

## CONTENTS

<i>Preface</i>	ix
<i>Acknowledgments</i>	xiii
<b>1</b> Player One	1
<b>2</b> Life Goes On	24
<b>3</b> Action!	44
<b>4</b> Life Gets Complicated	68
<b>5</b> The Perceiving Self	95
<b>6</b> Choosing	121
<b>7</b> The Future Is Not Written	145
<b>8</b> Harnessing Indeterminacy	170
<b>9</b> Meaning	194
<b>10</b> Becoming Ourselves	220
<b>11</b> Thinking about Thinking	248
<b>12</b> Free Will	273
Epilogue: Artificial Agents	295
<i>Figure Credits</i>	301
<i>Bibliography</i>	303
<i>Index</i>	319

# 1

## Player One

Are we the authors of our own stories? Or is our apparent freedom of choice really an illusion? These questions were brought home to me recently as I was watching my son play a video game—one where you wander around an open world, meeting interesting denizens of one type or another (and killing quite a few of them). As I watched, his character entered a tavern and approached the bartender, who offered a generic greeting. The game then threw up some options for things you could say in reply to get information about the prospects for fortune and glory in the surrounding territory.

In this exchange, my son's possibilities for action were limited by the game, but he *was* really making choices among them, and these choices then affected how the conversation went and what would subsequently unfold. His decisions were based on his overall goal in the game, the tension between his goals of taking some immediate action or to keep exploring, his need to have enough information to make a decision with confidence, the risk of biting off more than he could chew and losing his hard-won stuff: all these considerations fed into the decisions he made. He had his reasons and he acted on them, just like you or I do every day, all day long.

The bartender, in contrast, was not making choices. He was a classic “non-player character,” an NPC. His responses were completely determined by his programming: he had no degrees of freedom. His actions were merely the inevitable outcome of a flow of electrons through the circuits of the game console, constrained by the rules encoded in the

software. Even the more sophisticated NPCs in the game, including the monster that eventually caramelized my son's avatar, were similarly constrained. The monster's actions—even in the fast-moving melee—were determined by the software programming and mediated by the electronic components in the console.

Thus the NPCs only *appear* to be making choices. They're not autonomous entities like us: they're just a manifestation of lots of lines of code, implemented in the physical structure of the computer chips. Their behavior is entirely determined by the inputs they get and their preprogrammed responses. We, in contrast, are causes of things in our own right. We have *agency*: we make our own choices and are in charge of our own actions.

At least it seems that way. It certainly feels like we have “free will,” like we make choices, like we are in control of our actions. That's pretty much what we do all day—go around making decisions about what to do. Some are trivial, like what to have for breakfast; some are more meaningful, like what to say or do in social or professional situations; and some are momentous, like whether to accept a job offer or a marriage proposal. Some we deliberate on consciously, and others we perform on autopilot—but *we* still perform them. Of course, our options may be more or less constrained (or informed) by all kinds of factors at any given moment, but generally we feel like the authors of our own actions.

And we interpret other people's behavior in terms of their reasons for selecting different actions—their intentions, beliefs, and desires that make up the content of their mental states. We constantly analyze each other's motives and habits and character, looking for explanations and predictors of their behavior and the decisions they make. Why people act the way they do is ultimately the theme of most entertainment, from Dostoyevsky to *Big Brother*. All this rests on the view that we are not just acted on—we are actors. Things don't just happen to us, in the way they happen to rocks or spoons or electrons: *we do things*.

The problem is that, if you think about this view for too long, it becomes difficult to escape a discomfiting thought. After all, like the NPCs, our decisions, however complex they may be, are mediated by the flow of electrical ions through the circuits of our brains and thus are constrained by our own “programming,” by how our circuits are configured.

Unless you invoke an immaterial soul or some other ethereal substance or force that is really in charge—call it spirit or simply mind, if you prefer—you cannot escape the fact that our consciousness and our behavior emerge from the purely physical workings of the brain.

There is no shortage of evidence for this from our own experience. If you've ever been drunk, for example, or even just a little tipsy, you've experienced how altering the physical workings of your brain alters your choices and the way you behave. There is a whole industry of recreational drugs—from caffeine to crystal meth—that people take because of the way that physically tweaking the brain's machinery in various ways makes them feel and act. The ultimate consequence in some cases is addiction—perhaps the starkest example of how our actions can sometimes be out of our control.

And, of course, if the machinery of your brain gets physically damaged—as occurs with head injuries, strokes, brain tumors, neurodegenerative disorders, or a host of other kinds of insults—or its function is impaired in other ways, as in conditions such as schizophrenia, depression, or mania, then your ability to choose your actions may also be impaired. In some situations the integrity of your very *self* may be compromised.

We all like to think that we are Player One in this game of life, but perhaps we are just incredibly sophisticated NPCs. Our programming may be complex and subtle enough to make it *seem* as if we are really making decisions and choosing our own actions, but maybe we're just fooling ourselves. Perhaps “we” are just the manifestations of genetic and neural codes, implemented in biological rather than computer hardware. Perhaps we are the victims of a cruel joke, tragic figures in the grip of the Fates. As Gnarl Barkley sang, “Who do you, who do you, who do you think you are? Ha ha ha, bless your soul, you really think you're in control.”

## Robots with Personality

In my 2018 book *Innate* I described how we all come pre-wired with a set of innate psychological predispositions. At the most basic level, we all share the profile of human nature. Evolution has shaped the behavior

of our species just as much as that of any other. Human nature is encoded in our DNA in a genetic program that specifies the building and wiring of our human brains.

However, the details of that genetic program inevitably vary among individuals. I use the word “inevitably” because there is no way that this variation could not exist. Every time DNA is copied in a cell, including when sperm or egg cells are made, some small number of copying errors or mutations arise. New variations in the DNA sequence thus enter the gene pool in every generation, and—if their effects are tolerated—they can spread through the population over time, leading to the accumulation of genetic variation that we observe.

This leads to the differences that we observe in people’s physical traits, such as how tall they are or the shape of their faces or various aspects of their physiology. This variation occurs just the same in the physical structure of their brains and the way they function. The fact that all these traits are affected by genetic variation explains why people who are related to each other resemble each other more than do unrelated people, both physically and psychologically. So, even though the “canonical” human genome (which doesn’t really exist anywhere) encodes a program to build a canonical human brain, your particular genome encodes a program to build *a brain like yours*.

But not exactly like yours. The program in your genome does not encode one particular outcome, specified down to the level of individual nerve cells or synaptic connections between them. It does not encode the outcome at all, in fact: it just encodes a set of biochemical rules that, when played out over the complicated processes of development, will tend to result in an outcome within a certain range. Exactly how these processes played out in your specific case was also affected by all kinds of random events during development that added considerable variation. If you ran the program again, you would not get exactly the same outcome. Even the brains of identical twins who share the same genetic program are quite distinct from each other already at birth.

All this means that the way your brain is wired is affected by millions of years of common evolution, by the specific genetic variations that you carry, and by the unique trajectory of development that occurred

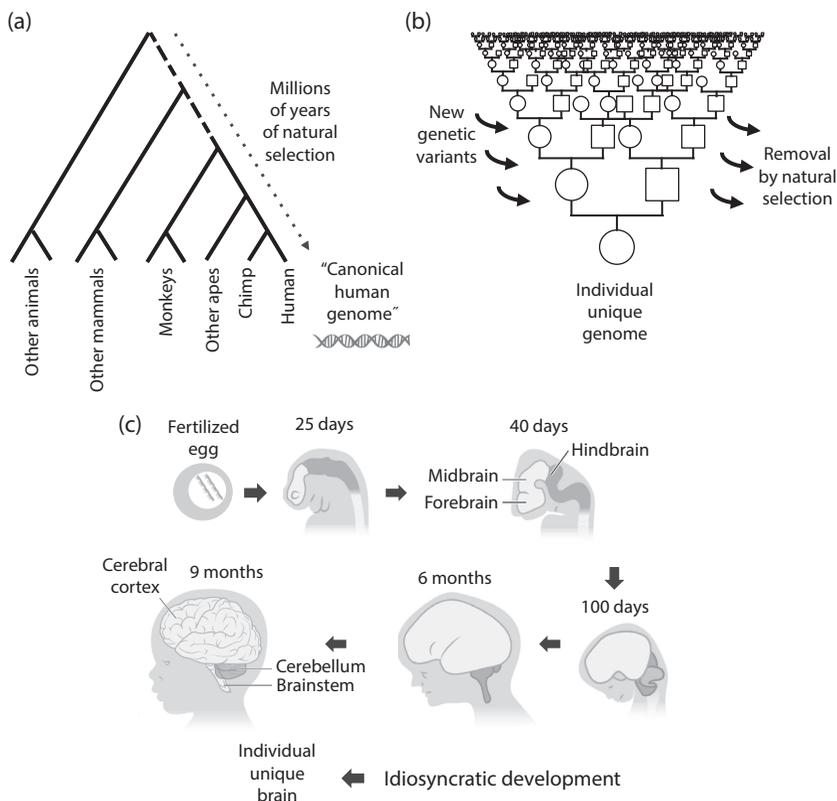


FIGURE 1.1. The making of you. (a) Like all species, the genome of *Homo sapiens* has been shaped by millions of years of evolution, selecting for all the traits that comprise “human nature” generally. (b) Your individual genome is a unique version of the canonical human genome, reflecting the processes of mutation and selection in your specific ancestors. (c) The outcome of brain development in any individual is idiosyncratic, shaped by genetic variation and the unique trajectories of development itself. Our individual natures (or innate predispositions) are thus variations on the theme of human nature generally.

while your brain was developing (see Figure 1.1). And the way your brain is wired affects how it works and how you will tend to behave.

We can think of this variation like that in the internal tuning of the behavioral controls of a robot. Imagine you and I were asked to build an autonomous robot that has to make its way in the world—finding

fuel, avoiding threats, interpreting sensory information, assessing situations, and deciding among possible actions.

There are all kinds of things we would have to build in our robot for it to accomplish its tasks. It would need some sensors, of course, to detect things in the environment, and it would require motors so it could move around and perform various actions. It would need to be programmed to move toward fuel and away from threats, but it might also require some fancy circuitry for it to recognize which is which. And what would happen if fuel supplies and something threatening happen to be in the same place? It would have to weigh the opportunity versus the risk and make a decision accordingly about where to move. And it would be good if that decision were informed by how much fuel it had left at the time. So some way to monitor its internal states and use them to inform decisions would certainly be beneficial.

