

## Sizing Up Free Will: The Scale of Compatibilism

Stuart T. Doyle

*Third Force Reconnaissance Company, USMC*

Is human free will compatible with the natural laws of the universe? To “compatibilists” who see free actions as emanating from the wants and reasons of human agents, free will looks perfectly plausible. However, “incompatibilists” claim to see the more ultimate sources of human action. The wants and reasons of agents are said to be caused by physical processes which are themselves mere natural results of the previous state of the world and the natural laws which govern it. This paper argues that the incompatibilists make a mistake in appealing to such non-agent sources of human action. They fail to realize that free will may exist at one scale, but not at the scales where they look. When free will is considered from the correctly scaled perspective, it does seem compatible with determinism and natural laws.

Keywords: free will, scale, compatibilism, incompatibilism, determinism

I am the master of my fate:  
 I am the captain of my soul.  
 William Ernest Henley  
*Invictus*

Most people feel like their decisions and actions follow from their own free will.<sup>1</sup> But most scientists believe that all events in the world evolve according to natural laws<sup>2</sup> which do not follow from anyone’s free will. There is an apparent conflict here which is inevitably interesting to cognitive scientists, philosophers, psychologists, and physical scientists. In 1994, George Howard wrote in *The Journal of Theoretical and Philosophical Psychology* that although psychological research into human free will had started to make experimental and descriptive

---

I would like to thank Brian Doyle (US Army), for review and feedback on the central argument. Correspondence concerning this article should be addressed to Stuart T. Doyle, 3rd Force Recon Co, 1630 S Broad Street, Mobile, Alabama 36605. Email: stuartdoyle1@gmail.com

<sup>1</sup>Survey results show a majority of respondents endorsing notions of free will. See the survey study by Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer, and Jason Turner, 2005.

<sup>2</sup>I say “natural laws” to include the possibility of fundamentally probabilistic laws, which would not be deterministic, yet would pose a similar problem for many so called “hard incompatibilists.” See Pereboom, 2001.

progress, there remained a “stuck” feeling in the tension between free will and determinism (Howard, 1994, p. 76). He and others in the field were grappling with the logical coherence of their chosen topic of psychological research (Slife, 1994, p. 1) Howard came to this conclusion on the matter: “I am a determinist (in a restricted sense) and a free willist (self-determination is one among several causes of human action)” [p. 76]. According to Howard, this understanding is what enabled his experimental research program (p. 76). Howard’s conclusion was that free will is compatible with determinism because we are the things that determine our freely willed actions. Though he didn’t use the word, this was what’s known as a “compatibilist” position on free will. There has been an overlapping debate in the field of philosophy, where compatibilist positions have been elaborated and “incompatibilist” arguments have formed in opposition. Howard’s view on free will fits squarely into a category called “source compatibilism” (since he focuses on being the source of one’s own actions).

This conception of free will which Howard found to work well in psychology has not been treated gently in the world of philosophy. There is a prominent argument against source compatibilism, which convinces many philosophers that positions such as Howard’s are far from tenable (McKenna and Coates, 2018, § 5.4.4). I will be arguing against the incompatibilists and their best argument, in favor of compatibilists such as Howard and his philosopher counterparts. But I’ll go farther than Howard by incorporating an account of the ability to make decisions other than the ones we actually end up making. The question, “Do I have free will,” is often conceived as two separate questions. The first, “Am I the real source of my decision?” and the second, “Could I have chosen otherwise?” will both be answered in this paper.

Though primarily philosophical, the free will debate has direct consequence for psychological theorizing. If free will is a human capability, then a complete psychology must include a description of it, not merely a description of what people think about it. If philosophy produces an answer to the question of whether or not we have free will, it will be producing a description of the human mind and behavior. This will be an act of psychology by means of philosophy. As Carol Dweck and Daniel Molden put it in their chapter on the overlap between the philosophy and psychology of free will, “The definition or nature of free will is ultimately a philosophical question; whether people believe they have free will is a psychological one; and whether people actually have free will is in the terrain somewhere in between” (Dweck and Molden, 2008, p. 44).

The general terminology of compatibilism versus incompatibilism frames most of the debate around free will and natural laws. In this paper, I will argue that compatibilism is the right way to go, and that what many consider to be the strongest incompatibilist argument is really not strong at all. I will reach that conclusion through consideration of scale; consideration of how the facts change when we zoom in or zoom out from a given perspective. But first, I should briefly

justify my choice to work within the framework of the compatibilism debate. And then, I will describe more specifically the kinds of compatibilism and incompatibilism I will be talking about.

