AUTONOMY, CONSCIOUSNESS, AND FREEDOM

DANIEL C. DENNETT

THE AMHERST LECTURE IN PHILOSOPHY

LECTURE 14, 2019

http://www.amherstlecture.org/

Autonomy, Consciousness, and Freedom

Daniel C. Dennett

PREFERRED CITATION

Dennett, Daniel C. "Autonomy, Consciousness, and Freedom." *The Amherst Lecture in Philosophy* 14 (2019): 1–22. <http://www.amherstlecture.org/dennett2019/>.

ABSTRACT

Free will is best understood as a particularly powerful – but vulnerable – ability to control oneself. This autonomy is independent of both determinism and indeterminism. Causation and control are not the same thing, as examples of remote control and ballistic trajectories reveal. Sam Harris has said. "A puppet is free as long as he loves his strings." He is right; loving your strings involves protecting them from those who would turn you into a puppet, and for this you need to preserve the privacy of your thinking. The appreciation of privacy has been inflated by many into a desire for absolute unpredictability, which is not required for the one variety of free will worth wanting.

The Amherst Lecture in Philosophy (ISSN: 1559-7199) is a free on-line journal, published by the Department of Philosophy, Amherst College, Amherst, MA 01002. Phone: (413) 542-5805. E-mail: alp@amherst.edu. Website: http://www.amherstlecture.org/.

Autonomy, Consciousness, and Freedom

Daniel C. Dennett

Tufts University

IS FREE WILL AN ILLUSION? A few years ago, Scott Adams' cartoon character Dilbert enunciated the issue vividly: "Free will is an illusion. Humans are nothing but moist robots. Just relax and let it happen."[1] Ever since this appeared, I've been using the phrase "moist robots." It's a wonderful term. Dilbert is right; we're moist robots. That's all we are, no magical mystery stuff in us. We're collections of biochemical micro-gadgets and that's all we are made of. But we're autonomous. That is my topic.

We're moist robots but should we "just relax and let it happen?" No, and that is the main point I want to make today. That's wrong. Yes, we're moist robots. No, we should not just relax and let it happen. The question of whether free will is an illusion I'll postpone until the end. It all depends, of course, on what you mean by free will. I don't want to argue about that now, so we're going to set that aside. Free will may or may not be an illusion, but *autonomy is not an illusion*.

Autonomy is real and it's important, and it can be defined quite clearly and strictly. Curiously enough, the concept that we need to make sense of autonomy is a concept from physics, but more particularly from engineering, and that is *degrees of freedom*, which has nothing to do with free will, you might think. How many degrees of freedom in a robotic arm (in an automobile assembly line, for instance? It depends on how many joints the arm has and how each joint can vary. Roughly speaking, the degrees of freedom of a thing line up pretty well

1 Scott Adams, *Dilbert*, March 18, 2012.

with how many *moving parts* it has, how many different *ways* can it be. An old-fashioned alarm clock has three degrees of freedom. First, you can reset the time it shows (and then leave that degree of freedom clamped until you notice that it is no longer telling the correct time). Second, you can reset the time the alarm should sound (and on an old-fashioned alarm clock, with a twelve hour face, there is no degree of freedom for setting AM or PM). Finally, of course, there's the ON/OFF degree of freedom. Today alarm clocks have many more degrees of freedom: dozens of possible sounds to choose from, or music from any of hundreds of radio or internet sources, snooze buttons, automatic resetting of time to international standards, daylight savings time . . .

Each degree of freedom is an opportunity to control. How many degrees of freedom in your arm for you to control? Lots. More than in most industrial robots, but not as many as in an octopus tentacle. There is a choice for each degree of freedom:

> you can "clamp" or "burn" it by setting it and not letting it change until you set it again;
>
> you can let it swing free, in effect, out of control.
>
> Or you can actively control it, monitoring its state and the state of the world and making adjustments in real time.

A familiar diagram shows the degrees of freedom in a flying vehicle – a plane, say, or a drone (fig. 1). You've got right, left, up, down, yaw, pitch, and roll. If all you're talking about is motion through three-dimensional space, those are the degrees of freedom that are up for control. Each degree of freedom is an opportunity for – or a need for – control. And so, obviously, the *remote control* device for a drone must have a button or knob or joystick for each degree of freedom that is under the control of the drone-operator. If you've got more degrees of freedom than you have control buttons for, then those are going to be degrees of freedom that you're not going to control. They're going to be out of control.

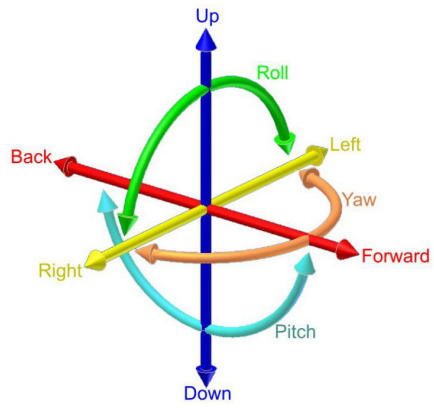

FIGURE 1
© Emissarydrones.com

In *Consciousness Explained*, I went to great lengths to try to explode the idea that inside your head there is what I called the *Cartesian Theater*.[2] That's supposedly where *you*, the inner witness, sits looking at a screen and pushing buttons to make your body move. This idea of the inner homunculus, the inner witness, sitting there controlling the body by taking in incoming information from screens and speakers and so forth is a very powerful and seductive idea about what consciousness is like – and it's all wrong. There is no Cartesian Theatre in your head! But the remote control panel for a drone is a perfect stand-in, an externalized control center complete with buttons and knobs to push *and a screen to look at* and *audio to listen to*. All the control work done by a remote controller, a land-based drone pilot, depends on the feedback loops that give the controller current information about the effects of all the joystick pushes and button presses. And that is the *work* that your brain must do for you to maintain control of your body, but it doesn't do it by looking at an inner screen or listening to inner sounds; you do it by looking at the world and listening to, and touching, and smelling the things your sense organs permit you to sense.