A very fancy robot might also be able to learn from experience; for example, it might learn that there tends to be fuel in some particular spot or that some kind of otherwise innocuous stimulus (a rustle in the robot grass perhaps) signals a hidden threat. Now imagine we give our robot another goal: not just to survive but also to find robot love and reproduce. Then it would have to balance the short-term goal of ensuring it has enough fuel with the longer-term goal of finding a mate, all while not getting destroyed by a bigger robot.

All those functions—some means for inferring what is out in the world from the data gathered by its sensors; integrating both external and internal multiple signals to derive a picture of the whole situation; comparing that with the data in its memory bank to help inform its next action; weighing threats versus opportunities, short-term versus long-term goals, and good versus bad outcomes; and eventually picking one action to perform while inhibiting all other possibilities—would have to be configured into its circuitry.

With so many circuits and parameters that could vary, it is inevitable that the way you would tune your robot would differ from how I would tune mine. You might set the threat sensitivity a little higher and the reward sensitivity a little lower. I might tune the circuits in my robot with a different balance between short- and long-term goals. All these

settings would manifest as different *patterns of behavior* over time and across contexts. Your robot might appear more cautious than mine. Mine might show more perseverance: it might be willing to work longer for a delayed reward. The robots could differ in how much evidence they need to make a decision (impulsivity), how much they value mating opportunities (sex drive), and even how salient they find novel objects or situations (curiosity). In short, our robots would have personalities, just like you and I do.

And just like you or me, they would not have had any hand in choosing those traits. Even if the robots learn over their lifetime and adapt to the various scenarios they encounter, all this learning is also physically embodied in the configuration of their circuits at the moment they are faced with a decision. The sense of fatalism that this realization engenders is aptly summarized by prominent free will skeptic Sam Harris:

Take a moment to think about the context in which your next decision will occur: You did not pick your parents or the time and place of your birth. You didn't choose your gender or most of your life experiences. You had no control whatsoever over your genome or the development of your brain. And now your brain is making choices on the basis of preferences and beliefs that have been hammered into it over a lifetime—by your genes, your physical development since the moment you were conceived, and the interactions you have had with other people, events, and ideas. Where is the freedom in this? Yes, you are free to do what you want even now. But where did your desires come from?<sup>1</sup>

The essence of the problem was captured by the famously pessimistic (or some might say realistic) philosopher Arthur Schopenhauer, who said, "A man can do what he will, but not will as he will."<sup>2</sup> Even if we are making choices right now, those choices are not free from all kinds of prior causes or influences, over which we had no control.

1. Sam Harris, *Free will* (New York: Free Press, 2012), 44.

2. Arthur Schopenhauer, *Essay on the freedom of the will* (New York: Dover, 1960), 6.

## The Machine

As a neuroscientist, this kind of existential worry is an occupational hazard. But it gets worse. The more we learn about the mechanisms of perception and cognition and, in particular, of decision making and action selection, the more *mechanistic* it all seems and the less there seems to be for *the mind* to do. How can we even think that *we* are making choices at all, when we can see that the process is the result of just a bunch of gears turning in the machine? What reason is there to think that an entity is in charge?

And, thanks to modern technology, we can actually see the figurative gears turning. Using a variety of neuroimaging tools in humans and animals to track the activity of different neural circuits or brain areas, it is possible to tease out the types of information they carry and the cognitive operations they perform as the organisms or individuals make decisions or select actions. We can, for example, distinguish patterns of neural activity that correlate with (and seem to internally “represent”) the accumulation of evidence about something in the world, the degree of certainty attached to some signal, the confidence level in a belief, the adoption of a new goal, the rewards associated with a positive outcome, the learning that happens in response to such rewards, the emotional signals that accompany decision making, the gradual formation of habits, the real-time switch from habitual to goal-directed or exploratory behavior as circumstances change, and on and on. We can see the thinking happening.

We can even, in some circumstances, predict an incipient action before the individual performs it. There are many experimental setups using rodents or monkeys where researchers can track patterns of brain activity, observe a threshold being approached that will result in an action, and even predict (not with complete accuracy but significantly better than chance) what action it will be—whether a rat will turn left or right in a maze, for example.

In humans there is a famous example where an action was not only predicted ahead of time but also before the subject even became consciously aware of having chosen to do it. In these experiments, performed

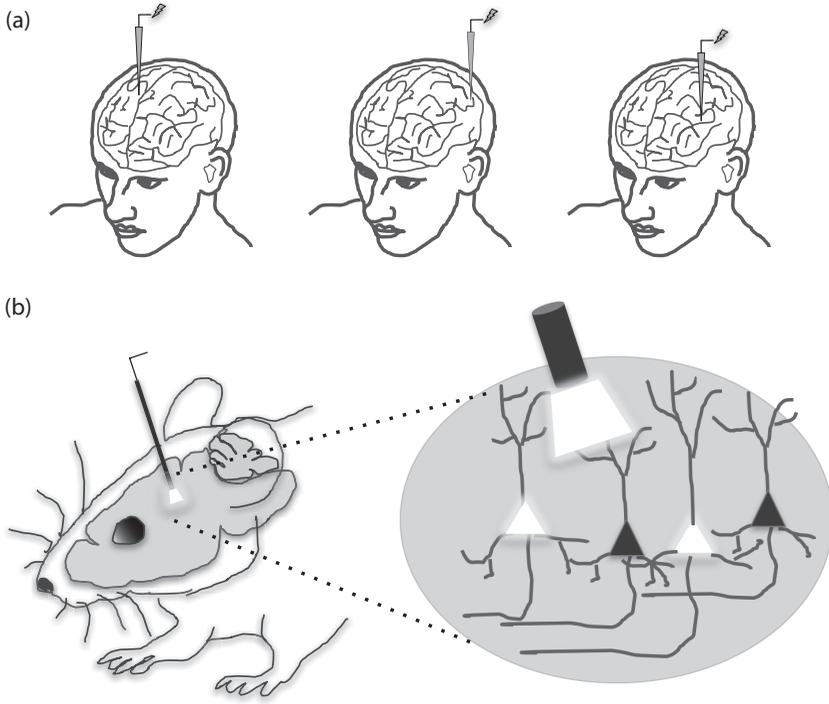
by Benjamin Libet and colleagues in the 1980s, subjects had to randomly decide to move their fingers while watching a clock and while their brainwaves were being recorded by an electroencephalograph. The striking result: the onset of brain activity leading to a movement preceded the reported timing of the conscious awareness of the intention to act by several hundred milliseconds.

Although not relevant to truly deliberative decisions, these findings can still shake your faith in your conscious mind really being in control of your actions. Is the rest of the brain just flattering us, making us feel that we're in charge, like a wily civil servant expertly managing his elected boss?

If pulling back the curtain to expose the neural machinery of decision making at work were not enough of a threat to our egos (in both senses of that word), it is also possible to intervene in the machine—to drive patterns of neural activity from the outside—and *cause* the individual to behave in certain ways.

Famous experiments carried out in the 1940s by neurologist Wilder Penfield and his colleagues in human subjects undergoing brain surgery (who were awake and aware throughout the procedure) showed that stimulating different parts of the cerebral cortex with electrodes could produce all kinds of sensations, emotions, urges, memories, or movements of various parts of the body (see Figure 1.2). This work contributed greatly to the mapping of functions across the brain and reinforced the view of a complex electrical machine *producing* the contents of the mind, rather than being controlled by that mental content.

Similar experiments are possible in animals, but, as in humans, they're a bit crude. Just poking an electrode into a part of the brain and zapping it activates all the neurons in that area in a nonspecific fashion. The brain then attempts to make sense of that mini-explosion of activation, but this process is very different from how neural signaling normally happens. Indeed, within any little chunk of brain, there are hundreds of different types of nerve cells connected in intricate microcircuits designed to carry out diverse sorts of computations. Just blasting them all at once is thus not hugely informative about how these computations mediate cognitive operations.



**FIGURE 1.2.** Brain stimulation. **(a)** Direct stimulation of brain areas in awake subjects can lead to involuntary movements (motor cortex, left), sensory percepts (visual cortex, middle), or even activation of memories (temporal lobe, right), depending on the area stimulated. **(b)** Optogenetic techniques in rodents allow much more specific activation of genetically targeted subsets of neurons, providing a powerful platform to dissect the functions of neural circuits in awake, behaving animals.

Seventy years after Penfield's experiments, the study of the neural systems that control behavior in animals was revolutionized by the invention of molecular tools that allow researchers to drive the activity of very specific subsets of neurons in an animal's brain while observing its behavior in real time. Like most techniques in molecular biology, this one—called optogenetics—borrows from nature. It uses a protein made by blue-green algae that sits in the membrane of the cell and responds to light by opening up a channel through which electrically charged atoms (or ions) can pass. That protein is related to ones that we

use in our eyes to detect light, and it is exactly the opening of ion channels in the membrane that drives nerve cells to “fire” or send a sharp electrical signal.

Researchers including Karl Deisseroth, Edward Boyden, and others realized that if they cloned the algal gene that encodes this protein and transferred it to mammalian neurons, they could effectively turn “on” the neurons with exquisite temporal precision by shining a blue light on them. Hooking the piece of DNA that codes for this light-responsive channel protein (called channel-Rhodopsin) to the DNA codes that regulate the expression of all kinds of different genes in the mouse brain enabled them to generate lines of transgenic mice expressing channel-Rhodopsin in extremely specific subsets of neurons in different brain regions.

Shining a light on the relevant bit of the brain—accomplished by threading a minute fiber optic cable through the skull—allows researchers to activate just that specific subset of neurons within the circuit and study the effects on behavior of the animal. Using this technique, specific sets of neurons were identified that, when activated, drive all kinds of behaviors—from general locomotion to more subtle motor actions like reaching or grasping, from aggression to mating, from freezing in fright to lunging attacks on prey that are not present, from eating to sleeping to looking after pups, and on and on.

But this research reaches far beyond directly activating particular actions from the animal’s repertoire of behavior. It has made it possible to dissect the cognitive machinery involved in choosing among actions, weighing options, signaling rewards and punishments, judging the reliability of sensory information, assigning a level of certainty or confidence to a decision, using past memories to guide current actions, and selecting one option while inhibiting every other possibility. It is even possible, as my colleague Tomás Ryan and others have done, to implant false memories in an animal’s brain that will influence its future behavior. This is not just remote control of what the animal is doing: it is control of what the animal *is thinking*.

