On free will or the lack thereof

An interview with Robert Sapolsky

By Alexandra Mikhailova and Daniel Friedman

Cite as: Sapolsky, R., Mikhailova, A., Friedman, D.A. (2022). On free will or the lack thereof. An interview with Robert Sapolsky by Alexandra Mikhailova and Daniel Friedman, *ALIUS Bulletin*, *6*, https://doi.org/10.5281/zenodo.7394901

Robert Sapolsky
sapolsky@stanford.edu
Department of Biology,
Stanford University,
USA

Alexandra Mikhailova
smikhailova@ucdavis.edu
Center for Neuroscience
University of California, Davis,
USA

Daniel Friedman
danielarifriedman@gmail.com
Department of Entomology
& Nematology, University
of California, Davis, USA

Abstract

In this interview, Robert Sapolsky outlines his view on Free Will and related topics. The discussion anticipates his upcoming book *Determined: The Science of Life Without Free Will*. Various topics are covered at the intersection of neuroscience with philosophy, education, and the criminal justice system.

Keywords: *Free Will*, *Human Behavior*, *Legal System*, *Philosophy*, *Complexity*

In your career you have worked across areas: doing empirical field and laboratory work, crafting various kinds of articles, and teaching legions of undergraduates. Also you have written several popular science books (Sapolsky *et al.*, 2004; 2005; 2017), variously bringing attention to ulcers, hormones, and learning. Recently you have appeared on various podcasts such as with Prof. Andrew Huberman (Huberman & Sapolsky, 2022), Here We Are (Mauss & Sapolsky, 2022), and Freakonomics Radio (Levitt & Sapolsky, 2021). To pick up here with this interview where some of these previous conversations have left off, on Freakonomics Radio you said: *"I don't think we have any free will whatsoever. I think we are the outcomes of the sheer random, good and bad biological luck that each of us has stumbled into"*, and on the Huberman podcast you denied that humans have even a "*shred of free will*". For clarity in this interview, can you provide your usage or definition of free will here? What are your definitions of consciousness and awareness?

These are great questions that I'm spending half my waking hours obsessing over. Philosophers/legal people define free will along the lines of, having had alternative options when you carry out some behavior, understanding that there are those options, understanding the implications of what you've chosen, having the ability to veto impulses to behave in particular ways (what usually gets called "free won't"). I come from way out in left field, in terms of how completely biological my perspective is. A behavior happens, something as simple as you decide to bend your finger. You can probably isolate the one motor neuron that initiated that action (or least a small handful of neurons). They activated because a neuron(s) upstream activated it, which did so because upstream from it...

So, let's take a neuron that has just triggered a behavior in this sense. Show me that it did so without any inputs from any upstream or surrounding neurons. Show me that it would have done the exact same thing at that moment, regardless of how much the person slept last night, their blood glucose levels, levels of various in the bloodstream around that time. Show me that it would have done the same even among subliminal cues and primes that psychologists show will alter behavior. Then, show me that it would have done the same regardless of the person's childhood, fetal life, genomic makeup. Show me that the neuron's behavior was not shaped by any of these influences, and you've demonstrated free will. This is obviously reductive to a silly extent, but you get the idea.

Philosophers and legal people spend a huge amount of time focusing on, "Did the person intend to take that action?" and in courtrooms that's one of the hallmarks of what would qualify for legal/criminal responsibility/expression of free will. And this ignores the 99% of determinative events that explain, "And where did that intent come from in the first place?" It's like judging a movie after only seeing its final three minutes.

> "Show me that the neuron's behavior was not shaped by any of these influences, and you've demonstrated free will."

Soon you will be publishing a new book, *Determined: The Science of Life Without Free Will* (Sapolsky, forthcoming), adding another discourse into the fray regarding free will (O'Connor *et al.*, 2018). To judge a book by its title in advance of reading—in the polysemous word "Determined" we can see allusion to both sides of the debate. From one view, if the outcome of a situation has been determined, then no agency can be exerted at all. In this view, a deterministic perspective, free will indeed does not exist. On the other view, for example the case of a person feeling *determined* to achieve a goal, determination seems like a canonical case where applied free will provides the motivation and grit to succeed. While surely the book itself will be addressing these questions as well—What motivated or influenced you to write this book now? What do you hope the book's impact will be?

I'm glad you perceived the (intended) ambiguity in the title. Why am I writing it now? I published this book, *Behave: The Biology of Humans at Our Best and Worst* in 2017 (Sapolsky, 2017), and did a ton of public talks about it. It basically goes through how our behavior is the outcome of events in the brain one second before, hormone levels that morning, experience in previous months, childhood, fetal life, genes, cultural evolution, all these things over which we have no control. And I'd be shocked by how many audience members, during questions, would show that the idea of there being no free will, or at least so little as to be irrelevant to interesting stuff, was novel. It seems like an obvious implicit conclusion from that book. So given those responses, I decided to make it a little more explicit.