Keith Frankish and I are engaged in creating an extended thought experiment to help us understand both autonomy and consciousness by starting with a remote controlled drone and then emancipating it, step by step. If we can put all the controls executed by a human drone pilot on board the drone itself, so that it is no longer under the control of any agent but itself, then we think we will answer some important questions about consciousness. That will be a topic for another occasion, since we are trying to do it right, with lots of attention to the details, and that means supplementing our philosophical thinking tools with a lot of empirical and theoretical work in cognitive neuroscience, robotics, and artificial intelligence. I mention it now because I am going to slide over all the questions one might well ask about those details – many of which are still not just unanswered but unasked by us – so I can focus on the issue of autonomy and the particular problems for self-control raised by human consciousness.

When I was working with Rod Brooks at MIT's Artificial Intelligence Lab on Cog, the humanoid robot project, I got a vivid introduction to degrees of freedom and control. Cog's arms were very loosely-goosey for a robot, and its hands and fingers had many degrees of free-

2 Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown, and Company, 1991).

dom. It was quite wonderfully spooky. One day I took one of my Tufts teaching assistants over to the Cog lab to see what we were doing. Matt Williamson, the graduate student who had created the "series-elastic actuators" that controlled Cog's arms, had taken one of Cog's arms off its shoulder and C-clamped it to a workbench where he could more comfortably work on it. It was powered on and under control, and Matt invited my TA to shake hands with it. So she reached over and shook hands with this bodyless robot arm – and screamed. "It's alive!" It had responded so naturally, so dynamically to the pressure of her hand that she couldn't squelch the automatic inference drawn by her own perceptual systems. Cog's humanlike hands and arms, with many degrees of freedom to control, were a very interesting innovation and also an engineering problem. Rod wanted Cog to "appreciate" that, if it wanted to thread a needle, for instance, it should simplify the task by burning some degrees of freedom by bracing one arm against a table, turning that hand into a fixed vise, in effect. If you clamp several degrees of freedom, you don't have to think about them. You can concentrate entirely on controlling the degrees of freedom of your other hand. This idea of clamping degrees of freedom is a useful idea. Sometimes you've got more degrees of freedom than you want and it's a good idea to clamp some of them so you don't have to think about them.

Let's take our drone and clamp all the degrees of freedom except one: up/down. What have we got? A virtual flagpole. Tie the flag to it, hit the up button, raise the flag. Let it stay up there all day, hit the down button at sunset, bring it down. Not a very economical solution to a simple problem but when you clamp all those degrees of freedom you still have something worth thinking about. The example reveals that even a recreational drone has already installed on board some very sophisticated control machinery. You, the drone pilot, don't have to worry much about stability. When you stop pushing the *up* button the drone will just stop and hover where it is, indefinitely. You don't have to control the speed of the individual rotors, for instance, but leave that to the onboard stabilizer system, which can do it much better than you can. (In fact, without all that onboard control machinery, recreational drones would be too hard to control to be much fun.) Our nervous systems make the same bargain: you don't have to think about controlling the individual muscles in your limbs in order to walk; that's all packed off into highly reliable and versatile control systems that you needn't waste any *conscious* cognition time on – and couldn't if you wanted to, without special training. It's handled unconsciously.

What, then, is *autonomy*? Autonomy is self-control as contrasted with remote control and with being out of control. I take it that the difference is obvious, real – not an illusion – and fundamental. The first point I want to make is that *control* and *causation* are not the same thing. I think a lot of confusion in the free will literature is due to the fact that people don't see the difference. So I'm going to try to make it more obvious. Let's think about autumn leaves falling from the trees. Are they *controlled* by anything? Aren't they being controlled by the wind? And by gravity? No, they're being *caused* by the wind and by gravity to take the trajectories they take, but they're not being *controlled* by the wind or by the gravity because gravity is not a controller nor is the wind. Gravity doesn't care, has no plans, has no goals, is not *trying* to make the leaves go where they go. They are just going wherever they're going. They are caused to go where they go. But they are not controlled. They are *out of control.*

Now let's think about a slightly different case: grains of sand being sorted by the waves on the beach. They may make some beautiful patterns, wonderful ripples and this may begin to look like control. It's regular causation caused by the periodicity of the waves and the size of the grains and so forth, but the individual grains of sand are not controlled. They end up going where they go in what is basically a chaotic – or you might even say random – way, under the purposeless influences of the many forces that impinge on them. A Galton board which shows how ball bearings falling from a spout will form a normal distribution is a well known simple demonstration of how such a pattern can be generated by an uncontrolled process.) Suppose you went down to the beach and when you looked at the ripples in the sand what you saw spelled out was "Jesus is coming. Look busy." You would know that there was a controller that had done that. No chance confluence of waves is going to make that pattern. That's the result of a controlled operation. In other words, control requires a controller, that is to say: an agent. And a controller requires feedback information in position to be used by the controller to modulate control. Without feedback, you don't have control.