It’s hard not to look at this growing body of work and see only the machine at work. Driving this circuit or that one either directly causes an action or influences the cognitive operations that the animal—mouse

or human or anything else—uses to decide between actions. If we were dissecting a robot in this way we would apply engineering approaches to understand the kinds of information being processed, the control mechanisms configured into the different circuits, and the computations that lead to one output or another. There does not seem to be any need for something like *a mind* in that discussion. There is no real need for *life*, for that matter.

If the circuits just work on physical principles, then who cares what the patterns of activity *mean*? Why does it matter what the mental content associated with a particular pattern of neural activity is, if it is solely the physical configuration of the circuitry that is going to determine what happens next? We may have set out, as neuroscientists, to explain how the workings of the brain generate or realize psychological phenomena, but we are in danger of explaining those phenomena away.

### It's All Just Physics at the End of the Day

If the neuroscientists have it bad, pity the poor physicists, whose existential angst must run much deeper. Where neuroscientists can at least hold onto the view that the circuits in the brain are doing things (whether “*you*” are or not), some physicists claim that even that functionality is an illusion. After all, the brain is made of molecules and atoms that must obey the laws of physics, just like the molecules and atoms in any other bit of matter.

These small bits of matter are pushed and pulled by all the forces acting on them—gravity, electromagnetism, the so-called strong and weak nuclear forces that hold atoms together—and where each atom goes is fully determined by the way those interactions play out. These processes are no doubt complicated, as they would be in any system with so many atoms simultaneously acting on each other, and in practice how the system will evolve is unpredictable—but it is still all driven by the physics. Even at the lower levels of subatomic particles, how the system evolves is captured by the equations of quantum mechanics in a way that many would argue theoretically leaves no room for any other causes to be at play.

So, then, what does it matter what you are thinking? You cannot push the atoms in your brain around with a thought. You cannot override the fundamental laws of physics or exert some ghostly control over the basic constituents of matter. According to this view, the very idea of mental causation—of the content of your thoughts and beliefs and desires mattering in some way—is a naive superstition, a conceptual hangover inherited from philosophers like the famous dualist Rene Descartes.

Here is the late Stephen Hawking on the subject: “Biological processes are governed by the laws of physics and chemistry and therefore are as determined as the orbits of the planets. Recent experiments in neuroscience support the view that it is our physical brain, following the known laws of science, that determines our actions and not some agency that exists outside those laws . . . so it seems that we are no more than biological machines and that free will is just an illusion.”<sup>3</sup> Brian Greene, another well-known physicist and author, agrees: “Free will is the sensation of making a choice. The sensation is real, but the choice seems illusory. Laws of physics determine the future.”<sup>4</sup>

There are two main flavors of this kind of physical determinism. In the first, the low-level laws of physics rule completely: every aspect of the way the universe and everything in it evolves is fully determined by how these interactions play out. There is no room for any other force and, in particular, no role for any kind of randomness or indeterminacy. This model can be summed up as follows:

current state + laws of physics → next state

The consequences of this view are stark. If you keep on working through from one state to the next, you quickly realize that the current state predicts not just one step ahead but also two, or three, or actually an infinite number. And you can work backward just as easily as forward. If this is really the whole picture, then the entire history of the

3. Stephen Hawking and Leonard Mlodinow, *The grand design* (New York: Bantam Books, 2010), 32.

4. Brian Greene (@bgreene) on Twitter, June 5, 2013, <https://twitter.com/bgreene/status/342376183519916033?lang=en>.

universe up until now and for the rest of time was *predetermined* from shortly after the Big Bang. Indeed, our conception of time as having a direction goes out the window. The whole universe, over all time, is simply given, as a block: there is no real difference between the past and the future. There are no possibilities—only what has happened and what will happen. This view is known as *hard determinism*.

The implied softer version differs in allowing some randomness or indeterminacy to exist. It holds that the future is *not* fully predetermined by the current state (and certainly not by the initial state of the universe). Here, the past and the future are very different: the past is fixed while the future is a branching web of possibilities, only one line of which will be realized at any choice point.

However, even though the branch that is taken is not predetermined in this model, it is still decided by the low-level interactions of all the atoms and molecules. It is just that some of those interactions are a bit random. You might sum up this view like this:

current state + laws of physics + randomness → next state

The debate over whether there really is any true randomness in physical events has been raging since the days of Einstein and Bohr. When you get down to the quantum level of subatomic particles, weird things happen, and even though the weirdness can be fully accounted for in the equations that physicists use, allowing them to make exquisitely precise predictions, there is no consensus at all about what these equations imply about the fundamental nature of reality.

We'll return to this topic of randomness later. For now, what are the implications of this softer version of determinism? It is often summed up with the pithy line: "every event has a cause." This doesn't seem to align with the idea of random events happening, which would seem not to have a cause, by definition. What this statement really seems to imply is that everything that happens—at a system level—is caused by the interactions of particles at the lowest level, even if some randomness is at play there.

Yet, that view seems to be just as problematic for the idea of organisms like us being in charge of anything that happens. The future may

not be written, but if what happens is still decided by how the physical forces play out at the minutest scale of matter, there doesn't seem to be much scope for us to be in control. Even neuroscientist Patrick Haggard, a leader in the study of volition, agrees: "As a neuroscientist, you've got to be a determinist. There are physical laws, which the electrical and chemical events in the brain obey. Under identical circumstances, you couldn't have done otherwise; there's no 'I' which can say 'I want to do otherwise.'"<sup>5</sup>

In hard determinism, there are no causes. The universe just inexorably unfolds according to the laws of physics. If nothing could ever be or have been different, then you cannot point to one thing being a certain way and say it caused something else. The concept just doesn't apply. In *soft determinism*, there are causes—some things could be different, depending on how that little bit of randomness plays out—but all the causes are located at the lowest levels. That lowest level is deemed to be the bedrock of reality.

Some physicists, like Sean Carroll or Sabine Hossenfelder, may be magnanimous enough to allow that descriptions at higher levels of organization are "useful ways of talking about" complicated systems. We can productively do chemistry or biology or psychology with theories and methods that remain at those higher levels. But Carroll maintains that the real truth—the whole truth—resides at the lowest level, with the fundamental physical interactions of the smallest particles. If you had a complete accounting of what is going on down there, then you would not need any other information to fully predict what the system will do: everything happening at the higher levels simply derives or emerges from the low-level dynamics. Every other description is just a kind of coarse-grained picture, a *simplification* or statistical averaging that allows our puny minds to grasp how various systems—like cells or brains or minds—behave, despite all the underlying complexity.

5. Patrick Haggard, Neuroscience, free will and determinism: "I'm just a machine," interview by Tom Chivers, *The Telegraph*, October 12, 2010, <https://www.telegraph.co.uk/news/science/8058541/Neuroscience-free-will-and-determinism-Im-just-a-machine.html>.

Given the phenomenal successes of modern physics in confirming the predictions of quantum mechanics with eye-watering precision, it is not surprising that the focus has been on continuing to develop and test such theories while not worrying too much about what they mean for the nature of reality. The admonition to “Shut up and calculate!” by quantum physicist David Mermin is effectively the motto of the field. Let the philosophers worry about what it all means, especially for metaphysical concepts like free will.

## The Blame Game

Philosophers, for their part, have been debating the implications of deterministic theories of the physical universe for free will for thousands of years, at least as long ago as Democritus and Epicurus in ancient Greece. That these debates continue today with unabated fervor tells you that they have not yet resolved the issue.

In fairness, free will is a uniquely vexing problem. The phenomenon we are trying to explain—our own experience of having the power to make choices—seems inherently at odds with what we know about how everything else works in the universe. The scientific rejection of the idea of an immaterial soul or spirit that is somehow pulling the strings has left us scrambling to explain instead how the machine could pull its own strings. And the progress of physics into the wonderful weirdness of the quantum realm has only deepened the mysteries of what the machine and the world around it are made of in the first place.

But if philosophy can be excused for not having provided an answer, one might at least have hoped for some consensus on what is the right question. The popular framing, “Do we have free will?” is undermined in an obvious way by a lack of agreed-on definitions. If you define the capacity of free will as being able to make decisions in a way that is necessarily *free from every prior cause*, then you have set an unattainable standard, one that could only be met by supernatural means. Alternatively, if your criterion is merely that a person is doing things *based on causes internal to his or her physical self*, then you have not met the chal-

lence of physical determinism but merely sidestepped it with appeals to complexity and unpredictability.

Less obviously, the question “Do we have free will?” is more deeply undermined by a lack of clarity of the terms “we” and “have.” We cannot profitably approach the question of whether *you* have free will until we have answered the much more fundamental question, “What kind of thing are you?” The contrasting criteria cited earlier are founded on differing conceptions of the nature of the self, where the philosophical footing is equally treacherous. Without a shared understanding of what everyone is talking about, it’s not surprising that the debate seems to go round and round interminably.

Another barrier to a clear explication of the arguments around whether free will exists is that they are often approached from the direction of their *consequences* for our positions on moral responsibility. If people are not really in control of their actions—if we are nothing more than physical automata, mounting a wonderfully sophisticated but ultimately empty simulacrum of free will—then how can we be worthy of praise or blame? How can we defend judgment or punishment? The stakes here could not be higher. The idea of moral responsibility is the foundation not only of our legal systems but also of all our social interactions. We are constantly thinking about what we should or shouldn’t do in any given circumstance and probably spend even more time thinking about what other people should or shouldn’t do (or should or shouldn’t have done).

But tying the discussion of free will to the issue of moral responsibility muddies the waters. Questions of moral responsibility are crucially important, of course, but they are confounded by all kinds of additional issues: the nature and origins of our moral sensibilities, the evolution of moral norms, the legal philosophies underpinning our justice systems, and the complex and innumerable pragmatic decisions that societies and individuals have to make to keep our collective existence stable. Asking what kind of free will *we want* that will let us maintain our positions on moral responsibility can become almost a theological exercise in motivated reasoning. It means we are looking for a palatable answer

instead of trying to understand what really *is*. It is coming at the question from the wrong end, picking an answer we like and seeing what edifice of arguments we need to build to support it. Instead, I would like to know what kind of free will *we actually have*.