I am taking the side of the compatibilists on the topic of free will. There are a couple of very influential compatibilist theories which describe freely willed choices and actions as resulting from either the desires of the self, or from reasons that the self has for acting. These are source compatibilist theories, in that they focus on the source of action. Harry Frankfurt put forth the theory which explains freely willed action as issuing from desires that mesh with important elements of a person's psychology. The idea is that a person who acts of her own free will acts from desires that are nested within elements of herself. Frankfurt says that the subject's actions emanate from herself rather than from something foreign (Frankfurt, 1971, p. 17).

John Martin Fischer proposes another influential source compatibilist theory, which has become prominent. He speaks of "reasons" as the source of freely willed actions. He claims that action is responsive to rational reasons in a free agent; that if the reasons were different, the action would be different as well. His view is not merely that an agent would be responsive to reasons in some counterfactuals, but rather that her responsiveness to reasons in the counterfactual is evidence that her actual conduct is also responsive to rational considerations (Fischer, 1994, p. 166).

I am not arguing that either of these source theories of free will is exactly right or complete. I am making a case in this paper against the negation of the *kind* of theories exemplified by Frankfurt and Fischer above. That is, I am arguing against *source incompatibilism* and the thought experiments meant to support it. After I make my case for the plausibility of source compatibilism, I will discuss how sourcehood is related to another major question about free will. The ability to have multiple paths open to oneself, or "the power to do otherwise" is often seen as a separate question from that of sourcehood (Vihvelin, 2018, § 2). It will turn out that sourcehood is very important in figuring out the issue of the power to do otherwise. So before I can make any progress on that issue, I need to argue that we are the sources of our own actions, against the source incompatibilists. Michael McKenna and Justin Coates aptly describe source incompatibilism in their encyclopedia review:

If determinism is true, then for any person, what happened in the past prior to her birth that, when combined with the laws of nature, provides causally sufficient conditions for the production of her actions. But if this is so, then, while it might be true that an agent herself provides a source of her action, that source, the one provided by her, itself has a further source that originates outside of her. Hence, she, as an agent, is not the ultimate source of her actions....

1. A person acts of her own free will only if she is its ultimate source.
2. If determinism is true, no one is the ultimate source of her actions.
3. Therefore, if determinism is true, no one acts of her own free will. (McKenna and Coates, 2018, § 2.2)

In support of source incompatibilism, some thought experiments called "manipulation cases" have been developed in order to crystalize the incompatibilist intuition. These manipulation cases ask us to imagine an agent whose thoughts and actions are manipulated by some evil scientist. Though apparently not free, this manipulated agent could meet the free will criteria put forth by either Frankfurt or Fischer. So manipulation offers a counterexample to source compatibilist theories. And manipulation cases have been taken farther by Derk Pereboom in his Four-Case Argument for incompatibilism, in which he tries to show that normal reality is not different in any relevant way from a manipulation case (Pereboom, 2001, p. 112). I will describe Pereboom's argument in detail in the "Manipulation Cases and Real Life" section of this paper. In short, he says that having our actions determined by physical events outside of our control is the same in all relevant ways as having our actions controlled by an evil scientist, because the physical events and the scientist are both ultimately out of our control. Pereboom's argument has been widely regarded as very effective. McKenna and Coates say of manipulation cases, especially Pereboom's: "Why is determinism any different from a manipulation case? The burden, it seems, is on the compatibilist to show how it is that manipulation cases differ from a normal deterministic history" (McKenna and Coates, 2018, § 5.4.4). One of the conclusions of this paper will oblige McKenna and Coates by doing exactly that: I will show how manipulation cases differ from a normal reality. And more generally, I will show why source incompatibilism is malformed. All of this will rely on a certain understanding of what I am calling scale, which I will now describe.

### *Scale Changes Things*

Many facts and properties depend on scale. Take, for example, the color of a red apple. At macro scale, where we usually think of apples, it is truly red. But if we zoom far enough into the micro scale (technically to the pico scale), where atoms dominate, the surface of the apple is truly not red. None of the atoms is red. The concept of "red" does not even apply to atoms. Likewise, we can zoom out to the mega scale, where the apple is just a part of some grand entity or process. The agricultural industry is not red; the consumer demand which brought our apple (with all of its color) into existence is not red.

