Was Samuel Johnson a Robot?

GEORGE MAKARI

Determined: A Science of Life without Free Will, by Robert Sapolsky, Penguin Press.

Free Agents: How Evolution Gave Us Free Will, by Kevin J. Mitchell, Princeton University Press.

While on a stroll, Samuel Johnson and James Boswell fell into debate. Had Bishop Berkeley really proven the nonexistence of matter? With a swift kick, Johnson sent a stone flying. "I refute it, *thus*," he declared triumphantly.

Who is not cheered by this hurrah for common sense? And who in our present moment dares trust such a judgment? So many formerly unimaginable transformations have taken place, constantly expanding what we accept as predictable and obvious. Surrounded by flux, we now must consider how incredible possibilities might be as real as a rock. With the slow-motion collapse of the postwar order, a renewed fear of global pandemics, economic progress leading to ecological decimation, and the rise of computers that seem ever more human, where is solid ground?

Amid this general turbulence, a debate that long occupied a dusty corner of philosophy has come into public view. The contest between determinism and free will is hardly new; it is at least as old as Western philosophy. In recent decades, it has been waged under the banner of compatibilism. Compatibilists, of which there are a wide array, propose different ways to accommodate both free will and a fully determined world. Incompatibilists, who also come in different shapes and sizes, don't.

We now have a heightened stake in this tug-of-war. With artificial intelligence poised to become ever more powerful and with bots that eerily mimic our language and behavior, questions regarding human nature and its difference from machines become urgent.

Enter a distinguished neuroscientist, Robert Sapolsky, a "hard incompatibilist" who has written a book to prove that "*the world is deterministic and there's no free will*." Samuel Johnson's foot, he would have us believe, struck that stone not due to a quick-witted thought, but because of preconscious neural and hormonal causes. It was no different from the blind thrust of a motor. Nothing happens because of meaning or conscious intention. Therefore, ethical systems based on the guilt and innocence of individuals are ridiculous.

◆ ◆ ◆

Sapolsky's widely publicized book is a fascinating cultural artifact, a window into our tumultuous present, and, perhaps, an unintended example of a paradigm on the road to collapse. In the early 1980s, he trained in the laboratory of Bruce McEwen, the trailblazing Rockefeller University researcher who worked on the neurobiology of stress. Stress, McEwen and others recognized, could alter hormones, genetic expression, and brain anatomy. This catchall concept allowed investigators to consider kinds of social and environmental influences. Whether an animal was exposed to hunger, fright, isolation, violence, or a host of other pressures, its neurobiological reactions could be tracked. Stress in, altered biomarkers out. McEwen considered himself a "molecular sociologist"; his research methods cohered nicely with behaviorism and E. O. Wilson's sociobiology, where whatever happened in the black box of the mind was not relevant.

Sapolsky took this elegant, if limited, experimental approach into the wild and spent years studying baboon stress responses. For example, he would sneak up on the unlucky primates and blow-dart them into unconsciousness, then measure physiological differences related to things like social hierarchies. This work won him accolades and furthered our understanding of the chemical modulators of anxiety and aggression. When not in places like Kenya doing fieldwork, this charismatic, wild-haired Stanford University professor also wrote articles and books for the public.

Popular science writing is a difficult craft. Deep dives into data, methods, and skeptically tested, hedged, probabilistic conclusions

don't make for fat story lines. If you like tales with a beginning, middle, and end, the problem with a lot of science is that it's mostly middle. Results are tentative, in need of reproduction, and correlations that emerge in complex domains like the mind/brain generally are not conclusive. Despite wonderful exceptions, scientific work rarely contains eureka moments.

Successful science writers manage this narrative trouble by veering off into the investigators' personalities or ushering us into the adventure of scientific inquiry. The worst popularizers grab hold of our interest with hyped-up claims that temporarily sell books, but quickly fail in terms of veracity. In the process, frothy media exaggerations blur the line between advertising and solid knowledge. The consequences of this became clear during the pandemic, when we realized that a great deal of a critical social resource, confidence in scientific authority, had been squandered.

For years, Robert Sapolsky could not be accused of such recklessness. Yes, he had a penchant for speculating far beyond his data to consider things like psychiatric illness, the causes of violence, and even the sources of ethnonationalism. But he took care to mark the differences between conclusions reaped from animal models and inferences about human beings. Yes, he entertained shocking propositions of social engineering, like considering the "promise of administering oxytocin [the hormone associated with maternal bonding] far and wide" or "banishing testosterone" (i.e., men) to diminish human violence. However, he wisely rejected these Orwellian ideas due to the risk of unintended consequences.