## Back to the Start

It's fashionable these days to claim that "free will is an illusion!": either it does not exist at all, or it is really not what we think it is. I am not willing to give up on it so easily. In this book I argue that we really *are* agents. We make decisions, we choose, we act—we are causal forces in the universe. These are the fundamental truths of our existence and absolutely the most basic phenomenology of our lives. If science seems to be suggesting otherwise, the correct response is not to throw our hands up and say, "Well, I guess everything we thought about our own existence is a laughable delusion." It is to accept instead that there is a deep mystery to be solved and to realize that we may need to question the philosophical bedrock of our scientific approach if we are to reconcile the clear existence of choice with the apparent determinism of the physical universe.

But if we want to solve this mystery, humans are the absolute worst place to start. It is a truism in biology to say that nothing makes sense except in the light of evolution—and this is surely true of agency. Instead of trying to understand it in its most complex form, I go back to its beginnings and ask how it emerged, what the earliest building blocks were, and what the basic concepts should be. How can we think about things like purpose and value and meaning without sinking into mysticism or vague metaphor? I argue that we can do so by locating these concepts in simpler creatures and then following how they were elaborated over the course of evolution, increasing in complexity and sophistication as certain branches of life developed ever-greater autonomy and self-directedness.

Indeed, before tackling the question of free will in humans, we have a much more fundamental problem to solve. How can any organism be said to *do anything*? Most things in the universe don't make choices.

Most things—like rocks or atoms or planets—don't do anything at all, in fact. Things happen to them, or near them, or in them, but they are not capable of action. But *you* are. You are the type of thing that can take action, that can make decisions, that can be a causal force in the world: you are an agent. And humans are not unique in this capacity. All living things have some degree of agency. That is their defining characteristic, what sets them apart from the mostly lifeless, passive universe. Living beings are autonomous entities, imbued with purpose and able to act on their own terms, not yoked to every cause in their environment but causes in their own right.

To understand how this could be, we have to go right back to the beginning, to the very origins of life itself (see Figure 1.3). This is the trajectory that I sketch out in this book.

From the chemistry of rocks and hydrothermal vents—the chemistry of the evolving planet itself—life emerged as systems of interacting molecules, interlocked in dynamic patterns that became self-sustaining. The ones that most robustly maintained their own dynamic organization persisted, replicated, evolved. They became enclosed in a membrane—a tiny subworld unto themselves—exchanging matter and energy with their environment while protecting an internal economy and reconfiguring their own metabolism to adapt to changing conditions. They became autonomous entities, causally sheltered from the thermodynamic storm outside and selected to persist.

A new trick was invented: action, the ability to move or affect things out in the environment. Information became a valuable commodity, and mechanisms evolved to gather it from the environment. With that came the crude beginnings of value and meaning. Movement toward or away from various things out in the world became good or bad for the persistence of the organism. These responses were selected for and became wired into the biochemical circuitry of simple creatures.

As multicellular creatures evolved, a class of cells—neurons—emerged that specialized in transmitting and processing information. Initial circuits acted as internal control systems, designed to coordinate the various muscles or other moving parts of the multicellular animal, defining a repertoire of useful actions. At the same time, neurons coupled

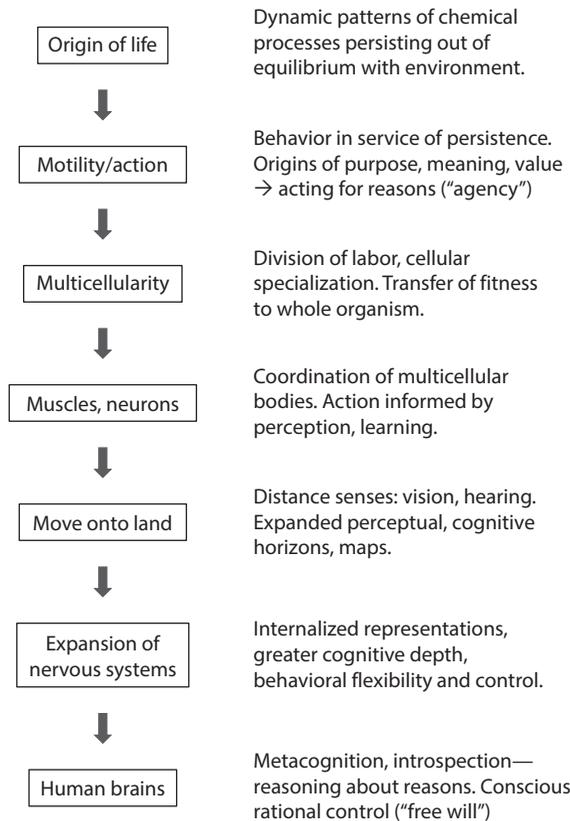


FIGURE 1.3. The evolution of agency and free will. The major stages of evolution of perception, cognition, and behavioral control.

various sensory signals to specific actions in this repertoire, hardwiring adaptive instincts for approach or avoidance.

With the elaboration of the nervous system, this kind of pragmatic meaning eventually led to semantic representations. Perception and action were decoupled by layers of intervening cells. Instead of being acted on singly and immediately like a reflex, multiple sensory signals could be simultaneously conveyed to central processing regions and operated on in a common space. Circuits were built that integrated, amplified, compared, filtered, and otherwise processed those signals

to extract information about what was out in the world and what that meant for the organism. More and more abstract concepts were extracted—not just about things but also types of things and types of relations between them. Creatures capable of understanding emerged.

Meaning became the driving force behind the choice of action by the organism. That choice is real: the fundamental indeterminacy in the universe means the future is not written. The low-level forces of physics by themselves do not determine the next state of a complex system. In most instances, even the details of the patterns of neural activity do not actually matter and are filtered out in transmission. What matters is *what they mean*—how they are interpreted by the criteria established in the physical configuration of the system. Animals were now doing things for reasons.

That causal power does not come for free: it is packed into the organism through evolution, through development, and through learning. It is encoded in the genome by the actions of natural selection. And it is embodied in the physical structure of the nervous system in the strength of neuronal connections that express functional criteria in relation to a hierarchy of aims of the organism. There is nothing here that violates the laws of physics; it just demands a wider concept of causation over longer timeframes and an understanding that the dynamic organization of a system, which encodes meaning, can constrain and direct the dynamics of its component parts.

And yes, your actions are at any given moment constrained by all those prior causes. Yet you could just as well say, more positively, that they are *informed by* prior experience. That is precisely the property that sets life apart from other types of matter: living things literally *incorporate* their history into their own physical structure to inform future action. For those who would argue this impinges on the freedom of the self to decide at any moment, I counter that it is this very process that enables the self to exist at all. There is no self in a given moment: the self is defined by persistence over time.

And though you are configured in a certain way that reflects all this history, you are not hardwired. We humans have the remarkable capacity for introspection and metacognition. We can inspect our own

programming, treating goals and beliefs and desires as cognitive objects that can be recognized and manipulated. We can think about our own thoughts, reason about our own reasons, and communicate with each other through a shared language. We can access the machine code running in our brains by translating high-level abstract concepts into causally efficacious patterns of neural activity. This gives a physical basis for how decisions are made in real time, not just as the outcome of complex physical interactions but also *for consciously accessible reasons*, and it provides a firm footing for the otherwise troublesome concept of mental causation.

So, if you want to know what kind of thing you are, you are the kind of thing that can decide. Not just a collection of atoms pushed around by the laws of physics. Not a complex automaton whose movements are determined by the patterns of electrical activity zipping through its circuits. And not an NPC, unknowingly driven by its programming. You are a new type of thing in the universe—a self, a causal agent. In the game of your life, you are Player One.

What follows is thus a full-throated defense of the idea of free will. Despite many claims to the contrary, the latest science—whether physics, genetics, neuroscience, or psychology—does not in fact imply that we have no choice or control over our actions. It's true that we are learning more and more about the mechanisms underlying our cognition and behavior—from neural systems and circuits down to the level of cells and molecules or even atomic physics. But even though our cognitive systems have a physical instantiation, their workings cannot be *reduced to* this level. We are not a collection of *mere mechanisms*. As we will see, the nervous system runs on meaning.

The fact that our capacities for cognitive control are grounded in definable biological systems does, however, have important implications for issues of moral and legal responsibility, though these are notably more subtle than the typical absolutist framing. I return to consider these and related issues in the final chapter.

Along the way, I offer a perspective on life that centers agency as its defining characteristic. What distinguishes living organisms is that *they do things, for reasons*. They behave in a truly purposeful manner. This is

not an illusion or just a convenient way of talking or thinking about them: it's the right way of thinking about them. Causation does not all bubble up from the bottom, nor is it all instantaneous. The way things are organized can and does govern the way complex systems behave. Living organisms accumulate causal power by coming to embody aspects of their history in their own structure, either through evolution or over the course of their individual lifetimes. The story of agency is thus the story of life itself, and that is where we begin, in chapter two.

## INDEX

Note: page numbers in italics refer to figures.