So we see that a property as simple and real as color can not always survive analysis which jumps between different scales. I do not think this should lead us to conclude that the apple is not actually red. It should tell us that when trying to answer whether or not the apple is red, our analysis should stick to the macro apple scale; not the micro atom scale, and not the mega system scale. We should avoid the common assumption that there is truth or insight to be found by zooming in or zooming out. There is a tendency to act as if to investigate means nothing more than to zoom in or out. But the apple proves that there is also truth to be

lost by zooming. Truths of one scale may not apply on other scales. The truths of different scales can contradict each other if lifted from their homes, and clumsily mashed together in a scale-free theory.

I use the word "scale" to refer to a frame of inquiry which explicitly or implicitly treats a thing or a type of thing as a whole unit of analysis. Wholes often have quantitatively and qualitatively different properties than their parts. To identify which scale we're looking through at any given time, we need to ask what we are taking as a unit of analysis. What are we treating as a whole? In the example above, when we ask about the color of an apple, we are at apple scale, because we treat the apple as a whole. In apple scale, atoms are merely parts of a whole thing, and the agricultural industry is not a thing, but a nominal short-hand for many whole things. The properties which can be described at apple scale are mostly very different from the properties which can be described at either atom scale or industry scale. There are multiple identifiable reasons why wholes are often quantitatively and qualitatively different from their parts. I will briefly mention a few of them for background, but these particular reasons are not strictly necessary to the main point I will make in this paper. I will argue that people have the property of free will, although their parts do not. I can make this argument without identifying *why* this property changes across scales. But I'll now describe three general categories of scaling mechanisms to help with the intuition.

*Size.* Many properties of many things depend directly on their size. An atom can't be red because the atom is smaller than the wavelength of red light. A bigger object can't have the structure or behavior of an atom because nuclear and electromagnetic forces decrease exponentially with distance. Size directly changes the properties of even abstract objects. A large cube can not have the same ratio of volume to surface area as a smaller cube. And abstract size scaling clearly has implications for physical scaling. For example, all else being equal, a larger object can not radiate its heat as quickly as a smaller object can. And this general physical size scaling mechanism has implications for many properties of particular physical objects. For example, the earth has the property of having its own magnetic field, which could not be true if the earth had a smaller volume to surface area ratio.

*Non-Ergodicity.* Measures of groups do not necessarily reflect the individuals which contribute to the measurement. This can apply to individual molecules which make up a gas, individual people who make up a society, individual neurons which make up a brain, and many other collections of things. It also applies to individual temporal parts of singular things over time.

*Organization.* The properties of a whole depend heavily on the way its parts are arranged relative to each other, and relative to the larger structure. The whole often has important functional properties, while being composed of parts which completely lack those properties. A nonlinear dynamical system may be constructed of simple parts. The system is chaotic, fundamentally unpredictable. But in isolation, each part is perfectly simple and predictable. For example, an ideal double-rod pendulum is made of only two functionally simple parts. The behavior

of the whole double-rod pendulum is rich, complicated, and nonperiodic. The reason that the functional properties of the whole double-rod pendulum are different from those of its parts is that each of the two parts has inputs on each other. The input from the first part affects the behavior of the second part, which affects the behavior of the first part. Described in this way, the double-rod pendulum is a two-node network. Many larger networks also have properties which none of their nodes has. Simulated neural networks can perform tasks such as differentiating between cats and dogs, while each simulated neural node can do no such thing. The human brain is a network of networks, and it too has properties as a whole that none of its parts has.

I will argue that people have the property of free will, although their parts do not. The three general categories of scaling mechanisms briefly described here give a sense of how such a claim might be true. But I make no attempt to determine which mechanisms work, or how they work to result in differing properties at different scales. This would be an impossible task given the present state of cognitive neuroscience and philosophy of mind. There is no model which comes remotely close to fully explaining how human mind and behavior emerge from the brain and its parts. Neuroscientists have taken precise measurements of the synaptic wiring in *C. elegans*, a roundworm 1 mm in length, with a total of 302 neurons. With those data, researchers built simulations of the 302 neuron network at the cellular level, including the roundworm's entire muscular system and representations of its physical environment (Bryden and Cohen, 2004, p. 183). Despite the comprehensive measurements and simulations, it remains unclear to researchers how the actions of 302 neurons combine to generate the behaviors observed in the *C. elegans* roundworm (Lynn and Bassett, 2019, p. 324). Human brains have roughly 86 billion neurons (Azevedo et al., 2009, p. 532). The number of possible relationships between neurons grows exponentially as the number of neurons grows linearly. So to explain exactly why or how different properties come to be found at different scales of the brain would be completely impossible right now.