No more. *Determined* is less a work of meticulous epistemology than a Molotov cocktail. Down-to-earth colloquialisms—the text is littered with expressions like "bummer" or "total downer" or "yeah, right"—don't conceal the Olympian confidence that supports this author's conviction, which is this: the age-old argument between determinism and free will is not philosophical; it is a scientific question that has been resolved. Humans have no free will. Not less than we might have believed. Not a little. *None*. Our actions are all caused by physical forces outside of our awareness. Genetic, hormonal,

nonconscious neuronal and environmental factors choose, and then a short time later, we delude ourselves into thinking we have chosen.

To understand the vast implications of such a radical view, we need to go back four centuries. For it was in the early modern period that beliefs regarding determinism and free will, as well as the status of the executor of that conscious will, the mind, became foundational. *Determined*, though the author seems only partly aware of this, would overturn much of that and transform the way we live.

◆ ◆ ◆

With the advent of Western modernity, humankind was said to lose its soul. The synthesis of ancient Greek and Christian thought forged by St. Thomas of Aquinas no longer held sway over influential pockets of Western Christendom. Aquinas had proposed that alongside the material souls possessed by living things and animals, humans were uniquely endowed with a rational soul that accounted for thought, memory, and will. That soul was immaterial, hence a candidate for eternal life. And by the dawn of the seventeenth century, it was in trouble.

Thanks to what would later be called the scientific revolution, this supernatural notion of humanity's essence—and the culture it supported—began to crack. In Paris, Marin Mersenne, a brilliant Minim friar, hosted salons that helped generate responses to the challenges facing the rational soul. This devout man desperately wanted to preserve that soul, while accommodating science's increasing claims over nature. To do so, he became an advocate for Galileo's "mechanical" philosophy. Nature, Galileo proposed, could be seen as a machine. All natural events could be accounted for by prior, determined action. All, Mersenne concluded, except for one. The psyche was not a wind-up toy, but also had the power to do the winding up. Mechanical philosophy made sense for everything except inner life, where free will clearly existed.

Mersenne's younger colleague, René Descartes, followed a similar path, adopting mechanics for Nature but not the immaterial Cogito. Stars, planets, animals, and plants were part of a grand mechanism,

but inner life was of another sphere. Grave trouble lay ahead for the Frenchman when he tried to sort out how an immaterial entity could control a material body, but nonetheless, his dualist solution—a compromise between science and the Catholic Church—was gradually widely adopted.

This is the kind of compatibilism that Sapolsky loves to hate. Most thinkers committed to a naturalistic science of the inner world—I count myself as one—also reject it. But the author of *Determined* assumes *any* claim about the mind is Cartesian, and therefore supernatural hocus-pocus. That's wrong. Two of Mersenne's other distinguished guests, Thomas Hobbes and Pierre Gassendi, both came up with natural alternatives. They thought Descartes's dualism was a cowardly turning away from what was heretical but clear: cognitive functions were brain-based. Physical determinism and an intentional mind were compatible. How? Gassendi gave the only honest answer: *who knows?* With the study of inner life in its infancy, Gassendi allowed for continued work on this paradox without eliminating it. He originated an agnostic position by which the mind and brain were assumed to be dual aspects of the same substance, though how that worked remained a mystery.

The history of science is replete with examples of such generative dodges, by which investigation continues by bracketing unknowables. In this way, researchers don't prematurely discard complexity and instead allow for continued inquiries that may end up in surprising places. At Oxford, Gassendi's big shrug did exactly that. It influenced the father of neuroanatomy, Thomas Willis, to propose that some functions of the rational soul were *somehow* brain based. One of his medical students, John Locke, in his lecture notes, noted the God-given capacity of passive brain matter to somehow become active *mind*. Locke then repurposed the Old English word *mynd* to be, in contrast to Descartes's Cogito, embodied, "a thinking thing."

Locke's mind could be acted on—as a doctor, he well knew about delirium and other altered states—but it also could freely think, choose, and act. These functions were natural, as was their dysfunction, which now entered the realm of mental health or illness,

rather than soul possession. Mental life, for Locke, was composed of external inputs and internal reflection, not divinely endowed conceptions. Therefore, human understanding remained partial and contingent, and no justification existed for absolute religious or political authority. Similarly in the public square, freely choosing minds required toleration, since all claims were made by imperfect knowers. Progressive visions of poverty, crime, and some forms of illness also provided the hope that active minds in error could be reformed. Secular ethics were rebuilt around this inner agency's capacity to make moral choices. Egalitarianism and democracy made sense since all minds had the capacity for reason, no matter their creed or race. Locke's bigotry against Jews, Quakers, Catholics, black people, and others made it impossible for him to follow this tenet to its rightful conclusion, but gradually others did. After the French and American Revolutions, secularism spread based on social contracts predicated on enlightened minds that had some modicum of free will.