Let's think about firing a rifle bullet. You're controlling the direction of the gun barrel and, with the trigger, the time of the bullet emerging from the gun barrel. Are you controlling the course of the bullet after that? No. Where it goes after it leaves the muzzle is out of your control. Now if you're a really good shot, you may be able to calculate in advance the windage and so forth and you may be able to get it in the bull's eye almost every time. But you are unable to affect the trajectory of the bullet after it leaves the gun. So it is not a controlled

trajectory, it is a *ballistic* trajectory. It goes where it goes and, if your eyes were good enough to watch it and see that it was going "off course," you'd have feedback, but you wouldn't be able to do anything about it. Feedback is only useful information coming back to a controller if the controller also maintains an informational link back to the thing that's being controlled. You fired the gun. You caused that bullet to go where it went, but you did not control the bullet after it left the gun. Compare that with a *guided* missile. A guided missile, after it's launched, can still be controlled, to some extent, often to a great extent (as in a cruise missile). As you know, one of the chief inventions of technology in warfare in the last 50 years is the development of remote control missiles and, of course, remote control drones. *Remote control is real,* and readily distinguished from *out of control.* Autonomy is *non*remote control, *local* or *internal* control, and it is just as real, and even more important.

Now let's think about randomness for a minute. Think about flipping a coin. It's like shooting the bullet. After the coin has left your thumb, you can't control it. It's going to go where it's going to go and you can't make any further adjustments even if you can see that it's not going where you wanted it to go. It's no longer under your control. I want to make this particularly clear by imagining that you've decided with somebody to settle some huge conflict – who gets the million dollars, for instance – and you've decided to settle it with a coin flip at noon. You show up at the appointed place at noon and the other person shows up with a giant machine inside a heavy glass case, floating in a bath of mercury, with a control panel that measures humidity, barometric pressure and temperature inside the box. The machine has micrometer adjusting screws and dozens of dials. He says: "This is my coin flipper." Do not let him use it. Why not? Because he's trying to control the trajectory of the coin, and the whole point of a coin flip is to prevent either party from controlling it. Actually, he'd be very unlikely, even with the imagined high tech that I gave him, to be able to control whether the coin came down heads or tails. Coin flips are remarkably uncontrollable. Physicists have calculated that if you really want to predict coin flips, you have to know the location of every electron in the visible universe and its gravitational effect on the coin. Fair coins and dice are not exempt from causation but they are exempt from control. They are designed to be uncontrollable. And note that this has nothing whatever to do with whether physics is deterministic or indeterministic. Things can be *out of control* while being determined, and

things can be *controlled* while being determined. If you want to discover whether something that happened was controlled, you don't need to know whether determinism is true or false.

Coin flips are designed to be uncontrollable. Are there other things that are designed to be uncontrollable? Yes. We are. It's not that we're designed to be chaotic; it's that we're designed to be uncontrollable by others who want to control us. Way back in 1978, the Viking II orbiters of Mars were emancipated by the scientists at the Jet Propulsion Laborary at Caltech. They were made autonomous.[3] The spacecraft had been remotely controlled since their launch in 1975 by signals from a station on Earth all the way to Mars. But when the mission was extended beyond its scheduled completion date three years later, the engineers needed to cede control of some important factors to the spacecraft themselves, which were running low on gas to adjust their orientation, and had some other control problems that they had not been designed to handle on their own. The time lag (between 6 and 42 minutes round trip at the speed of light) between Mars and Caltech made remote control impossible, but looking at the wiring diagrams, the engineers found an unused connection between the two main computers on board that could be programmed remotely to send feedback and control signals internally! These autonomous Vikings were made capable of fending for themselves. They were able to protect their supplies of electricity in their solar-powered batteries by locally turning off instruments when they weren't needed. They could also fix gas leaks in their propulsion systems – which took split-second timing, and even detect and discount deceptive "visual" data (light reflecting off small specks nearby that made them appear to be stars) that would have thrown their navigation systems off.

You may have seen or heard the phrase "fast, cheap and out of control." a slogan of Rod Brooks. He was the director of the Cog project, as I have mentioned, but he had another project: designing (unmanned) vehicles for exploring planets. NASA doled out big grants to engineers to design such robots, and Rod's idea was: just as you don't have to send fragile people to the planet at great risk and expense, you don't have to send an artificial person, a big, fancy, expensive, robot. Instead, send a whole bunch of insect-like robots, fast, cheap and out of control. You drop them on the planet and let them run around more or less helter-skelter and the odds are pretty good that one of them at least will get you the data you want

3 Edward Hutchings Jr. "The Autonomous Viking," *Science* 219, no. 4586 (February 18, 1983) 803–8.

and radio it back to you. You can't control them from Earth and you don't care. You put some modest self-control capacities in them and then you just let them do their thing and send their data back when they are through. (Some of you may have seen the 1997 Errol Morris documentary film *Fast, Cheap & Out of Control*, which featured Rod Brooks along with several other inventive characters who had their own ways of being fascinating. I recommend it.)

Here's what I think I have shown so far. Whatever *free will* may be, if anything, autonomy in the modest, unphilosophical sense of self-control is a real phenomenon. We can distinguish physical systems or entities that are autonomous from those that are remotely controlled and from those that are out of control. This isn't metaphysics; it's engineering. These are reliable distinctions and at least this kind of autonomy has nothing to do with determinism. Trivially, in a deterministic world, everything is determined. Some things are determined to be remotely controlled, some things are determined to be uncontrolled, some things are determined to be autonomous or self-controlled – until they aren't! Determinism doesn't say that something that is self-controlled now will always be self-controlled. It doesn't say that something that is not controlled now will never be controlled. Determinism is simply mute on that topic. Determinism is remarkably neutral with regard to whether something is in control, can be in control, will come under the control of something else, is going to go out of control. Determinism does not say that something out of control can't be controlled.