What impact am I hoping the book will have? Obviously, I'm hoping it will solve global warming, suggest a cure for Alzheimer's Disease, save mountain gorillas from extinction, *etc.*, *etc.* Realistically, I'll be happy if at least a few people read it and find the arguments to be convincing.

What does free will feel like? Or, what would/could it feel like?

It's the most natural feeling in the world, totally reflecting this enormous human need for a sense of agency. It's something we desperately hold on to—loss of a sense of agency is at the core of the learned helplessness of depression, and of the hypervigilance of anxiety. It's so tough to take apart because we are working simply on the conscious level of attribution—it's so counterintuitive to think that an answer to the question, "Why did you do

that just now?", is one that shows that 99% of what is going on in us is subterranean. Who finds it natural to think that one of the explanations for why you just did something wonderful and altruistic is related to your oxytocin levels this morning, or how your cingulate cortex was constructed when you were a third trimester fetus? Our conscious sense of agency is usually post-hoc attributions to make sense of what you just did.

> Our conscious sense of agency is usually post-hoc attributions to make sense of what you just did.

Why do people—many of them quite honestly—feel that they have free will?

The answers are embedded in the one above—because, cognitively, it is so hard to understand the influences of all these things you can't see (back to hormone levels, fetal environment, *etc*). Emotionally, because it's depressing as hell to view ourselves as nothing more than the outcome of the biology, over which we had no control, and its interactions with the environment, over which we had no control.

The cognitive stances we take can be associated with real differences in personal outcomes. While the preceding sentence was mindful not to imply that such cognitive stances play a causal role in behavior, in various literatures it is common to read about the "impact" or "effect" that beliefs have on behavior and life outcomes (*e.g.* Conversano *et al.*, 2016; Sirgy, 2021). In your view, how is our personal stance on free will associated with our behaviors and context? Where are the causal arrows pointing—what leads to what?

Something that everyone assumes is that if people stopped believing in free will, everyone would run amok, there would be no constraint on behavior, madness in the streets. Moreover, some experimental studies have shown that when you lessen someone's belief in free will, they will cheat more in an economic game, will lie more, *etc*. Three responses to that:
a) those studies haven't been replicated;
b) when you look at people who come into the study ALREADY believing there is no free will, already having struggled with the implications of that, and so on, they are just as ethical and moral as those who do;
c) this notion of running amok if people stop believing in free will is a version of, "Why not, I'm not responsible for my actions."

A very large body of literature concerns a related conclusion of, "Why not run amok, after all I'm an atheist and don't think there is an ultimate Judgment." However, when you study it correctly, reflective atheists are exactly as moral as highly religious people.

In work such as the 1971 book *Beyond Freedom and Dignity* (Skinner, 1971) and *Walden Two* (Kulman, 2010; Skinner, 1973; 1948), B.F. Skinner popularized a behaviorist perspective, which in the human setting emphasizes the role of external reinforcement in social control. Skinner wrote: "Almost all major problems involve human behavior, and they cannot be solved by physical and biological technology alone. What is needed is a technology of human behavior." What do you think such behavioral technologies could look like today and tomorrow, and how can we design systems for human benefit?

That's going to be tough as hell, for the reasons outlined above. Skinner was pretty dogmatic in thinking that the important aspects are determined by the tendency (but not universality) for people's behavior being shaped by positive and negative reinforcements.

What generates the conscious flow of aware experiences that we have? In addressing this question, what role do you see for studies of altered states of consciousness?

I'm terrified by the subject of consciousness, on both a philosophical and biological level, in that I really can't understand what people are saying most of the time. Reflecting that, I don't think consciousness is actually very important to these issues—whether you just did something with conscious intent or otherwise, it is subject to the same reality—that behavior occurred because of neurobiological events a second ago, environmental triggers a minute ago, hormone levels that morning...all the way back to your genes and why they evolved that way. This allows me to completely sidestep my dimness about consciousness.

In a 2004 article "The Frontal Cortex and the Criminal Justice System" (Sapolsky, 2004), you memorably wrote "If free will lurks in those interstices, those crawl spaces are certainly shrinking" (Sapolsky, 2004). Here we might interpret these neural "interstices" as the physical spaces surrounding different cells in the brain, and more metaphorically as the gaps in how neural systems are currently studied. In the years since that article, there have continued to be

advances in functional, molecular, and systems neuroscience. How would you characterize the current state of exploration of those interstices? Over the past few years, where have we been searching where the light is, meaninglessly oversampling the same experiments or framings? And conversely, what methodologies or areas of the brain may still hold useful information?