- acetylcholine, 131, 140
- action: and confidence level in information, 255–56; control through metacognition, 21–22, 266–67; evaluating success of, in lower *vs.* higher animals, 255–57; in living creatures, as action by whole, 274; prefrontal cortex in initiation of, 261; rational, prefrontal cortex and, 258–59; System 1 and System 2 in, 236–37, 261. *See also* causality; self-control
- action, for reasons: ability to move and, 274; clinical evidence suggesting subconscious reasons, 249–52; in higher animals, as for their own reasons, 290; human power to articulate and adjust, 257, 281–82; natural selection's shaping of organisms for, 66–67, 281; transition to individual basis for, 94, 276. *See also* agency; choice; metacognition and introspection
- action potentials (spikes), 81, 84, 171, 197
- affordances, evaluation of, as part of perception, 128
- agency: as action based on meaning, 217; as action by whole self, not by parts, 215–16; as action informed by past experience and directed toward imagined future, 217–18; as built into organisms' structure over time, 21, 23; causal integrity and autonomy as criteria for, 214–15; in *C. elegans*, 93–94; as defining characteristic of life, 18–19, 22–23, 218, 273–74; definition of, 2, 22–23; evidence of, 216–18; evolution of, 19, 20, 273–78; holistic nature of, *vs.* action by machine, 216–17; as human characteristic, 2; and life as holistic, relational system, 292; multicellular creatures as locus of, 76; as proactive rather than reactive, 217; and reasons for acting, 218; in simple multicellular creatures, 93–94; in single-celled organisms, 60–62, 67, 68. *See also* action, for reasons; choice; free will
- aggregative multicellularity: advantages of, 73–74; differentiation of cells in, 73–74; higher level of organization created by, 75–76; mechanism to ensure genetically identical cells, 75; natural selection of, 74–75. *See also* slime mold (*Dictyostelium discoideum*)
- agriculture, effect on human culture, 72–73
- amoebas: constant activity in, 274; motion in, 50, 58, 79. *See also* slime mold (*Dictyostelium discoideum*)
- anterior cingulate cortex (ACC), 191
- archaea, and mitochondria origin, 71, 72
- Aristotle, 165, 289
- artificial intelligence (AI), 295–99; and advances due to deep learning, 295; capabilities of, 295–96; and consciousness, 277; elements required to emulate natural human intelligence, 296–99; and generative models, 296; neuroscience as inspiration for, 295; wisdom of creating, 299

- ATP (adenosine triphosphate), 29–30, 32  
ATP synthase, 29, 32  
autonomy, in single-celled organisms, 60–62, 67, 68  
avoidance reaction, in *Paramecia*, 57
- Bacon, Francis, 165  
bacteria: constant activity in, 274; energy production in, 29–30, 69–70; as food for amoebas, 58; light detection in, 98; and mitochondria origin, 70–71, 72; movement in, 50, 55–57, 56, 174; myxobacteria, 73; trading of genes, 38. *See also E. coli*
- Bandura, Albert, 240–41  
basal ganglia: in action selection, 137–39, 139, 187; in learning and memory, 125, 141; location of, 124–25  
Bateson, Gregory, 168  
behaviors: consummatory vs. appetitive, 134; as only intelligible over time, 290–91  
Bell, John, 157  
Berger, Hans, 181  
Bergson, Henri, 291  
Big Bang, quantum fluctuations in, 159  
Big Five personality types, 230–32  
bilateral body plan: basic structure of, 101; evolution of, 100–101; new niches opened by, 102  
biological determinism: and character as developed through effort and experience, 241–43, 242; claim that biological structure predetermines action, 146, 218–19, 220–21; clinical evidence for, 249–52; critique of evolution or genetics as deterministic, 221–26, 281–82; critique of nurture or environment as deterministic, 226–29; critique of personality traits as deterministic, 229–35; as inconsistent with experience, 248–49; and legal responsibility, 283–86; and misidentification of self, 248; and moral responsibility, 282–83; on shaping of individuals by environment, 220; on shaping of individuals by evolution and genetics, 220; supporters of, 220, 248. *See also* upbringing and environment  
biology: as historical science, 290; as simplification of underlying physics, 15–16 (*see also* reductionism)
- bipolar cells, 100, 105  
Born, Max, 156  
Boyden, Edward, 11  
brain: anomalies, and biological determinism, 284–85; cognitive exaptation in humans, 294; as distinct from machine-like physical systems, 195–96; evolution of, 101; influence of thoughts, beliefs and desires on, 212, 213; as product of complex interaction of individual DNA and environment, 4–5, 5; top-down/semantic causation similarity to computer program, 212–14; use of information meaning to guide action, 212, 213; use of meanings and representations, 195, 196. *See also* learning and memory, brain circuitry used in; meaning; upbringing and environment, effect on brain development
- Buridan, Jean, 181  
Buridan's ass, 181
- Cambrian period, and evolution of Bilateria, 101  
cAMP, 48, 58  
CAP protein, 47, 48  
carbon atoms, and life, 27  
Carroll, Sean, 15, 148, 152  
Caruso, Gregg, 147  
causal determinism. *See* determinism, soft  
causality: abstract thinking and, 254; Aristotle on, 165; as difference that makes a difference, 154, 168; final cause and, 289; “how” and “why” questions about, 165–66, 168; living systems’ accumulation of, through evolution, 279–80; living systems as stores of causally effective information,

- 168; spiral causation in living systems, 287–88, 288, 289; triggering *vs.* structuring causes, 168
- causal relations, understanding of: brain's development of, 121–22, 130–31; survival benefits of, 121
- C. Elegans*. *See* nematode worm (*Caenorhabditis elegans*)
- cell membrane: and causal insulation of cells, 33, 273–74; emergence of, 31–32, 35; sensors on, 52–53
- cell multiplication, emergence of, 36–37
- cells: ATP generation in, 29–30, 32; emergence of, 30–33, 31; energy source for, 29–30, 32; internal regulation by DNA, 33. *See also* metabolism
- Chabris, Christopher, 263–64
- character: definition of, 239; development through effort and experience, 241–43, 242; genetic influences on, 286; praise and blame used to mold, 242–43, 283; prosocial traits as virtues, 239–40, 243, 270; self-control traits as virtues, 240–41
- chemistry: of carbon atoms, and life, 27; organic, and origin of life, 30–33, 31, 35; as simplification of underlying physics, 15–16. *See also* reductionism
- choice (action selection): as action by self based on meaning, 144; biological signals affecting, 132–35; as context-specific, 134; habit and, 136, 142, 187, 188; in human beings *vs.* machines, 1–3; increased sophistication of, with improved perception processing, 121–22; as integrated with perception, 127–28; loop of perceiving, thinking, acting, and learning in, 126; as not reducible to physical causation, 22; as process, 143; simulation of possible futures in, 135–40, 276; as spiral of action and reaction, 143; in two-stage model of free will, 187–89, 188, 192. *See also* agency; decision-making; free will choice, and possibilities that “spring to mind”: chemical signals to broaden, 190–91; control of, 135–36; noise in neural circuits and, 187–91, 188; as product of learning, 186–87
- choice, brain circuitry for, 126–31, 137–40, 139, 187–88; as constantly active, 143; neuromodulators and, 139, 139–40; and post-action learning, 140, 141–43; provision of options, not decisions, 143–44
- choice, evolution of: importance implications of, xii; as key to understanding free will, xi
- Churchland, Patricia, 240
- Cicero, 241, 243
- Cnidaria, 83
- cockroaches, randomness in escape behavior, 175–76
- cognitive niche, 252–53
- cognitive realism, on mental causation, 269
- Cohen, Jonathan, 261
- command neurons, in *C. elegans*, 89, 90
- compatibilist view of free will: critique of, 153–54, 248; as form of determinism, ix–x; supporters arguments, 150–52
- cones, 107, 108
- confabulation, in split-brain patients, 249–50
- consciousness, 262–66; definition of, 262; origin and functions of, 264–66, 277–78; prefrontal regions and, 262–63, 263; regulation of information available to, 262–64
- cortex: in action selection, 137, 139, 140, 187; and goals, storage of, 135; in learning and memory, 126–31, 132, 141; regions of, 129
- creativity, brain processes in, 191
- Crick, Francis, 25, 37, 195–96
- culture: and accumulation of knowledge over generations, 277, 294; and development of consciousness, 277–78
- cytoskeleton, in slime mold, 58, 79

- Darwin, Charles, 28, 39, 102, 189
- decision-making: decisions about future decisions, 237–38; as fundamental characteristic of higher life, 22; personal interests and aptitudes as guide in, 237–38; as process, 143. *See also* choice (action selection)
- Deisseroth, Karl, 11
- Democritus, 146, 154–55
- Dennett, Daniel, ix, 147, 150–51, 192, 248, 277, 283
- Descartes, Rene, 13, 267
- determinism: causal isolation of cells and, 33; and causal role of higher-order constraints, 280–81, 287–88; on free will as illusion, ix–x, 3; inability to explain defining characteristics of life, x–xi; and legal responsibility, 283–86; and moral responsibility, x, 282–83; multiple realizability in neural circuits and, 210–12, 211; neuroscience’s view of, 8–12; physics’ view of, 12–16; as product of Western science’s methodology, 287; as seemingly inevitable conclusion, 2–3; tautological use of cause and effect in, 281. *See also* biological determinism; compatibilist view of free will
- determinism, hard (physical predeterminism), 13–14, 15, 145–46, 280; as challenge to agency, 146, 150; as fantastical, 280; history of concept, 146–47; and limits to precision, 161; and physical indeterminacy as inherent in system, 162, 287; and physics’ understanding of fundamental particles and forces, 12–16, 148–50; undermining of, by quantum physics, 158–64, 168–69, 280
- determinism, soft (causal determinism), 14–15, 145–46; as challenge to agency, 146; critique of, 164–65
- Dictyostelium discoideum* (slime mold). *See* slime mold
- DNA: in cell multiplication, 36–37; functions in cells, 34, 35, 36; genes exchanges in early organisms, 38–39; genes on, 36; information in, and cell regulation, 33; mutations in, natural selection enabled by, 38; record of evolutionary branching in, 78; stability of, 34, 36; structure of, 33–34, 36–37; as type of memory of natural selection, 39–41. *See also* heritability; human DNA
- dopamine, 110, 131, 140, 141, 234
- Doyle, Robert, 192
- dreaming. *See* sleep and dreaming
- Dretske, Fred, 167–68
- dualism, mind-body: critique of, 13; and determinism, 3, 267; scientific rejection of soul, 16, 144; as unnecessary postulate, xi, 267–68, 269
- E. coli* (*Escherichia coli*): active response to integrated information from sensors, 61–62; metabolism control mechanisms in, 46–47, 48; motion in, sensors guiding, 55–57, 56
- Einstein, Albert, 146–47
- electroencephalograph (EEG), 181
- Ellis, George, 164, 213–14
- endosymbionts, 71
- Epicurus, 155, 159, 164, 170, 193
- escape behavior, value of randomness in, 175–77
- eukaryotes: complex cell structure of, 70; mitochondria origin and function, 70–73, 72
- Everett, Hugh, 157
- evolution: of agency, 19, 20, 273–78, 279–80; of bilateral body plan, 100–101; of brain, 101, 110, 110; coevolution of individuals and environments, 238–39; and exaptation, 293–94; leaps in, 70; and living systems’ accumulation of causality, 279–80; of mammals, 105; of meaning as driving force behind action, 21; of metacognition and introspection in humans, 252, 276; and natural selection, 37–39; of nervous

