to assess. That sense of being utterly lost that neophytes have when they first get on the Web—choosing search engines, knowing what to trust, where home is, whom to believe, what sites to go to—arises because everybody is thirsting for reliable informants, signposts.

This geography of available information was established over centuries in the traditional media. You went to the *Times* and you read something there, and it had a certain authority for you. Or you went to the public library and read something in the *Encyclopedia Britannica.* These institutions had their own character, their own reputations, and their reputations were shared communally. It was important that your friends also knew that the *Times* or the *Encyclopedia Britannica* was an important place to look. Suppose somebody writes and publishes a volume called "Sammy's Encyclopedia of the World's Information"? It might well be the best encyclopedia in the world, but if people in general don't realize it, nobody's going to trust what's in there. It's this credibility issue that, as far as I can see, has not yet even begun to crystallize on the Web. We're entering uncharted waters there, and the outcome is hard to predict.

Human experience changed tremendously in the last century—and especially over the last decade. For instance, I'd guess that the average western-world teenager has heard more professionally played music than Mozart ever heard in his whole life (not counting his own playing and composing and rehearsal time). It used to be that hearing professional musicians in performance was a very special thing. Now *not* hearing professional musicians is a very special thing—there's a soundtrack almost everywhere you go. This is a huge change in the auditory structure of the world we live in. The other arts are similarly positioned. There was a time when just seeing written words was a rarity. Now everything has words on it. People can stand in the shower and read the back of the shampoo bottle. We are completely surrounded by the technology of communication—and that's new. Our species has no adaptations for it, so we're winging it.

There are lots of patterns in the world. Some are governed by the law of gravity, some by other physical principles. And some of them are governed by software. That is to say, the robustness of the pattern—the fact that it's salient, that you can identify it, that it keeps reproducing itself, that it can be found here, there, and elsewhere, and that you can predict it—is not because there's a fundamental law like the law of gravity that governs it but because these are patterns that occur wherever you have organisms that process information. They preserve, restore, and repair the patterns and keep them going. And that's a fundamental, new feature of the universe. If you went to a lifeless planet and surveyed all the patterns on it, these patterns wouldn't be there. They're the patterns you can find in DNA—those are the ur-patterns, the ones that make all the other patterns possible. They're also the patterns you find in texts. They have to have some physical embodiment in nucleotides or ink marks or particles and charges. But what explains their very existence in the universe is computation, the algorithmic quality of all things that reproduce and have meaning, and that make meaning.

These patterns are not, in one sense, reducible to the laws of physics, although they are based in physical reality. The explanation of why the patterns form as they do has to go on at a higher level. Douglas Hofstadter once offered a very elegant simple example: We come across a computer and it's chugging along, chugging along, chugging along. Why doesn't it stop? What fact explains the fact that this particular computer doesn't stop? And in Hofstadter's example, the reason it doesn't stop is that pi is an irrational number. What? Well, pi is an irrational number, which means that it's a never-ending decimal, and this particular computer program is generating the decimal expansion of pi, a process that will never stop. Of course, the computer may break. Somebody may come along with an ax and cut the power cord—but as long as it keeps powered, it's going to go on generating these digits forever. That's a simple concrete fact that can be detected in the world, the explanation of which cites an abstract mathematical fact.

Now, there are many other patterns in the world that aren't as arcane and have to do with the meaning we attach to things. Why is he blushing? There's a perfectly good explanation of what the *process* of blushing is: Blushing is the suffusion of blood through the skin of the face. But *why* is he blushing? He's blushing because he thinks she knows some fact about him that he wishes she didn't know. That's a complex, higher-order, intentional state, one that's visible only when you go to the higher, intentional level. You can't see that by looking at the individual states of the neurons in his brain. You have to go to the level at which you're talking about what this man knows and believes and wants.