The reasons *why* whole people have different properties than those of their parts is not a central issue of this paper. What's important here is that whole people *do* have different properties than those of their parts. When examining the properties of a person, it is important to be clear about whether we're actually looking at properties of the person, or of her parts. I call this scale. There are many scales at which we might look, since there are many parts, which are made of smaller parts, which are also made of yet smaller parts. We will find different truths at each scale because different kinds of properties exist at each scale, which have their own dynamics. The particular reasons that the properties and dynamics at any given scale differ from those of another scale may be obscure. They likely include reasons from some or all of the general categories: size, non-ergodicity, and organization. And there may be whole categories of reasons yet to be conceptualized. But regardless of the reasons for the difference, we can often know that there is a difference between scales.

If we want to answer the question of whether or not people have free will, we have to figure out where in the world of scale we should be looking for truths, from which we can build our answer. For the question of apple color, the right scale for the answer was simply the same as the question scale: the apple scale. I contend that the same is true of the free will question; that we need to look for answers on the same scale as that of the question: the human agent scale. I use the term "agent scale" to refer to a certain level of size and complexity, where a whole person is the unit of analysis. I do not use the word "agent" in order to beg the question of agency, but to use the established jargon in the compatibility discourse as described by McKenna and Coates to refer to an acting person who *prima facie* seems agentic (2018, § 2.2). I will argue that the agent scale is the right frame from which to answer the question of free will. I will also show that prominent compatibilists (Frankfurt and Fischer) have operated on this correct scale, while their incompatibilist opponents have been inappropriately zooming between scales in order to build their theories.

First, I have to argue that free will is like color; that there is such a thing as right and wrong scales at which to look for it. Then I will have to make a good case that compatibilist theories like Frankfurt's and Fischer's look for free will at the right scale, and that source incompatibilists look at wrong scales. If I can do this, it will constitute good reason to favor compatibilist theories. Though there may be other issues with the major compatibilist theories, the landscape will become much more favorable for compatibilism in general after giving good reason to reject the source incompatibilist critique, which has so far not been satisfyingly answered.

### *Free Will is Scale-Dependent*

Behavior has causes at many different scales which are simultaneously real and true. Here are a couple of rough examples:

1. Agent scale. I moved to a new city across the country because I wanted to. I wanted to move because I perceived a good reason to; and because I am the type of person that responds to reasons, and that does not fear the type of change entailed in moving.
2. Neurological Scale. I moved to a new city across the country because a specific combination of motor neurons in my brain and spinal cord fired action potentials in a certain sequence that caused certain muscle fibers to contract in a certain sequence. Those motor neurons were caused to fire by the neurotransmitters released from the firing of other neurons, which were also caused to fire by the firing of other neurons.

Let's think for a minute about the neurological scale so that we can see that free will is scale-dependent, that there are scales at which we should know not to look for free will, much like we know not to look for color at certain scales. Everyone already knows not to look for free will in a single brain cell. Though I am not aware of any compatibilists nor incompatibilists who have bothered to spell this out, neither camp seems to ever explicitly set out on the project of finding the cell with free will. This is because a single cell does not experience action or potential action as the object of desire, want, or contemplation; it does not experience any conscious desire or contemplation at all. So if we look for free will in cells qua cells, we will never find it, no matter how large the collection of cells we search. Note that I have said nothing of determinism (or probabilistic natural laws) and the mechanistic nature of cells. We already know not to look for free will at the neurological scale before we ever start thinking about compatibilism and incompatibilism. The concept of free will simply does not apply at that scale, much like the concept of redness simply does not apply at the pico scale. The concept of free will only applies once we look at the collection of brain cells as an emergent unified subject, which is to say, once we zoom out of the neurological scale, into the agent scale. We have wills. That is, we want things, we plan actions, and we often act in accordance with our own wants and plans. Will is real. The question is whether or not those wills of ours are free or not. The question of whether the will is free or not is only coherent at the scale where there is will. Neurons and molecules do not have wills, so we should not be impressed by the observation that they do not have *free* wills.