In fact, contrary to what many of my students say, when I ask them what bothers them about determinism, it doesn't say that you can never improve your circumstances, or that you are stuck with your foibles and shortcomings, or that you "can't make a difference" in the world. Maybe you *can* teach an old dog new tricks. If you can, then some dogs, at least, are determined to be capable of learning new tricks when they're old. So if one of the things that bothers you about determinism is that it *ties your hands*, you're just wrong. Maybe your hands are tied now. Probably not. (I don't see anybody with tied hands in the audience.) Maybe your hands are going to be tied for a long time, maybe you're going to untie them tomorrow, maybe your hands are never going to be tied. Determinism doesn't say one way or another whether your hands are tired, will be tied, can't be tied. It simply doesn't have anything to say on that topic – or on any of the topics for which "tied hands" is a metaphorical expression. If you're lost in the desert with no water and nobody to rescue you, you'll soon die. That grim

fact follow from the particular circumstances you're in, not from determinism in general. If you're lost in the desert with no water and nobody to rescue you, try dialing 911 on your cell phone. It might work. Be prepared. Determinism doesn't tie your hands.

Neither, by the way, does indeterminism. People sometimes say, "You can't change the past but you can change the future." No, you can't. From what to what? You can duck an incoming brick that was *going to* hit you, but then it's the *anticipated* future that you kept from being the *actual* future, thanks to your perceptual capacities and quick response. You can no more change the actual future than you can change the actual past. If indeterminism is true this gives you no more chance to win the lottery, or a coin toss, than if it isn't true.

Self-control is possible whether or not determinism is true. Where does consciousness come in? If we're going to emancipate our drone, what's the first thing we should do? Throw away the LED screen. If the drone is going to control itself, it's not going to have internal eyes with color vision and a vision system to look at the screen the way a human remote controller does. All the information that is rendered so vividly and conveniently for human drone controllers is already contained in the drone, and it's already in a format that's directly usable by something like a drone. There is no need for a Cartesian Theater in the drone, but still, there *could* be one. Consider the scene in the film *Men in Black* where Will Smith and Linda Fiorentino visit the morgue and see the "corpse" of the huge bald man lying on the shelf. Smith notices a little button or wire on the corpse's ear and pushes it. The whole face hinges open, and inside there's a little green man sitting in the control room. It turns out that this strange giant is actually just a puppet with an internal puppeteer. That's the Cartesian Theater. What the film shows is that the idea of a Cartesian Theater is not incoherent. There could be a Cartesian Theater. If we wanted to send people to the planet of the giant people, maybe the way to do it would be to send giant puppets (with real people in their control rooms) to pass as local inhabitants. Slow, expensive, and in control, but *possible in principle* as philosophers love to say! There is nothing incoherent with the idea; it's just false. That is, when we open up somebody's face and look inside (noninvasively, with fMRI scans, etc) we find there's nothing like that going on in the brain. That's an empirical point. There is also a *conceptual* point, which is, if there were a Cartesian Theater in our heads, we'd have to keep going, opening up the little controller's head, and so on. The conceptual point is that at some point you have to get rid of the Cartesian Theater, and get the control system

accomplished by things that are not themselves conscious agents. In fact, lucky for science, we dismantle the Cartesian Theater at the outset, and distribute its work load among lesser agencies in the brain. There could have been a Cartesian Theater but there isn't.

So now, if we're going to make our drone autonomous, how autonomous should we make it? This is a serious practical question today. It's one that I am spending a lot of time discussing with roboticists and people in AI around the world. How autonomous do you want your systems to be? Once again, Dilbert can help us out. A self-driving car says to Dilbert, "I find it offensive when you call me a self-driving car. That's my slave name, I prefer to go by the name Carl." Dilbert responds, "Shut up and drive me to work." And the car responds as it backs away: "Said the self-walking human."[4] This is an autonomous vehicle that's just a tad too autonomous. We don't want to make our autonomous vehicles that autonomous. And the reason is simple. Autonomy is dangerous.

Lucky us. I think everybody in this room is autonomous and, hence, dangerous. We're all dangerous. How can we bear to let such dangerous things remain free to roam the world? We don't want to have drones and cars out there that are that autonomous. How can we dare to let our children go out in the world as autonomous agents? And the answer is that we devote a lot of time and energy to prepare our children for this freedom. Notice that when we *launch* them, they are no longer in our control, even if we have a lot of feedback. We may not be able to use that feedback to modulate our responses well enough to guide their behavior in the ways we would like. When they are young, they have uncountable degrees of freedom, but are out of control in many regards. We try to control them when necessary – herding cats, as one says – and try to teach them how to control themselves. If all works out, they eventually become reliably safe and reliable autonomous agents, ready to confront the world of opportunities with some chance of surviving intact. When do they acquire free will? Wait. I'm getting there.