The intentional level is what I call the "intentional stance." It's a strategy you can try whenever you're confronted with something complex in nature. It doesn't always work. The idea is to interpret that complexity as one or more intelligent, rational agents that have agendas, beliefs, and desires and are interacting. When you go up to the intentional level, you discover patterns that are highly predictive, robust, and not reducible in any meaningful sense to the lower-level patterns at the physical level. In between the intentional stance and the "physical stance" is what I call the "design stance." That's the level of software.

The idea of abstraction has been around for a long time, and 200 years ago you could enliven a philosophical imagination by asking what Mozart's Haffner Symphony is made of. It's ink on pieces of paper. It's a sequence of sounds as played by people with various stringed instruments and other instruments. It's an abstract thing. It's a symphony. Stradivarius made violins; Mozart made symphonies, which depend on a physical realization but don't depend on any particular one. They have an independent existence, which can shift from one medium to another and back.

We've had that idea for a long time, but we've recently become much more comfortable with it, living as we do in a world of abstract artifacts that jump promiscuously from medium to medium. It's no longer a big deal to go from the score, to the music you hear live, to the recorded version of the music. You can jump back and forth between media very rapidly now. It's become a fact of life. It used to be hard work to get things from one form to another; that's not hard work anymore; it's automatic. You eliminate the middleman. You no longer have to have the musician to read the score, to produce the music. This removal of all the hard work in translating from one medium to another makes it all the more natural to populate your world with abstractions, because you find it's hard to keep track of what medium they're in. And that doesn't matter much anymore—you're interested in the abstraction, not the medium. Where did you get that software? Did you go to a store and buy an actual CD and put it in your computer, or did you just download it off the Web? It's the same software, one way or another. It doesn't really matter. This idea of medium neutrality is one of the essential ideas of software, or of algorithms in general. And it's one that we're becoming familiar with, but it's amazing to me how much resistance there still is to this idea.

An algorithm is an abstract process that can be defined over a finite set of fundamental procedures—an instruction set. It is a structured array of such procedures. That's a very generous notion of an algorithm—more generous than many mathematicians would like, because I would include by that definition algorithms that may be in some ways defective. Consider your laptop. There's an instruction set for that laptop consisting of all the basic things its CPU can do; each basic operation has a digital name or code, and every time that bit sequence occurs, the CPU tries to execute that operation.

You can take any bit sequence at all and feed it to your laptop as if it were a program. Almost certainly, any sequence that isn't *designed* to be a program to run on that laptop won't do anything at all—it'll just crash. Still, there's utility in thinking that *any* sequence of instructions, however buggy, however stupid, however pointless, should be considered an algorithm, because one person's buggy, dumb sequence is another person's useful device for some weird purpose, and we don't want to prejudge that question. (Maybe that "nonsense" was included *in order* to get the laptop to crash at just the point it crashed!) One can define a more proper algorithm as one that runs without crashing. The only trouble is that if you define algorithms that way, then probably you don't have any on your laptop, because there's almost certainly a way to make almost every program on your laptop crash. You just haven't found it yet. Bug-free software is an ideal that's almost never achieved.

Looking at the world as if everything were a computational process is becoming fashionable. Here one encounters not an issue of fact but an issue of strategy. The question isn't "What's the truth?" The question is "What's the most fruitful strategy?" You don't want to abandon standards and count everything as computational, because then the idea loses its sense; it doesn't have any grip anymore. How do you deal with that? One way is to try to define, in a rigid centralist way, some threshold that has to be passed, and refuse to call a process computational unless it has properties A, B, C, D, and E. You can do that any number of ways and it will save you the embarrassment of having to say that everything is computational. The trouble is that anything you choose as a set of defining conditions will be too rigid. There will be processes that meet those conditions but aren't interestingly computational by anybody's standards, and there are processes that don't meet the standards but nevertheless are significantly like the things you want to consider computational. So how do you deal with the issue of definition? By ignoring it—that's my suggestion. Same as with life! You don't want to argue about whether viruses are alive or not; in some ways they're alive, in some ways they're not. Some processes are obviously computational. Others are obviously not computational. Where does the computational perspective illuminate? Well, that depends on who's looking at the illumination.