- system, and development of internalized representations, 20, 20–21, 276; of nervous system, as utilitarian, 293; new niches opened by, 101–2; of power of choice, xi, xii; as purposeless, 69, 78, 289; of world of ideas in humans, 294. *See also* multicellular organisms, evolution of; natural selection; vision, evolution of
- evolutionary explanation of free will:  
benefits of, 18; overview of, 19–22, 20
- exaptation, 293–94
- eyes: anatomy, 106–9, 108; evolution of, 100, 105. *See also* vision
- eyespot, 99–100
- faces, recognition of, 108, 114, 200, 206, 207
- fairness, innate sense of, 240
- Farahany, Nita, 283–84
- Feynman, Richard, 163
- fish, vision-based escape circuits in, 103
- flagella, 50, 55, 56, 83
- forebrain: and biological signals affecting action selection, 134; integration and analysis of sensory input, 124, 276; learning and memory functions, 125; structures in, 124–25
- fossil record, 77–78
- free will: as capacity for conscious, rational action, 282; as choice by self as defined by history, character, and goals, 244–47, 246, 278–79; as choice informed but not determined by past action, 21, 23, 279; constraint-free, as absurdity, 244–47, 246, 279; determinism's rejection of, 12–16; and evolution of layers of processing between perception and action, 67, 204, 205, 275; as evolved function with biological basis, 272; as fundamental truth demanding explanation, 18; humans' meeting of common-sense definition of, 279–80; imprecise definition of issue, 16–17, 278; as issue distinct from moral responsibility, 17–18; level of cognition and metacognition required for, 68–69; Libet's experiment on decision-making and, 183–86; moral responsibility as driver of debates on, 278; neuroscience's rejection of, 8–12; as uniquely vexing problems, 16; unresolved philosophical debate on, ix, 16. *See also* agency; choice; compatibilist view of free will; determinism
- free will, two-stage models of, 187–89, 188; present as “thick” in, 192–93, 194; selection of possible actions in, 187–89, 188, 192
- Frith, Uta and Chris, 294
- function. *See* purpose
- game theory, on unpredictability as strategy, 175
- gap junctions, 81
- Gazzaniga, Michael, 249–50
- genes, 36; exchange of, in early organisms, 38–39; expression, in regulation of metabolism, 45, 46, 47; mutations, and biological determinism, 284–86
- germ cells, 76, 275
- Gibson, James, 128
- Gisin, Nicolas, 161
- Glenn, Andrea, 285
- goal-directed behavior, in hydras, 87, 88
- goals, storage in cortex, 135
- Godfrey-Smith, Peter, 120
- GPe, 138, 139
- GPI, 137–38, 139
- Greene, Brian, x, 13
- Grillner, Sten, 103
- habits, 235–39; bad, as typical focus, 235; and reinforcement learning, 235; as shift from model-based to model-free reasoning, 236; as shortcut for decision-making, 136, 142, 187, 188, 235–36; types of response, 236

- habits, types of: behavior cued by stimulus, 236; and coevolution of individuals and environments, 238–39; long-term, following general heuristics, 237; personality-based characteristic adaptations, 238; response types cued by situation types, 236–37
- Haggard, Patrick, 15
- Harris, Sam, 7, 220–21, 247, 248
- Hawking, Stephen, x, 13
- hearing: development of, and increase in internalized representation, 97–98; evolution in land mammals, 20, 95, 252, 275; in single-celled organisms, 54; topographic relationships in neural circuits of, 205–6; usefulness for predators and prey, 97–98; visual component of, 116
- Heisenberg, Martin, 193
- Heisenberg, Werner, 155, 162
- Heisenberg uncertainty principle, 162–63, 193
- heritability, 221–26; definition of, 221–22; of personality traits, 231–32; of psychological traits, 225–26; small variations within larger human nature, 224–25; studies on, 222–23
- Heyes, Cecilia, 277
- hippocampus: in action selection, 140; and biological signals affecting action selection, 132–35; learning and memory functions, 124, 125, 126, 128–30, 132; location of, 124, 129; role in regulation of body functions, 132–33; structure of, 128
- Hippocrates, 229–30
- histamine, 140
- Hoel, Erik, 190
- Hofstadter, Douglas, 257
- hormones, 81, 109–10, 133, 140, 197, 234
- Hossenfelder, Sabine, 15
- human DNA: basic human nature provided by, 3–4; interaction with environment, 4–5; variations creating individual differences, 4
- humans: ability to plan, imagine, and evaluate using internalized representations, 252–55; agency as characteristic of, 2; and common-sense definition of free will, 279–80; evolution of metacognition and introspection in, 252, 276; evolution of senses in narrow range useful for survival, 54; evolution of world of ideas in, 294; sources of information about ancestors of, 77–78
- humans, modeled as type of machine: determinism implied by, 2–3, 7; and individual behavior as product of system complexity, 5–7; in neuroscience, 8–16. *See also* artificial intelligence (AI)
- hydras, 85–88; as Cnidaria, 83; goal-directed behavior in, 87, 88; muscles in, 85–87, 86; nerve net in, 85–86, 86; response to external signals, 87
- hydrogen ion gradients, as source of energy for all life, 29–30
- hydrothermal vents, as likely origin of life, 29–30, 31, 34–35
- hypothalamus: ancient autonomic functions of, 109–10; evolution of, 110, 110
- imagination: of future, and agency, 217–18; of mental models, for evaluation, 254; use of internalized representations in, 252–55
- imagination, consciousness and, 264–65
- immune system, natural selection in antibodies, 189
- information, 62–66; abstract, manipulation in metacognition, 252–55; bits as unit of, 63; collection by single-celled organisms, 66, 66; in living systems, possibility of being different as defining feature of, 63–64; as “meaningful facts” in colloquial sense, 64; as necessarily physically instantiated, 62; processing and responding to, as characteristic of living organisms, 62; regulation of information available to consciousness, 262–64; relative, 65–66;

- technical meaning of, 63–64; value of, in creatures capable of movement, 274. *See also* neural circuits' generation of higher-level information; senses
- Innate* (Mitchell), 3–4, 222–23
- interneurons, in *C. Elegans*, 89–90, 90
- ion channels: ATP synthase as, 29–30, 32, 69; ligand-gated, 80, 81; in light detection, 98; and muscle movement, 79; in neurons, 80–81, 82, 84, 171, 172, 173, 197; and optogenetics, 10–11; in *Paramecia*, 57, 62; protein sensors as, 52, 65; voltage-gated, 80, 81, 84
- ion gradients of hydrogen: and origin of life, 30; as source of energy for all life, 29–30
- ion pumps, 29–30, 32, 69–70, 80
- James, William, two-stage model of free will, 187–89, 188, 192
- Jonas, Hans, 298
- Juarrero, Alicia, 166–67
- Kahneman, Daniel, 236–37, 250–51, 261
- Kant, Immanuel, 215
- Kaufmann, Stuart, 101–2
- Keijzer, Fred, 120
- Laland, Kevin, 277
- lampreys: brain regions controlling action selection, 133–34, 181; vision-based escape circuits in, 103
- language, and development of consciousness, 277–78
- Lao Tzu, 241
- Laplace, Pierre-Simon, 147, 158, 166
- learning and memory: benefits in guiding action, 122; evaluation of situations and potential responses as focus of, 131; in nematode worm (*Caenorhabditis elegans*), 92–93; reinforcement learning, 141–43, 235
- learning and memory, brain circuitry used in, 126–31, 132, 142; causal relations, understanding of, 130–31; comparison of sensory input with stored schema, 127, 130; creation of hierarchical categories, 127; and episodic memory, 130; and meaning, determining, 130; memory prioritization, 131; as model for artificial neural networks, 130, 141; and modeling of self, 131; neuromodulators and, 131; spatial maps, creation and use, 130; synaptic plasticity in, 126–27, 129–31, 141–42
- leeches, randomness in escape behavior, 176–77, 177
- legal responsibility: biological determinism and, 283–86; moral responsibility and, 17; necessity of free will for, 278
- Libet, Benjamin, 8–9, 181
- Libet's experiment on decision-making processes, 8–9, 181–83, 182; design problems in, 183–84; detection of brain activity in advance of conscious decision, 183; erroneous conclusions about free will, 183, 186; Schurger's reinterpretation of data in, 184–85, 185
- life: agency as defining characteristic of, 18–19, 22–23, 218, 273–74; chemistry of carbon atoms in, 27; determinism's inability to explain characteristics of, x–xi; emergence of, 19, 28–33, 31; lack of intent or purpose behind evolution of, 41–42; metabolism necessary for, 27–28; as process organizing cells over time, 24–27, 33, 274, 290–91
- life, purpose of: as de facto product of system structure, 42; function, meaning, and value as properties derived from, 43; as self-preservation through natural selection, 42–43, 288–89
- living systems: accumulation of causal power (agency) through evolution, 19, 20, 273–78, 279–80; action for reasons as defining feature of, 66–67, 281; components of, as processes rather than parts, 215–16, 274;

- living systems (*continued*)  
as holistic, relational system, 291–92;  
isolation from surroundings, 196, 273–74;  
as locus of concern, 298; new technologies for study as holistic, relational system, 292–93; spiral causation over time in, 287–88, 288; as stores of causally effective information, 168, 217–18
- locus coeruleus (blue spot), 190–91
- mammals: costs and benefits of being warm-blooded, 105; evolution of, 105; land, evolution of senses in, 20, 95, 105, 252, 275; tectum in, 109. *See also* visual image processing in mammal brain
- many worlds hypothesis, 157
- MAOA gene mutations, effects of, 284
- Maoz, Uri, 186
- maps. *See* spatial maps, internalized
- Margulis, Lynn, 71
- marine plankton, light detection in, 99
- materialism, reductive, on mental causation, 269
- “A Mathematical Theory of Communication” (Shannon), 63–64
- McAdams, Dan, 238
- McCulloch, Warren, 198–99
- meaning: act of interpretation required for, 66–67; causal power of, 290; as characteristic of information, 64; choice as action by self based on, 144; as contextual, 290; origin of, 19, 20, 43; for single-celled organisms, 66, 66–67; in single-celled organisms, as expressed in control policy, 203, 205; value for guiding behavior, 121. *See also* understanding
- meaning, as driving force behind action, 21, 212, 213; and breaking of chain of determinism, 21; evolution of, 21
- meaning, creation by neural circuits: as about something consequential, 202–3; brain areas involved in, 206; effect of immediate attention on, 212; grounding in organism’s learning from past experience, 194–95, 212, 217–18, 276; and increased internalization and personalization of perception, 121, 142; as inherently relational, 207–8; linguistic representations of, 206–7, 207; maps used to recognize and categorize information, 206–8, 207; and mental vs. neural representations, 208–10; and mind’s experience of actionable meaning, 209–10; and multiple layers of processing in complex systems, 203–4; and multiple realizability, 210–12, 211; as shaped by natural selection, 194–95, 203, 212, 217–18; and space of decision between picture of world and goals/intentions, 204, 205; use of internal representations, 203–4
- Meincke, Anne-Sophie, 291
- memory. *See* learning and memory
- mental causation: as brain responding to meaning, 268–69, 269; dualism on, xi, 267–68, 269
- mental models: imagining and evaluating, development of power for, 254; and metacognition and introspection, 254; of self, 131, 254–55; testing and updating of, 256. *See also* spatial maps, internalized
- Mermin, David, 16
- messenger RNA, 34
- metabolism: as characteristic of life, 27–28; chemical balance maintained by, 28; effect on environment, 44–45; materials and fuel required for, 27, 32
- metabolism control mechanisms, 44–47; in *E. coli*, 46–47, 48; molecular clocks, 49–50; in respiration pathway, 45–46, 46
- metacognition and introspection: control over action, 21–22, 266–67; and evaluation of actions and perceptions, 255–57; evolution in humans, 252, 276; human prefrontal cortex and, 258–61, 259; level required for free will, 68–69; manipulation of abstracted information in, 252–55;