We want to make our children autonomous in ways that are socially benign and amenable to the sorts of controls we approve of, not ON/OFF switches, but persuasion and education and appeals for common cause and the like. This brings me to a disagreement with a friend of mine, which has had considerable visibility on the Internet. Sam Harris published a book,

4 Scott Adams, *Dilbert*, January 24, 2019.

Free Will, and when it came out, he sent me a copy, with a jacket illustration showing the letters that make up the title hanging individually from marionette strings.[5] I told Sam I really didn't like the cover with those marionette strings, and also thought he'd made some serious mistakes in the book. And he reminded me that he'd asked me to read the penultimate draft of the book, and I had said I was too busy. I decided that I was indeed morally responsible to some degree for the sin of letting my friend commit in published form the errors I claimed he had made. So I owed him a belated critique, at the very least. I wrote my critique of his book, and it was fairly harsh, but he bravely and honestly put it on his website, where it resides to this day, along with his response, which is called "The Marionette's Lament: Response to Daniel Dennett."[6]

The title of his response reminded me of a wonderful line in his book. He says that compatibilism—a version of which I've defended for years—"amounts to nothing more than the assertion of the following creed: '*A puppet is free as long as he loves his strings*.'"[7] This sentence, it occurred to me, really does hit the nail on the head.

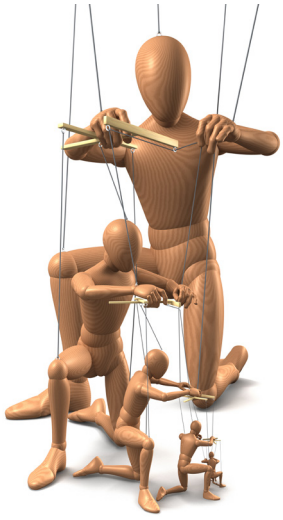

FIGURE 2
© Bizipix | Dreamstime.com

"A puppet is free as long as he loves his strings." Let's see if it's true. Here's a marionette, controlling a marionette, controlling a marionette, controlling a marionette, controlling a marionette (fig. 2). It's easy enough to draw and to imagine, but it's physically impossible. There's no way the degrees of freedom of the smallest marionette could be controlled via strings to the hands and so forth. It's not just that you couldn't do it with strings; you couldn't do it electronically either. The problem is just too hard, with too many

5 Sam Harris, *Free Will* (New York: Free Press, 2012).

6 Daniel C. Dennett, "Reflections on Free Will" (January 26, 2014), https://samharris.org/reflections-on-free-will/; Sam Harris, "The Marionette's Lament: A Response to Daniel Dennett" (February 12, 2014), https://samharris.org/the-marionettes-lament/. See also Harris's YouTube discussion of my compatibilism: "Sam Harris on Daniel Dennett's Compatibilism - Free Will" (March 29, 2013), https://www.youtube.com/watch?v=TXWDkwSyjpU.

7 Harris, *Free Will*, 20.

degrees of freedom, too little acuity. The precision you would need to get that smallest puppet to do anything other than just flop around is simply not available. If you want to have a marionette that can seem to be playing the flute, as in the famous Salzburg marionette theater, you better have just one marionette and a virtuoso marionetteer pulling the strings. So, this cascade of control is impossible—not metaphysically impossible but physically impossible. The laws of physics just do not permit the precision to make this possible.

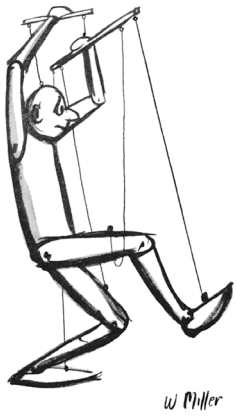

FIGURE 3
© Warren Miller | *New Yorker*, January 26, 1963

Here's another puppet, pulling its own strings (fig. 3). Is it possible? This is a 1963 New Yorker cartoon and while it's not the best way to control your life, it is possible. There could be an auto-puppet, a marionette that controls its own legs by moving its own arms. And how does it control its arms? With some internal "strings' or other control signals. (Some paraplegics actually enjoy arrangements of this sort.) The main thing to note about the auto-puppet is the smile on its face. It's not being controlled by anybody else; it's autonomous, and it loves its strings. You should learn to love your strings. Don't let anybody turn you into a puppet.

That is my main message today. In many philosophical debates about free will, people say things like "If determinism is true, we're all just puppets." (That seems to be the message of Sam Harris's book jacket.) And I am saying no, we're not just puppets but we could become puppets if we don't act carefully to preserve our autonomy. Our autonomy is on the line. In fact, it's becoming more on the line every day. People are inventing effective puppet strings right now, and getting people to attach those strings to themselves. The most effective puppet string yet invented is the smart phone. James Williams, a veteran and refugee from the world of app design and video game design, has published a small book, *Stand Out of Our Light*, with the subtitle "Freedom and Resistance in the Attention Economy."[8] What shocks him is how there is now a multi-billion-dollar competition among various giant com-

8 James Williams, *Stand Out of Our Light: Freedom and Resistance in the Attention Economy* (Cambridge: Cambridge University Press, 2018).

panies to pull your strings, to control your attention. Forget about controlling your legs or your hands. If they can control your attention, they can control you. Suppose you are doing something really important and you need to look up something on your smart phone. If you get distracted by a YouTube link or advertisement on the screen, your string has just been pulled. "Oh, that looks interesting," you remark to yourself, and off you go, abandoning, if only temporarily, your important project. Even if you don't bite, the people who would control you are gathering all the feedback they can, trying to learn all about you, so that they can design a better distractor to dangle in front of you tomorrow. (I tell my grandchildren about *anglerfish* that lie in wait, dangling a little wiggly worm-lure in front of their mouths, until *snap*! – their prey gets too close and becomes lunch. There are thousands of different species of anglerfish out there, I tell my grandchildren, and they must learn to be self-conscious about approaching anything that looks tempting.)