I describe three stances for looking at reality: the physical stance, the design stance, and the intentional stance. The physical stance is where the physicists are; it's matter and motion. The design stance is where you start looking at the software—at the patterns that are maintained—because these are designed things that are fending off their own dissolution. That is to say, they are bulwarks against the Second Law of Thermodynamics. This applies to all living things and also to all artifacts. Above that is the intentional stance, which is the way we treat that specific set of organisms and artifacts that are themselves rational information-processing agents. In some sense, from the intentional stance you can treat Mother Nature—that is, the whole process of evolution by natural selection—as an agent, but we understand that that's a *façon de parler*, a useful shortcut for getting at features of the design processes that are unfolding over eons of time. Once we get to the intentional stance, we have rational agents, we have minds, creators, authors, inventors, discoverers—and everyday folks—interacting on the basis of their take on the world.

Is there anything above that? Well, in one sense there is. People—or persons, as moral agents—are a specialized subset of the intentional systems. All animals are intentional systems. *Parts* of you are intentional systems. You're made up of lots of lesser intentional systems—homunculi of sorts—but unless you've got multiple personality disorder, there's only one person there. A person is a moral agent—not just a cognitive agent, not just a rational agent, but a moral agent. This is the highest level I can make sense of. Why it exists at all, how it exists, the conditions for its maintenance: these are very interesting problems. We can look at game theory as applied to the growth of trees—they compete for sunlight, it's a game in which there are winners and losers.

But when we look at game theory as applied not just to rational agents but to people with a moral outlook, we see some important differences. People have free will; trees don't. It's not an issue for trees in the way it is for people.

What I like about the idea that people are animals with free will is that it agrees with philosophical tradition (including Aristotle and Descartes, for instance) in maintaining that people *are* different—that people aren't *just* animals. Traditional theorists completely disagree, of course, on what that difference consists of. Although it's a naturalization of the idea of people, it does say they're different, and this, I discover, is the thing that most entices and upsets people about my view. There are those who want people to be more different than I'm allowing. They want people to have souls, to be Cartesian people. And there are those who are afraid that I'm trying to differentiate people too much from the other animals with my claim that human beings really are, because of culture, an importantly different sort of thing. Some scientists view this claim with skepticism, as if I'm trying to salvage for philosophy something that should fall to science. But in fact my view about what's different about people is a scientific theory; it stands or falls as an implication of a scientific theory, in any case.

Regarding my own role in cognitive science—as to whether I consider myself a philosopher or a scientist—I think I'm good at discovering the blockades of imagination, the bad habits of thought, that infect how theorists think about the problem of consciousness. When I go to a workshop or conference and give a talk, I'm actually doing research, because the howls and screeches and frowns that I get from people, the way in which they react to what I suggest is often diagnostic of how they are picturing the problems in their own minds. And in fact people have very different covert images about what the mind is and how the mind works. The trick is to expose these images, to bring them up into public view and then correct them. That's what I specialize in.

My demolition of the Cartesian theater, of Cartesian materialism, is just one of these campaigns of exposure. People often pay lip service to the idea that there isn't any privileged medium in the brain playing the role that Descartes assigned to the nonphysical mind as the theater of consciousness. Nevertheless, if you look closely at what they are thinking and saying, their views make sense only if you interpret them as covertly presupposing a Cartesian theater somewhere in their model. Teasing this out, bringing it to the surface and then showing what you might replace it with is, to me, very interesting work. Happily, some people have come to appreciate this as a valuable service that somebody like me, a philosopher, can perform: getting them to confront the hidden assumptions in their own thinking and showing how those hidden assumptions blind them to opportunities for explaining what they want to explain.