Since there are some wrong scales at which to look for free will, any theory which deals with free will needs to be checked for reliance on wrong scales. If any of the key steps in a theory depend on some intuition mined from one scale and misapplied to another, then the theory does not work. This is what is going on in the case of source incompatibilism. When one declares that for free will to obtain, the agent must be the *ultimate* cause of action, he implicitly demands that free will should make sense at scales where free will was never supposed to make sense. This becomes clear when we ask: What are the non-agent causes which disqualify the agent from being the ultimate cause? That is, which causes are *more* ultimate than the agent? There must be some more ultimate causes implied. Otherwise, the agent herself could be declared ultimate.

So what are these more ultimate causes? Philosophers seem to avoid enumerating them. But there are some that can be inferred. Most of them, especially in real life, are versions of, "Our brains are made of molecules; those molecules must obey the laws of physics; our decisions derive from brain activity" (Coyne, 2019, ¶ 4). This is the popular formulation of source incompatibilism, which is circulated by public intellectuals such as Jerry Coyne and Sam Harris (2012, p. 9). Clearly, this version of source incompatibilism jumps between scales in order to assess free will. It assigns the implicit title of more ultimate cause to the actions of molecules.



But free will is supposed to evaporate when we look at the molecular scale; we should expect it to, without causality and natural laws ever entering the question. So when the incompatibilist directs us (explicitly or implicitly) to look at the neurological or molecular scale and find no free will there, we should not conclude that there is no free will at agent scale. Remember the lesson from apple color: a property can be perfectly real (or descriptively true) at one scale, but not at others.

If the source incompatibilists were to search the correct scale for more ultimate causes of an agent's actions, what could they find? Let's find out, using McKenna and Coates' characterization of source incompatibilism as a starting point. "What happened in the past prior to [an agent's] birth, when combined with the laws of nature, provides causally sufficient conditions for the production of her actions" (McKenna and Coates, 2018, § 2.2). What actually did happen in agent scale dynamics before my birth? My parents decided to have a child. Lee Harvey Oswald shot John F. Kennedy. Steve Jobs built a computer. My mother composed some music. Suppose we list the whole set of agent scale events that happened before my birth. Does the whole set or any part of the set predict that I would move across the country on the day that I did? Does it explain why I did not move on another day or to a different destination? No, an all-knowing demon would need to reference more granular events, far below the agent scale in order to identify a chain of causes that would predict and explain my recent actions. The whole set of agent scale events that happened before my birth could have been instantiated by slightly different atomic scale arrangements. Every statement about what people thought and did before my birth could still be true, and I could have ended up with a slightly different genome. The precise genome I did end up with is better predicted and explained by events on the biochemical and microbiological scales, not the agent scale. And a small difference in my genome would have made my mind and body slightly different, which would cause different decisions, which would compound over the course of my life, leading me to presently be in a completely different situation, having moved at different times to different destinations.

Finding the source incompatibilist's more ultimate causes of my actions inevitably requires searching non-agent scales. Nothing like a sufficient cause can be found at agent scale before my birth. The source incompatibilist's more ultimate causes must be at scales where we know not to look for will. There is no will in the microbiological scale dynamics that caused the details of my genome, but this fact can not tell us that the will is not free at its own scale. It is important to note that the microbiological scale truth is not that the will is not free; it is that there is no will. That is a huge difference. If we look to the microbiological scale truth to inform us of the nature of the agent, we should conclude that the agent has no will, not that the agent's will is not free. But we know first-hand that we do have wills; we consciously want to do certain things and not others, and then we do those things which we want to do. So we should not look to truths from any scale where there is no will to inform us of the nature of the agent. It would lead us to

contradict what we know to be true; it would be inappropriate zooming, analogous to looking for redness in an atom on the surface of an apple. But this wrong move is inevitable in the search for source incompatibilist's more ultimate causes of our willed actions.

### *Manipulation Cases and Real Life*

Now that I have shown what is in principle wrong with the source incompatibilists' implicit reliance on more ultimate causes of action, I will make clear where this error shows up in their best argument. As mentioned above, the consensus as seen by McKenna and Coates is that manipulation cases pose the greatest challenge to compatibilists. And the most forceful of these manipulation case arguments is Derk Pereboom's (2001, p. 112) Four-Case Argument for incompatibilism. As briefly described earlier, a manipulation case is a thought experiment where the more ultimate cause of an agent's action is another person, who manipulates the agent. The idea is that we are supposed to intuit that a manipulated person is not free, yet she might fit compatibilist criteria such as wanting to act as she does, or having reason to act as she does. In a good manipulation case, the agent is manipulated into wanting to do, or having reason to do, the action which she is manipulated to do. Pereboom ties such a case to normal reality by using four cases which get progressively more normal and realistic.