I think that this is perhaps the greatest risk to human political freedom that we've ever seen. The capacity of individuals and companies to distract you and to channel, to clamp your degrees of freedom so that you just don't think about things that you really should be thinking about because you're so distracted by all these other things which you can't help looking at, thinking about, instead. The competition for your attention strikes at the heart of your freedom, your ability to think for yourself.

"What do I do now?" That's the question that every living thing faces every moment of its life. We're the only ones that can actually say the question out loud or even to ourselves. But every living organism is a self-controller that has to have an answer to the question: what do I do now. There is a big difference in the range of answers a self-controller may have to decide among. "Should I tumble or not, should I feed or flee?" might sum up the basic options of a bacterium, while hundreds of degrees of freedom are clearly available to mammals and birds. Compare the activities of the most versatile robots or drones to the animacy of warm-blooded creatures. All robots (even Cog, with its remarkable arms) appear quite phlegmatic and stolid – they are, after all, *inanimate*. No robot is "free as a bird" because roboticists economize on degrees of freedom for which they would then have to devise controls. But for all their exploratory liveliness and even playfulness, the degrees of freedom of a bird are dwarfed in turn by the degrees of freedom of a normal human being. This comes out clearly in a definition of intelligence by Jean Piaget, a bit revised and improved by Guy Claxton.

Intelligence is knowing what to do, when you don't know what to do.[9]

Think about it. Many animals know what to do because they have a limited variety of things they know *how* to do – either innate or learned talents – and they've got these well optimized, and prioritized, a repertoire of likely appropriate actions to perform at various times. But if you give them a situation for which their pasts – and the pasts of their ancestors--have not prepared them, they're clueless, they're hopeless. Real human intelligence, in contrast, is knowing what to do when you don't know what to do. What to do of course is to *think*, to become, for a time, a problem solver. You drop everything and put on your thinking cap and see if you can figure out a way out of the problem you've encountered.

There are a few impressive examples of such apparently novel problem solving in other species, but they are mostly controversial. Wolfgang Köhler's box-stacking chimpanzees have been duly famous for over a century, but how much practice or familiarization with the available materials did they have?[10] Betty the New Caledonian Crow bends a piece of wire to make a hook to grab an otherwise unreachable basket with food in it.[11] Perhaps the best examples are Bernd Heinrich's ravens.[12] Heinrich raises the ravens from the egg to make sure they have never had any experience with the problems or the solutions before giving them a test: recover the food dangling from the string. Ravens solve it; crows do not.

We're the only species in which most members can regularly solve simple novel problems: How do I get the spoon that fell behind the radiator? How do I get a pill out of this newfangled plastic pill-package without a pair of scissors? How can I get my neighbor to re-

---

9 Guy Claxton, "Intelligence from the Ground Up" (lecture, Understanding Intelligence Ethics in the Age of AI conference, New College of the Humanities, London, March 22, 2018).

10 Wolfgang Köhler, *The Mentality of Apes*, trans. Ella Winter, 2nd ed. (London: Routledge and Kegan Paul, 1927). For related video, see "Kohler Chimpanzees," YouTube, May 21, 2013, https://www.youtube.com/watch?v=FwDhYUlbxiQ.

11 See Bob Yirka, "Betty the Wire-Bending Crow May Have Been Less Insightful Than Previously Believed," Phys.org, https://phys.org/news/2016-08-betty-wire-bending-crow-insightful-previously.html, which has both the video and a discussion of the grounds for skepticism.

12 Bernd Heinrich, *Mind of the Raven: Investigations and Adventures with Wolf-Birds* (New York: HarperCollins, 2000). For related video, see "Ravens: Testing Intelligence," *Nature* (Educational Broadcasting Corporation, 2001), https://www.pbslearningmedia.org/resource/nato8.living.reg.resou.testintel/nature-ravens-testing-intelligence/.

trieve his pet python from my mailbox? And the MacGyvers among us delight us with their more ingenious innovations. We are virtuoso problem-solvers, but we pay a big price for our talent. Our abiding question concerns not just what I *do* now, but also what I *think about* now. How many degrees of freedom do you have? Thousands? No, you have millions or billions of degrees of freedom because you have human consciousness and you have language.

FIGURE 4
© Saul Steinberg, *New Yorker*, October 18, 1969

Consider the Saul Steinberg *New Yorker* cover (fig. 4) that I used as the jacket illustration on my book, *Sweet Dreams*, my favorite picture of consciousness.[13] The man in the museum is looking at a painting by Braque and he's guessed that it's Braque and that reminds him of baroque and then barrack and then bark and poodle and Suzanne and . . . This is his stream of consciousness. Steinberg's thought balloon is a wonderful metaphor for what's going on in his head, but it's only a metaphor. I would say that the problem of consciousness can be summed up succinctly: if that's what's *metaphorically* going on in his head, what's *literally* going on in his head that makes that a good metaphor? What are the details in neuroscientific terms that make this an accurate metaphor for what he's thinking about, in all its details? But let's set that question aside, challenging as it is, because on this occasion I want to use the picture to illustrate a different point.