◆ ◆ ◆

Having rid us of our souls, science now dictates, according to Sapolsky's logic, that we lose our minds. With them go the post-Enlightenment culture built around liberty. There is no capability, he insists, for the exercising of ideas, volitions, or desires. Instead, the professor promises to teach us "the science of why there is no free will, and the science of how we might best live once we accept that." This kind of talk might work with baboons and undergraduates, but it is, of course, nonsense. Even at its most elastic, "science" at present cannot falsify free will, solve the mind/brain problem, or use its methods to provide us with guides to a mindless, good life. Nonetheless, this neurobiologist sets out to prove his point by hightailing it to chaos theory, emergent complexity, and quantum indeterminacy. His conclusion? If we "*put all the scientific results together from all the relevant scientific disciplines*," he writes, "there's no room for free will." Room? Perhaps, a new room should be built for uncertainty. Call it the Gassendi suite.

The arguments in *Determined* read like a highway pileup. This examination of determinism and free will never defines either term. Messy, polemical, filled with straw-man arguments, and unscholarly in its dependence on secondary sources, the book engages in either/or thinking and a lack of nuance that obscures a host of more reasonable positions. Dubious definitions lead to downstream conclusions that are, in essence, language games.

Mental life holds no interest to this behaviorist. Samuel Johnson's witty, lower-extremity motion, we are to believe, was determined by the exact forces that led the highly trained footballer, Mia Hamm, to kick a whizzing cross coming at her at one hundred kilometers an hour, which is the same as Tolstoy's plotting out *War and Peace*, which is no different from the reactions of subjects in the famous experiments conducted by Benjamin Libet, where quick repetitive choices seemed to be signaled in the brain, seconds before thought. Really?

For this reader, it was impossible not to get hung up on a more immediate paradox. Doesn't the author's writing of this opus negate his own claims? How did these pages get written? Were his words *literally* chosen by his genes, hormones, and environmental conditioning? Or did an intentional self by the name of Robert Sapolsky have *something* to do with the pages of his book? And if he didn't, who gets the royalties?

Sapolsky admits to an "almost religious awe" for ideas, but responsible science depends on skepticism, the rules of empiricism, testing variables, hypothesis, and confirmation. It allows for the building of tentative models from facts, but reserves truth claims for verified results. *Determined* doesn't stick to those methods, yet it keeps invoking science as a bludgeon that seems to be in the author's sole possession. Those who don't agree with him ride with the witches at midnight.

All this has the characteristics of a late-stage ideology—while it is riddled with contradictions, reductive determinism still has blinkered partisans who act as if they command unquestionable truths. And so, it comes rather as a relief that as *Determined* appeared, so too

did a book by Kevin Mitchell called *Free Agents: How Evolution Gave Us Free Will*. Mitchell reveals what may be the source of Sapolsky's supreme confidence: he is preaching to the converted. The idea that free will was an illusion, Mitchell discovered, was "fashionable" with his neuroscience colleagues.

Mitchell makes quick work of these assumptions. Yes, a commitment to determinism is central to any science, but failing to distinguish between *kinds* of determinism drives some to accept a specific variety which leads to preposterous conclusions about the mind, the self, agency, and our limited but very real capacity for choice. Sapolsky's reductive determinism—known in mind/brain debates as eliminative reductionism—erases mental causes by attributing them entirely to the nonconscious brain. There is no top-down regulation by which ideas cause anything. In this way, psychology melts into neuroscience, but why stop there? Ultimately such reductionists end up talking about particle physics. That kind of determinism is hardly true to biology, since it eliminates any specific causal power inherent in living things. Rather, it forces us to attribute my choice of dinner tonight—much less all future events—to the Big Bang. Summing up, Mitchell writes, "the claim is absurd."

*Free Agent*s presents a fresh synthesis of how the same core sciences that Sapolsky uses—neurobiology and evolutionary theory—offer an account for willful causation. Open about where facts end and models begin, Mitchell builds our trust by deploying the conditional tense. Like Sapolsky, Mitchell thinks science should resolve the determinism-free will debate, but he is a compatibilist. Accounting for much more biological data than his rival, Mitchell constructs a compelling, coherent account for the evolutionary development of agency, not just in humans but also other living things. He goes about the hard work of accruing data that show how biological entities, from microorganisms to primates to humans, slowly evolved to possess complex, encoded loopback effects. To survive, primitive organisms model, predict, and act in a world that constantly acts on them. Processing information, Mitchell argues, is a biological commonplace. After millions of years of evolved feedback loops, organisms are "configured to *do things for*

reasons." Active causation, he argues, is to be found within the living, and its apex is our capacity to think, choose, and create.

Mitchell's book was heartily endorsed by the prominent neuroscientist Cori Bargmann and the cognitive psychologist Steven Pinker, but will this paradigm win the day? *Free Agents* is far more convincing than *Determined*, but the idea of humankind as automatons may continue to be hoisted up by Sapolsky's ideological allies. It has happened before.