Well, naturally, as a professorial scientist, I think that we now understand everything and nothing about the brain. All that new findings have been doing has been done to reveal more and more of the subterranean influences on our behavior—Wow, I had no idea that biology/genes/hormones/*etc.* have something to do with THAT behavior.

Where things are really problematic is that the scientific findings are on the group level, not the individual level – yes, we know with certainty that, for example, someone with their frontal cortex destroyed will behave in socially inappropriate ways—but good luck predicting which person with that damage will become a serial murderer, which will just say rude things to the host at a party. A lot of the legal scholars who weigh in with rejecting the idea that free will is a myth start at this point – if all this scientific insight can't tell me anything about what THIS individual is going to do, it has no place in the courtroom, and you sure haven't disproved the existence of free will.

My response is that we already know that, say, for every increase in the list of adversity that someone experienced in childhood (did they observe physical, sexual, or psychological abuse, did they experience it, was a family member incarcerated, was the person raised in poverty, was there substance at home… there are formal scales for quantifying this) there is about a 35% increase in the likelihood of them committing some serious (including criminal) anti-social behavior. Any time you already know enough science to say, "Someone with this background has a 93.5% chance of having been incarcerated by age 25," it doesn't matter that you can't predict events on an individual level with 100% accuracy—you don't need that to know that this is a screwed system in which people are held responsible for the uncontrollable biology that sculpted them.

In the same 2004 article (Sapolsky, 2004), you explored some of the ways in which neuroscience and philosophy intersect with justice systems, pointing to how different concepts of free will can implicitly and explicitly bear on legal proceedings. What promising and discouraging developments are you seeing at the intersection of neuroscience and law today?

Overall, I'm discouraged. I've developed a hobby of working with public defender offices to teach juries about the brain and how little control we have over who we are. I think I've worked on 11 cases, and lost ten of them. People are not open to these ideas, especially when looking at vivid pictures of corpses being shown to the jury…

Beyond the areas of law and justice: what does education look like in a world where the stance *Determined* is used as a handbook for the holistic design of learning systems?

Ugh, this is where I run into a wall. I am 100% convinced that we have no free will at all, and have thought that since I was an adolescent. But amid that, I have absolutely no idea how we are supposed to function with that insight, whether on an individual or societal level. I can truly think that way for maybe five seconds at a time. I can't visualize what things are supposed to look like if people actually understood that punishment, reward, blame, praise, any sense of someone deserving anything whatsoever, are all gibberish. I think the most important implication of that mindset, though, is that hating anyone for anything they've done is like hating an earthquake for the damage it caused, or hating a virus because it came up with a really clever spike protein and thus caused a pandemic. If we can achieve that mindset, that's some major progress.

> “I am 100% convinced that we have no free will at all (…). But amid that, I have absolutely no idea how we are supposed to function with that insight, whether on an individual or societal level.”

Struggles with mental health, sense-making, and cognitive deterioration are becoming mainstream topics. Some of these conversations can often be distilled down to different perspectives that people take on individual choice and

decision-making. What are the consequences of centering or excluding free will from these vital conversations related to individual cognition and behavior?

See above—the consequences of the two different views are monumental, and I sure can't figure out how best to run the world on a rejection of free will. The infinitely unsimple simple bit of advice is, keep that absence of free will in mind when you judge anyone about anything.

> "The infinitely unsimple simple bit of advice is, keep that absence of free will in mind when you judge anyone about anything."

After *Determined* – What are the immediate next steps, and deeper directions, for those studying free will as students or researchers? For example, are there some other topics to shift attention towards, or new philosophical tangles that can now be addressed in a better way?

This is going to keep me busy enough. Just to prepare myself, in surveys, more than 90% of philosophers firmly believe there is free will, essentially 100% of judges do, *etc.*, *etc.*, so this book is sure not going to transform much of anything.

The scientific landscape is changing for researchers in all positions, reflecting an overall turn towards collaborative approaches to education and research (Wray, 2002; Thomas & Zaytseva, 2016; Milojević *et al.*, 2018). Given your view into how career paths in science are evolving, what kind of skills or educational systems do you think will be useful for researchers in the coming years? What incentives or structures could be in place to improve career pathways for trainees in the era of Open Science (Allen & Mehler, 2019) or Decentralized Science (DeSci) (Friedman *et al.*, 2022; Hamburg, 2022; DeSci Foundation, 2021)?

My personal bias is that one of the most important things to teach, is that people realize the very limited use of reductionism in science. By reductionism, I mean basically the view that if you want to understand how something works, you can break it down into its component parts, understand how each part works, add them back together, and you will have an understanding. Instead, people need to appreciate how important

chaoticism and emergent complexity are—whether at the level of neural network function, or at the level of genetic crossing at fertilization—you see that reductionism can't explain the most interesting stuff.