Suppose this fellow walks into the museum, and looks at the painting; Steinberg has represented his stream of consciousness over, say, a minute. Now let's suppose the fellow walks out of that room and walks back in five seconds later, to look at the painting again. Is his stream of consciousness going to be the same? No, it's going to be all different (aside from a few recollected bits, let's grant, from his first musing). And if he walks out again and comes

13 Daniel C. Dennett, *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness* (Cambridge: MIT Press, 2005).

back, its going to be all different again. We never have exactly the same stream of consciousness twice. Why? Because we have so many degrees of freedom; each thought can remind us of another topic, which stirs up another reaction and so forth. We're drowning in degrees of freedom which are, remember, *opportunities* for control. But you simply cannot control all the degrees of freedom you've got. You'd go crazy trying. Sam Harris puts it this way: "Thoughts and intentions emerge from background causes of which we are unaware and over which we exert no conscious control."[14] That is an overstatement, for as he himself goes on to point out,

> A creative change of inputs to the system – learning new skills, forming new relationships, adopting new habits of attention, may radically transform one's life.… This understanding reveals you to be a biochemical puppet, of course, but it also allows you to grab hold of one of your strings."[15]

We're autopuppets after all, apparently, but not entirely. At any time we can pull only some of our strings. But that's not none. What we all have to learn when we grow up is that we are blessed – or cursed – with more degrees of freedom than we can handle, and what we have to learn is not the science or algorithm of self-control but the *art* of self-control. We need to learn – and almost all of us succeed in learning – how to marshal our attention and clamp degrees of freedom that we're not going to need for a while so that we can concentrate on the things that really matter. This is not something you are born with; it's a bag of tricks you have to acquire or invent. This is something that you can get better at. It is the key to maintaining your freedom in this contemporary twenty-first-century world. It is the only kind of free will worth wanting.

I said earlier that we are designed to be uncontrollable-by-others. All organisms have this as a design ideal, but often evolution invents ways for parasites to control hosts, and predators to control prey in ways that benefit the alien controller, not the organism itself. This wasn't the deliberate design of any intelligent creator, it was the design of evolution. After language evolved in one species, *Homo sapiens*, this made us systematically hard to

---

14 Harris, *Free Will*, 5.

15 Harris, *Free Will*, 46–47.

control from the outside. Nobody can yet get enough feedback about the current state of your mind to control you, and if you learn of the attempts, you are well equipped to take counter-measures. Faced with that prospect, the best that anybody can do if they want to control you is to try to clamp all but a few degrees of freedom so they can actually entrain aspects of your stream of consciousness and keep you thinking about the topics they want you to think about. There are times when this can be done openly and honestly and with the full cooperation of the target, for instance in controlled psychological experiments. The subject is instructed to perform a task (and to concentrate on it) but kept in the dark about the point of the study – and is typically paid or otherwise rewarded for cooperation. If the experiment is well-designed, this benign and voluntary clamping of degrees of freedom is accomplished, and nice clean data emerge that can be evaluated agains the predictions made (secretly) by the experimenter. But such well-designed attention-clamping is also part of the ecology of the mind of the twenty-first century, and much of it is not benign at all.

What is the important difference between being distracted from the thinking task you have set yourself by a beautiful sunset and being distracted from the thinking task you have set yourself by a YouTube video of a beautiful sunset? The latter is caused by a *controller,* a would-be puppeteer; the former just happens to happen. Ricky Skaggs, the folk singer, says, "I can't control the wind but I can adjust the sails." That's as good as a free will motto, but notice: if Eolus the Greek god of the winds exists then Skaggs has a problem, because Eolus is an agent and Eolus will adjust the winds while Scaggs is adjusting the sails. Skaggs has a problem and he better keep his itinerary secret from Eolus if he can because there's an arms race between Scaggs as a sailboat-controller and Eolus as a wind-controller.

This brings me to a famous experiment much discussed by philosophers and others writing about free will. This is the experiment by Soon et al., an improvement on Benjamin Libet's even more widely discussed work.[16] Subjects are put in an fMRI machine where they are shown a series of computer screens with different letters on them, each screen lasting 500msec – half a second. Subjects are asked to decide (on a whim, without any planning or purpose) to push either a left finger button or a right finger button. And subjects are also

16 Chun Siong Soon et al., "Unconscious Determinants of Free Decisions in the Human Brain," *Nature Neuroscience* 11, no. 5 (May 2008): 543–45.

instructed to remember what letter they were looking at on the screen at the moment they decided which gratuitous act to perform, left or right. Then subjects are shown a screen with all the letters they have seen and asked to push a button indicating which one coincided with the "onset of intention." As Soon et al. put it:

> At some point, when they felt the urge to do so, they were to freely decide between one of two buttons, operated by the left and right index fingers, and press it immediately. In parallel, they should remember the letter presented when their motor decision was consciously made.[17]

By the way if you actually are a subject in that experiment you'll find it an extremely annoying experiment. It's a very unnatural act. But in any case it turns out that the fMRI data gathered up to ten seconds before the choice permits a computer to predict whether the left or right button *will be* pushed. The standard interpretation of this result is that Soon and his colleagues can predict some "free" decisions people make ten seconds in advance of their conscious experience of deciding. Or, as some have said, your brain chooses long before your mind thinks it chooses. Is this evidence that we don't have free will? No, but it does point to a possible future threat to our free will. The result is very interesting, to be sure, but it isn't surprising, is it, that when given a task that requires a "random" spur-of-the-moment choice, people let some unconscious process churning in their heads tip the scales somehow. And one has to be careful about the word "predict" here. The data analysis (by a trained-up deep-learning algorithm) is very compute-intensive and actually takes quite a long time. That's fine for *scientific* prediction; the subjects' choices on each trial can be recorded and the records sealed until a week later, if need be, when the deep-learning algorithm issues its results. It they match, then a scientific prediction has been successfully made. But it is not fine for real-time prediction. Some day the process of gleaning this information from the data may occur fast enough to permit real-time prediction (which is required for control), and that will change everything.