◆ ◆ ◆

While the fiery showman Julien Offray de La Mettrie got nowhere with his 1748 work *Man-Machine*, it became a dry run for an influential group of determinists a century later. In 1847, a number of young German scientists banded together with a credo: no more speculations about minds. Given the mystified conceptions of *Geist* that had proliferated, eliminating that realm was a relief, and led them to propose that all life processes could be entirely explained by chemistry and physics. The Biophysics Movement attracted renowned investigators like Ernst Brücke, Hermann von Helmholtz, and Emil du Bois-Reymond; empowered by new tools like enhanced microscopy, they made a slew of new physiological and anatomical discoveries.

Yet as the end of the century neared, optimism about the power of such reductive determinism waned. Embarrassment followed from extravagant efforts to use the biophysics mantra to explain poetry or ethics. Sigmund Exner tried to construct psychic life from reflexes and physiology. William James and others began to decry such efforts. Finally, du Bois-Reymond felt compelled to admit that their approach left seven *Welträthsel*, insoluble "world-puzzles," which tellingly included the origin of intelligent thought and the question of free will.

While du Bois Reymond and his allies ceded free will and mental causation back to Gassendian uncertainty, medical disciplines did not yield so easily. Propped up by the scientific prestige of the biophysics movement, many in psychiatry—which first broadly considered determinative causes coming from the brain, the psyche, and the environment—clung to this narrow view. Degeneration theory

sprang forth from this soil. Genealogical charts—often quite fantastical—dictated who was fated to be mad and bad. Scientific racism linked degeneracy to Jews, and inescapable biological inferiority to non-Western, colonized peoples.

Determined shows no interest in these dark manifestations of reductive determinism, or in the way they influenced, for example, psychiatric genetics. The founder of that field was a German scientist deeply committed to the view that those with mental illness, criminality, alcoholism, homosexuality, and other "degenerate" behaviors were wholly determined by genes. Despite the lack of clear evidence, there was no uncertainty, no individual possibility, no hope. Dr. Ernst Rüdin, with his Nazi colleagues in the Aktion T4's "involuntary euthanasia" program, then murdered somewhere between one hundred thousand and three hundred thousand victims.

Instead of considering the risk of inadvertently supporting such evil, the Stanford professor reserves his outrage for those who believe in. . .crime? Yes, crime. Being wholly at the mercy of factors out of their control, humans cannot be held responsible for their actions. To do so is cruel. Sapolsky is rightly distressed by the way destructive environments can distort and destroy the lives of the disadvantaged. His solution, however, steals from them, from all of us, any shred of agency and meaning, much less any hope for change, and thereby does those he would help great harm.

From murder and rape to fraud, from Adolf Hitler's mass murder to Bernie Madoff's swindling to Nelson Mandela's righteous stand that landed him in jail, all are as responsible for their actions as the sky is for rain. And this is not just an opinion delivered from the ivory tower. Chillingly, Sapolsky serves as an expert witness, schooling judges and juries about how science "proves" defendants, awash in neural and environmental inputs, have no choice. So far, his expertise has not convinced the courts, but *Determined* ends with a brash prediction. Decades from now, Sapolsky prophesizes, the truth will be obvious. Then the Californian professor presumably will be revered, while the rest of us will be regarded as "ignorant" and "heartless," like "goitered peasants who thought Satan caused seizures."

Perhaps. But another future also may be in the offing. We will soon be encountering machines that advise us about our emotions, remember our childhoods, imitate our mannerisms and speech, and seem close to human. By Robert Sapolsky's logic, that's no problem since we too are mindless, amoral machines. But what then is the fate of a democratic culture built around "life, liberty and the pursuit of happiness"? What becomes of the right to vote, to choose, individually or as a citizen, about our bodies or our families, or our lives? Any such action, based on an intention, would be a delusion. Would "crime" under Sapolsky vanish or simply take a new name? He admits that we would need protection from violent humans and so would quarantine them. For how long? Life? In advance of their violent acts? Why wait for them to commit a crime if it is *already* determined?

Determined is a dizzying, terrifying work. On the one hand, it demonstrates what it looks like when a scientific paradigm staggers forward despite being shot through with anomalies. Festooned, lauded advocates, having devoted themselves to this way of thought, still stand tall and evince no doubt. By itself, this is merely the human comedy—vanity, blind allegiance, and power—as played out in science. However, when the paradigm in question goes to the core of men and women's freedom and morality, it is no joke. Irresponsible scientism of this sort has turned monstrous before, and nowhere has this been more dangerous than among those who claim to hold "truths" that dehumanize others.

Driven by vertiginous technological advances in artificial intelligence and wild profits, our information and social worlds are on the verge of great disruption. If decades hence, historians remember *Determined*, will it be as a reductio ad absurdum of its own claims, or as a symptom of a failing social consensus regarding human freedom and the coming of a new machine-man?