- and model-based reasoning, 254; and power to articulate and adjust reasons, 257
- midbrain, in action selection, 187
- Miller, Earl, 261
- mind: of animals, 209; experience of sensation at level of meaning, 209–10; and mind-body dualism, xi, 267–68; as physically instantiated in neural activity, 208–9, 267; as poorly defined concept, 208. *See also* mental causation; mental models
- mitochondria: and evolution of multicellular creatures, 76–77, 94, 275; origin and function of, 70–73, 72
- model-based reasoning, 254
- molecular clocks, 49–50
- molecules, emergence of, 28, 30, 31, 34–35
- moral responsibility: as complex issue distinct from free will, 17–18; determinism and, 282–83; as driver of free will debates, 278; justification of, in compatibilism, 150–51; necessity of free will for, 278; as problem in determinism, x
- Morgenstern, Oskar, 175
- mosquitoes, sensors used to track prey, 53–54
- motor neurons, in *C. Elegans*, 89, 90
- movement in multicellular organisms: coordinating across cells, 79–81; and development of muscles and neurons, 77, 274; distinguishing movement of self from movement of visual field, 119–20; in hydra, 85–87, 86; in nematode worm (*Caenorhabditis elegans*), 89–91, 90; and value of information, 274
- movement in single-celled organisms, 50–60; as active response to integrated information from sensors, 60–62; in amoebas, 50, 58, 79; in bacteria, 50, 55–57, 56, 174; evolution of sensors to guide, 51–54; as purposeful, 67; random motion, 51; sensing and moving toward or away from stimulus, 54–60, 56, 59; types of propulsion, 50
- Mudrik, Liad, 186
- multicellular organisms: advantages of, 76–77; aggregative multicellularity as type of, 73–76; cell replication without separation as type of, 76; chemical signaling in, advantages and disadvantages of, 79–80; and coordination of individual cells, 77; and economies of scale, 77; electrical signaling in, 79–81; and evaluation of information over longer time frames, 67, 68–69, 124, 252, 275, 293; internal representation of information in, 67, 203–4; as locus of agency, 76; movement in, and development of muscles and neurons, 77; simple forms of cognition in, 95; specialization of cells, 76, 275. *See also* movement in multicellular organisms
- multicellular organisms, evolution of, 19, 20, 275; development of layers of processing between perception and action, 67, 204, 205, 275; and energy barrier, 69–73, 275; as leap, 70; reasons for, 69; slime mold as model for, 58, 59; symbiosis and, 70–73, 72, 275
- multiple realizability, 210–12, 211
- muscles: coordination of action in, 79, 83; electrical impulses controlling, 79; evolution of, 83, 275; in hydras, 85–87, 86; structure of, 79
- myxobacteria, 73
- Nash, John, 175
- natural selection, 37–39; discovery of functional architectures by, 167; DNA as type of memory of, 39–41; and extension of causation over time, 167–68; as feedback loop shaping living organisms, 66–67, 166–68; higher-level principles constraining, 166–67; as incorporation of information about past events into structure, 39–41, 49–50, 167–68; positive and negative selection, 39–41, 40; as progressive process, 167; purpose of life as self-preservation

- natural selection (*continued*)  
through, 42–43; shaping of nervous system by, and generation of meaning, 194–95, 203, 212, 217–18; shaping of organisms to act for reasons, 66–67, 281; value for living things as defined by, 66–67
- neglect, in brain lesion patients, 249
- nematode worm (*Caenorhabditis elegans*), 88–94; intermediate stages of information processing in, 91; learning and memory in, 92–93; movement in, 89–91, 90; nervous system in, 88–92; nervous system layers, 89, 90; neurons monitoring internal states, 91–92; simple form of internalized representation in, 97, 252; structure of, 88–89
- neocortex: architecture of, 110–111, 112; evolution of, 110, 110, 276; expansion mechanism of, 111, 114; identification of meaning and mapping as functions of, 121; visual processing in, 109, 111–14
- nervous system: hierarchical, as conducive to learning, 93; as highly adaptive by design, 173–74; of hydra, 85–86, 86; of nematode worm (*Caenorhabditis elegans*), 89–92, 90
- nervous system, evolution of: and development of internalized representations, 20, 20–21, 276; as utilitarian, 293
- nervous system, noise in, 170–75; and adaptive behavior, 173–75; and artificial intelligence, 299; benefits for learning, 190; and choice in low-stakes decisions, 180–86; circuits controlling, 189–90; and cognitive permutations, 191; control mechanisms for, 175, 190–91; and creativity, 191; locus coeruleus (blue spot) and, 190–91; neurons' integration of inputs and, 197–98, 202; noise reduction strategies, 172–73; and possibilities that “spring to mind,” 187–91, 188; sleep and dreaming as means of maintaining, 175, 190; sources of, 170–72; and space for learning, 174–75; and two-stage model of free will, 187–89, 188, 192; and unpredictability as adaptive behavior, 175–77, 177; uses of, 175, 202; and variation in brain development, 226–27
- neural circuits: ability to accumulate evidence, 180; competition among patterns in, 180; configuration in critical state, 173; logical operations performed by, 198–99; mind as product of, 208–9, 267; network configuration as constraint on response to inputs, 178–80, 179; as specifically designed for selection of action, 280; and spiral causation in living systems, 287–88, 288, 289; topographic relationships in anatomy of, 204–8. *See also* learning and memory, brain circuitry used in; meaning, creation by neural circuits
- neural circuits' generation of higher-level information, 198–202; computational efficiency of, 200–201; condensing of data groups into named things, 201, 253–54; influence on action, 275–76; maps used to recognize and categorize information, 206–8, 207; and metacognition, 253–55; necessity of layers of processing in complex systems, 203–4; pattern recognition and comparison, 200; and space of decision between picture of world and goals/intentions, 204, 205; use of categories to generate understanding, 201–2. *See also* visual image processing in mammal brain
- neuroimaging: analysis of neural circuits using, 8; and neurological activity predicting action, 8–9
- neuromodulators, 131, 138, 139, 190–91, 197, 234
- neurons: assemblies/ensembles of, 128–29; cell structure, 82–83; and coarse graining of information, 198; communication between, 82; coordination of multiple outputs, 83; dendrites and axons, 82;

- evolution of, 19–20, 20, 275; generation of electrical signals, 80–81, 171–72, 197; integration and interpretation of inputs, 83, 197–98, 202; noise in signals of, 170–72; noise reduction strategies, 172–73; organisms' normalization into responsive range, 175; refractory period, 81; resting potential, 81; small range of inputs to, 197; synaptic transmission by, 84, 171; types, in *C. elegans*, 89, 90; variability in response, 173. *See also* nervous system
- neuroscience: and clinical evidence suggesting subconscious reasons for action, 249–52; creation of action and thought by stimulation of neurons, 9–12, 10; deterministic view of human choice, 8–12; optogenetics in, 10–11. *See also* neuroimaging
- neurotransmitters, 81, 82, 84, 171–72, 197
- Newton, Isaac, 146, 158
- Noble, Denis, 189
- Noble, Raymond, 189
- noradrenaline, 140, 190–91
- norepinephrine, 234
- normativity, 289–90
- ocelli, 103
- On the Freedom of the Will* (Schopenhauer), 220
- opsins, 98
- optogenetics, in neuroscience, 10–11
- organelles, 70
- osmotic pressure, 80
- Pals, Jennifer, 238
- Paramecia: active response to integrated information from sensors, 61–62; movement in, sensors guiding, 57
- Pauling, Linus, 191
- Penfield, Wilder, 9
- perception: action selection integrated with, 20, 104, 127–28; as actively interpreted even in simple organisms, 55; coevolution with means of action, 102; complexity of, and complexity of thought, 95; interpretation of, knowledge required for, 125; neural circuits to evaluate accuracy of, 256; pressures for improvement in, 76–77; processing according to expectations, 116–19, 117. *See also* hearing; senses; vision
- perception, increased processing complexity in: and decoupling from action, 20, 67, 91, 109–15; and increased internalization and personalization, 121, 142; and more sophisticated choice mechanisms, 121–22
- personality, definition of, 239
- personality traits, 229–35; classifications of, 229–31; effect of upbringing and environment on, 228–29, 231–32; heritability of, 231–32, 235; as not deterministic of action, 235; as product of variation of multiple neural parameters, 232–35, 233; study of brain areas involved in, 234
- pharmaceutical industry, focus on biological components rather than whole, 292
- photoreceptors: in animals, 106–7, 108; in simple organisms, 98–99
- physical indeterminacy: and freedom to choose in thick present, 192–93, 194; quantum physics and, 158–64, 168–69. *See also* free will, two-stage models of; nervous system, noise in
- physical predeterminism. *See* determinism, hard
- physics: relational views on, 291; understanding of fundamental particles and forces in, 148–49. *See also* quantum physics
- physics, hard determinism implied by, 12–16, 148–49; inability to define life, x; as reductionism, x
- Pitts, Walter, 198–99
- plankton, light detection in, 99
- planning ability: brain's development of, 122, 142; survival benefits of, 122; use of internalized representations in, 252–55