“ People need to appreciate how important chaoticism and emergent complexity are—whether at the level of neural networl function, or at the level of genetic crossing at fertilization—you see that reductionism can't explain the most interesting stuff. ”

References

Sapolsky, R. M. (2005). Monkeyluv: And other essays on our lives as animals. Scribner. *Simon & Schuster*

Sapolsky, R. M. (2004). Why Zebras Don't Get Ulcers: The Acclaimed Guide to Stress, Stress-Related Diseases, and Coping (Third Edition). *Henry Holt and Company* https://play.google.com/store/books/details?id=EI880S_3fZEC

Sapolsky, R. M. (2017). Behave: The Biology of Humans at Our Best and Worst. *Penguin*. https://play.google.com/store/books/details?id=tPTGDgAAQBAJ

Huberman A. (2021). Dr. Robert Sapolsky: Science of Stress, Testosterone & Free Will | Huberman Lab Podcast #35. *Youtube* [Internet] https://www.youtube.com/watch?v=DtmwtjOoSYU

Mauss, S. Robert Sapolsky | Here We Are Podcast Ep. 376 | Hosted by Shane Mauss. *Youtube* [Internet] https://www.youtube.com/watch?v=ajnmb2b5Alo

Levitt, S. D. (2021) Robert Sapolsky: "I Don't Think We Have Any Free Will Whatsoever." *Freakonomics* https://freakonomics.com/podcast/robert-sapolsky-i-dont-think-we-have-any-free-will-whatsoever/

Sapolsky, R. M. (*in progress*) Determined: The Science of Life Without Free Will. https://www.goodreads.com/book/show/58902324-determined

O'Connor, T., Franklin, C. (2018). Free Will. Zalta EN, editor. *The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab,* Stanford University https://plato.stanford.edu/archives/sum2022/entries/freewill/

Conversano, C., Rotondo, A., Lensi, E., Della Vista, O., Arpone, F., Reda, M. A. (2010). Optimism and its impact on mental and physical well-being. *Clin*

Pract Epidemiol Ment Health; 6:25–29 https://doi.org/10.2174/1745017901006010025

Sirgy, M. J. (2021). Effects of Beliefs and Values on Wellbeing. *In: Sirgy MJ, editor.* The Psychology of Quality of Life: Wellbeing and Positive Mental Health. Cham: *Springer International Publishing*. pp. 245–262. https://doi.org/10.1007/978-3-030-71888-6_11

Skinner, B. F. (1971). Beyond Freedom and Dignity. *Hackett Publishing*.

Kuhlman, H. (2010). Living Walden Two: B. F. Skinner's Behaviorist Utopia and Experimental Communities. *University of Illinois Press*. https://play.google.com/store/books/details?id=cuj-x03b9lQC

Skinner, B. F. (1973). Walden (one) and Walden two. *The Thoreau Society Bulletin.*; 1–3. http://www.jstor.org/stable/23399198

Skinner, B. F. (1948). Walden Two. *Hackett Publishing*.

Sapolsky, R. M. (2004). The frontal cortex and the criminal justice system. *Philos Trans R Soc Lond B Biol Sci.;* 359: 1787–1796. https://doi.org/10.1098/rstb.2004.1547

Wray, K. B. (2002). The Epistemic Significance of Collaborative Research. *Philos Sci.*; 69: 150–168. https://doi.org/10.1086/338946

Thomas, J., Zaytseva, A. (2016). Mapping complexity/Human knowledge as a complex adaptive system. *Complexity*. 21: 207–234. https://doi.org/10.1002/cplx.21799

Milojević, S., Radicchi, F., Walsh, J. P. (2018). Changing demographics of scientific careers: The rise of the temporary workforce. *PNAS*. 115: 12616–12623. https://doi.org/10.1073/pnas.1800478115

Allen, C., Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biol*; 17: e3000246. https://doi.org/10.1371/journal.pbio.3000246

Friedman, D., Applegate-Swanson, S., Choudhury, A., Cordes, R. J., El Damaty, S., Guénin-Carlut A., *et al.*(2022). An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons. https://doi.org/10.5281/zenodo.6320575

Hamburg, S. A. (2022). Guide to DeSci, the Latest Web3 Movement. In: *Future* [Internet] https://future.a16z.com/what-is-decentralized-science-aka-desci/

DeSci Foundation. (2021). Why we need to fundamentally rethink scientific publishing. In: *Medium* [Internet]. https://desci.medium.com/why-we-need-to-fundamentally-rethink-scientific-publishing-43f2ae39af76