So, what's the moral of the story? Only this: In the future, don't play *rock-paper-scissors* for money with Soon et al. with your head in an fMRI machine! That's the only moral. Take

17 Soon et al., 543.

heed. You don't want others reading your mind. Your task, as a responsible adult, is to preserve your privacy. This good idea – *preserve your privacy* – has been inflated by philosophers over the centuries, and turned into the requirement of absolute, "in principle" unpredictability. That's supposed to be what is required for free will. For instance, my dear departed friend Jerry Fodor, once said:

> One wants to be what tradition has it that Eve was when she bit the apple. Perfectly free to do otherwise. So perfectly free, in fact, that even God couldn't tell which way she'd jump.[18]

In other words, "one wants" a miracle: one wants to be perfectly unpredictable. Why? Galen Strawson chimes in (writing in another review of my book): "He doesn't establish the kind of absolute free will and moral responsibility that most people want to believe in and do believe in. That can't be done, and he knows it."[19] Strawson is right about one thing; it can't be done and I know it. But why would "most people" want to believe in *absolute* free will? I can't think of a good reason, but I hypothesize that what people *really* want – and they are right – is as much invulnerability to remote control as possible, and they recognize that having a poker face, keeping one's stream of consciousness private, is the first order of business. Philosophers have then made a typical philosophical move: they have inflated practical inscrutability into "in principle" inscrutability, and thereupon declared that nobody has free will worth wanting unless they can play *rock-paper-scissors* with God and break even. Philosophers have created a whirlwind of anxiety by amplifying a perfectly good requirement into an impossibility. Well, it keeps their classrooms full.

---

18 Jerry Fodor, "Why Would Mother Nature Bother?" review of *Freedom Evolves*, by Daniel C. Dennett, *London Review of Books*, March 6, 2003), https://www.lrb.co.uk/the-paper/v25/n05/jerry-fodor/why-would-mother-nature-bother.

19 Galen Strawson, "Evolution Explains It All for You," review of *Freedom Evolves*, by Daniel C. Dennett, *New York Times*, March 2, 2003, https://www.nytimes.com/2003/03/02/books/evolution-explains-it-all-for-you.html.

One of my favorite books is by Lee Siegel, a philosopher and magician, called *Net of Magic*.[20] It's about Indian street magic and he includes an autobiographical reflection worth pondering:.

> "I'm writing a book on magic," I explain, and I'm asked, "Real magic?" By *real magic* people mean miracles, thaumaturgical acts, and supernatural powers. "No," I answer: "Conjuring tricks, not real magic." *Real magic*, in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is *not real magic*.[21]

I think this quote sums up my work in philosophy quite succinctly. I've spent my life dealing with people who want real magic – real consciousness, real free will – and I've been trying to show them that they should in each case settle for a bag of tricks. They are amazing bags of tricks, and they have the advantage of being real. Free will as "real magic" does not exist. Sam Harris is right about that. So what?

Free will as responsible autonomy does exist. But it is under threat today thanks to our increasing ability to read and direct minds. So, love your strings and protect them from puppeteers and you will have all the free will worth wanting.

---

20 Lee Siegel, *Net of Magic: Wonders and Deceptions in India* (Chicago: University of Chicago Press, 1991).

21 Siegel, *Net of Magic*, 425.

References

Claxton, Guy. "Intelligence from the Ground Up." Lecture at the Understanding Intelligence Ethics in the Age of AI conference, New College of the Humanities, London, March 22, 2018.

Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown, and Company, 1991.

———. "Reflections on Free Will." Sam Harris's official website, January 26, 2014. https://samharris.org/reflections-on-free-will/.

———. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge: MIT Press, 2005.

Fodor, Jerry. "Why Would Mother Nature Bother?" Review of *Freedom Evolves*, by Daniel C. Dennett. *London Review of Books*, March 6, 2003. https://www.lrb.co.uk/the-paper/v25/n05/jerry-fodor/why-would-mother-nature-bother.

Harris, Sam. *Free Will*. New York: Free Press, 2012.

———. "The Marionette's Lament: A Response to Daniel Dennett." Sam's Harris's official website, February 12, 2014. https://samharris.org/the-marionettes-lament/.

Heinrich, Bernd. *Mind of the Raven: Investigations and Adventures with Wolf-Birds*. New York: HarperCollins, 2000.

Hutchings, Edward, Jr. "The Autonomous Viking." *Science* 219, no. 4586 (February 18, 1983: 803–8.

Köhler, Wolfgang. *The Mentality of Apes*. Translated by Ella Winter. 2nd ed. London: Routledge and Kegan Paul, 1927.

Siegel, Lee. *Net of Magic: Wonders and Deceptions in India*. Chicago: University of Chicago Press, 1991.

Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. "Unconscious Determinants of Free Decisions in the Human Brain." *Nature Neuroscience* 11, no. 5 (May 2008): 543–45.

Strawson, Galen. "Evolution Explains It All for You." Review of *Freedom Evolves*, by Daniel C. Dennett. *New York Times*, March 2, 2003. https://www.nytimes.com/2003/03/02/books/evolution-explains-it-all-for-you.html.

Williams, James. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press, 2018.