- Poldrack, Russell, 271–72
- polygenic scores, and legal responsibility, 285–86
- Popper, Karl, 137
- prefrontal cortex: behavior regulation, 260; constant communication with other brain regions, 260; effects of damage to, 259–60; evolution of, 259; and human metacognition and introspection, 258–61, 259; and initiation of action, 261; and working memory, 260
- prefrontal regions, and consciousness, 262–63, 263
- present, as “thick” in two-stage models of free will, 192, 280
- primary visual cortex, 108, 113–14
- priming, psychological, 250–51
- process philosophy, 291
- prokaryotes: characteristics of cells, 70; evolution of, 70–73, 72, 275
- proteins, as sensors, 52–53
- purpose: of cell functions, as continued life of cell, 44; meaning and value as defined by, 66, 289–90; origin of, 19, 20, 42, 43
- purpose of life: as de facto product of system structure, 42; function, meaning, and value as properties derived from, 43; as self-preservation through natural selection, 42–43, 288–89
- quantum decoherence, 158
- quantum physics: and determinism, 148–49; and determinism, undermining of, 158–64, 168–69, 280; and Heisenberg uncertainty principle, 162–63; influence on larger systems, 158–61, 160; nature of particles in, 155–57
- quantum to classical transition, 158–61, 160
- Raine, Adrian, 285
- rational action, prefrontal cortex and, 258–59
- reasons for acting. *See* action, for reasons
- receptor proteins, 82
- reductionism: indeterminacy in physical systems and, 164, 281; mechanistic view of life, *x*; physicists’ strict determinism as, *x*
- reinforcement learning, 141–43, 235
- representations, internalized: development of, 20, 20–21; and human ability to plan, imagine, and evaluate, 252–55; increase in, with development of vision and hearing, 97–98, 252; spatial maps created by simple creatures as rudimentary form of, 95–97. *See also* spatial maps; vision, internalized representations of
- repressor proteins, 47, 48
- reproduction: asexual, 37, 38–39; DNA and, 36–37, 38, 39–41, 40; and natural selection, 37–41
- reptiles, evolution of, 105
- respiration pathway, control mechanisms for, 45–46, 46
- retina, 107, 108, 113
- retinal, 98
- retinal axons, 108
- retinal ganglion cells (RGCs), 106–9, 199–200
- RGCs. *See* retinal ganglion cells
- RNA: functions in cells, 34, 35, 35–36; structure of, 33–34
- RNA World hypothesis, 34
- rods, 107, 108
- Rovelli, Carlo, 155, 291
- Ryan, Tomás, 11
- saccades, 119–20
- Santo, Flavio del, 161
- Schopenhauer, Arthur, 7, 220
- Schrödinger, Erwin, 25, 36, 155–56
- scientific methodology, Western: determinism as product of, 287; focus on biological components rather than whole, 291–92
- second law of thermodynamics: life as process in opposition to, 25–26; and living organisms’ contribution to entropy, 27
- self: defined as pattern persisting over time, 21, 26, 245, 279; as defined by constraints

- of history, character, and goals, 244–47, 246, 279; development of power for mental modeling of, 131, 254–55; self-awareness, and necessity of including self in stored maps of world, 120; as web of relations, 291
- self-control, 268–72; brain injury and, 272; development in children, 268–69; genetic and environmental origins of, 271; as indication of free will, 282; individual variation in, 269–70, 282; moral value attributed to, 270–71; and success in life, 270, 271–72
- self evaluation, development of ability, 255–57
- senses: evolution with changing needs in new niches, 102; and origin of reasons for action, 274; value of, in creatures capable of movement, 274
- senses, human, evolution of narrow range useful for survival, 54
- senses in simple multicellular creatures: as limited to immediate environment, 95; spatial map created by, 95–97
- sensors, evolution of: and embedding of behaviors into genome, 53; selection only for information useful for survival, 53–54; in single-celled organisms, 52–54
- sensors in single-celled organisms: evolution of, 52–54; for light detection, 98; sensing and moving toward or away from stimulus, 54–60, 56, 59; spatial map created by, 55, 62; transfer of information by, 66
- sensory neurons, in *C. Elegans*, 89, 90, 90
- serotonin, 131, 140, 234
- Shannon, Claude, 63–64
- Shurger, Aaron, and Libet's experiment on decision-making, 184–85, 185
- signal transmission, Shannon's work on efficiency in, 63–64
- single-celled organisms: agency and autonomy in, 60–62, 67, 68; as complex, dynamic systems, 60; energy barrier limiting growth of, 69–73, 285; evolution of sensors in, 52–54; information, meaning, and value in, 65, 65–67; limited scope of perception and action in, 68; meaning for, as expressed in control policy, 203, 205; senses of smell, hearing, and sight in, 54; small range of senses in, 196. *See also* movement in single-celled organisms; sensors in single-celled organisms
- sleep and dreaming, as means of maintaining randomness in neural networks, 175, 190
- slime mold (*Dictyostelium discoideum*): life cycle of, 58, 59; mechanism to ensure genetically identical cells, 75; sensing and moving toward stimuli, 58. *See also* aggregative multicellularity
- Smolin, Lee, 159
- SNr, 137–38, 139
- soul, existence of. *See* dualism, mind-body
- spatial maps, internalized: brain circuitry used in, 121, 130; creation and use in learning and memory, 130; identification of meaning and mapping, 121; image processing in mammal brain and, 120, 204, 252; inclusion of self in stored maps of visual world, 120; rudimentary form in simple creatures, 95–97; in single-celled organisms, 55, 62; stored in primary visual cortex, 113, 119; use to recognize and categorize information, 206–8, 207. *See also* mental models; representations, internalized
- split-brain patients, confabulation in, 249–50
- sponges, 83
- stable attractor states, 211
- state space, restricted trajectories through, 177, 178, 200
- Steward, Helen, 194
- striatum: in action selection, 137–39, 139; in learning and memory, 141
- Stroop task, 258–59
- subjective experience, origin and functions of, 264–66

- superior colliculus, 109
- symbiosis, evolution of, 70–73, 72, 275
- synapses, 82
- synaptic plasticity, 126–27, 129–31, 141–42
- tectum, 110; in fish, 103; in fish, birds, amphibians and reptiles, 109; integration and analysis of sensory input, 123, 123–24; in lampreys, 123; linking of sensory input with action, 122–24, 123; in mammals, 109
- telencephalon, 110–11
- teleology, cosmic, scientists' rejection of, 166
- thalamus: in action selection, 187; learning and memory functions, 124, 125; in vision processing, 108, 113
- theory of relativity, 146–47
- time: and agency as action informed by past and directed toward future, 217–18; behaviors as only intelligible over period of, 290–91; biology as study of life within, 290; as composed of flux, 291; and culture as receptacle for knowledge over generations, 277, 294; and decisions about future decisions, 237–38; and DNA as type of memory of natural selection, 39–41, 78; Einstein on, 146–47; expenditure of, as factor in animals' choice, 134, 136; free will as choice informed but not determined by past, 21, 23, 279; habit as savings of, 142, 235–36; imagination's ability to run in fast forward, 254; life as process organizing cells over time, 24–27, 33, 274, 290–91; meaning's extension in time, 290; and molecular clocks, 49–50; multicellular organisms' evaluation of information over longer periods, 67, 68–69, 124, 252, 275, 293; natural selection's extension of causation over, 167–68; and natural selection's incorporation of information about past events into structure, 39–41, 49–50, 167–68; present as “thick” in two-stage models of free will, 192–93; self defined as pattern persisting over time, 21, 26, 245, 279; simulation of possible futures in choice (action selection), 135–40, 276; spiral causation over time in living systems, 287–88, 288; time required for thinking, in humans *vs.* artificial intelligence, 297; visual systems' tracking and interpretation of information through, 105–6, 121. *See also* determinism; Libet's experiment on decision-making processes; natural selection; planning ability
- Tse, Peter Ulric, 212
- Tversky, Amos, 261
- understanding: brain's use of categories and relations to create, 201–2; of causal relations, brain's development of, 121–22, 130–31; of causal relations, survival benefits of, 121. *See also* meaning
- upbringing and environment, effect on brain development, 226–29; natural variations in brain development and, 226–28, 227; as small variations within context of human nature, 228–29
- upbringing and environment, effect on personality traits, 228–29, 231–32
- Urbilateria, 102. *See also* bilateral body plan
- value: as defined relative to a purpose, 66, 289–90; of information in creatures capable of movement, 274; for living things, as defined by natural selection, 66–67; origin of, 19, 20, 43; for single-celled organisms, 66, 66–67
- Verde, Clelia, 159
- video games, and human *vs.* machine choices, 1–3
- vision: internalized representations of, as later development, 104; in simple animals, as directly linked to survival response, 104; in single-celled organisms, 54
- vision, evolution of, 97–105, 100; early light-based warning systems, 103–4;

- early organs of light detection, 98–100, 100; evolution of eye anatomy, 100, 105; and increase in internalized representation, 97–98; in land mammals, 20, 95, 105, 252, 275; problem of explaining, 102; from single common ancestor, 102–3
- vision in animals: color perception, 107; image processing, pre-brain, 105–9, 108
- visual cortex, 108
- visual image processing in mammal brain, 109–15; and creation of topographic map, 204, 252; and decoupling of image from action, 115; distinguishing movement of self from movement of visual field, 119–20; as extraction of meaning, 118–19, 199–200; inclusion of self in stored maps of visual world, 120; and incorrect interpretations, 116, 117; as inferences, 115; maps of visual world stored in primary visual cortex, 113, 119; moving to change visual field, 115–19; role of expectations in, 116–19; visual information integrated into memory, 114
- von Neumann, John, 175
- Watson, James D., 25, 37
- What is Life?* (Schrödinger), 25, 36
- Whitehead, Alfred North, 291
- “why” questions, as “how” questions in broader frame, 66–67
- Wilczek, Frank, 155, 162–63
- Wisconsin Card Sorting Test, 259, 271
- world of ideas, evolution in humans, 294
- yeasts, single-celled, respiration pathway control mechanisms in, 45–46, 46
- Yuste, Rafael, 85
- zebrafinches, song learning in, 189–90
- zygote, 76