First is electronic stimulation of the subject's brain, causing her moment to moment decisions and actions in such a way that she still satisfies (according to Pereboom) what various compatibilists contend is sufficient for free will. That is, the manipulated subject experiences a desire to do what the scientist makes her do, and she has her own apparently coherent reasons for doing what the scientist makes her do. As a sophisticated manipulator, the scientist makes sure to induce these wants and reasons in the mind of the subject by means of complex and precise electrical stimulation of the brain. In this first case, compatibilist criteria for free will is shown to be insufficient, since we would agree that the manipulated subject is not free. Pereboom then presents a second case in which the manipulator tinkers with the agent ahead of time by "programming in" the relevant psychological details that will later lead to the agent's actions. We are meant to find no relevant differences between the two cases in terms of free will. For the third case, Pereboom takes another small step toward normal life: rather than direct electrode stimulation, the manipulation consists of indoctrination of the subject in her childhood, which causes her specific actions in adulthood. This process is said to not be different from the pre-programming case in any relevant way. And so here too, we are meant to conclude that the agent is not free. Pereboom claims that the best explanation for why these first three agents are not free and responsible is that the process of manipulation deterministically caused each of them to act as they did. And so, finally, the fourth case is that in which a normally

developed agent in a determined world acts (basically a description of normal reality). This case also is said to be no different from the others. In all four cases, what underwrites the lack of freedom, Pereboom concludes, is the fact that the source of the agent's actions can be traced back to originating conditions that were completely beyond her control.

Note that Pereboom's Four-Case Argument makes at least two points. The first is, like in other manipulation arguments, that compatibilist theories are shown by counterexample to be flawed. The second point is that normal reality is, in all relevant senses, just like the manipulation cases. The first point, I think is right, in that it at least shows that some compatibilist theories are incomplete. The second point is wrong. By considering scale, we can see why the second point is wrong and, in the same stroke, hint at what the compatibilist theories need to add in order to become more complete.

Pereboom's jump in scale, his inappropriate zooming, happens between cases three and four. The first three cases all manage to remain at the same scale in a subtle way. In the first three, causality flows perfectly from the manipulator's desires and reasons into the subject's actions. Where person-level things like desires and reasons can definitively, with perfect fidelity, cause actions, we are more or less looking at agent scale. This is where *will* can make sense. Here, we can appropriately talk about whether or not a will is free; because here, there is will. Depending on the character of the will we observe at this scale, we might judge it free or unfree. In Pereboom's first three cases, we judge the will of the subject unfree because there are desires and reasons other than the subject's own desires and reasons, which cause her actions. We can say this without jumping scales. To clarify by way of my analogy, we judge the color of an apple to be non-red if there are colors other than red (such as blue) all over the apple. We make this judgement without jumping scales; without making color a meaningless concept.

Pereboom tells us to imagine three controlled agents, which we agree are in fact not free. Then he points out truthfully that all three unfree agents are directed by deterministic mechanisms. And finally, he says that since all real agents are likewise directed by deterministic mechanisms, they are likewise unfree agents. Along the chain of his argument, he surreptitiously disposes of the control imposed by the evil scientist, which was the entire basis of our agreement about the unfreedom of the first three agents. Without the control exerted by the manipulator, we have no good reason to agree that all real agents are unfree. In the first three cases, the will of the evil scientist is the more ultimate cause; it is sufficient to predict and explain the actions of the manipulated agent. So we never leave the agent scale where we can coherently speak of will, and we validly judge the will of the manipulated agent to be unfree. The fourth case is completely different because there is no more ultimate cause of the agent's actions at agent scale. To predict and explain the actions of the agent, an all-knowing demon would have to search through scales where we can't coherently speak of will (such as chemical or neurological

scales). Lifting truths from these other scales does not produce a coherent answer to the question, "Does the agent have free will?" Answering "no" and answering "yes" are both incoherent when speaking of causality on non-agent scales.

Pereboom smuggles the intuition of imposed control from the manipulator at agent scale, misapplies it at other implicit scales (such as the chemical, biological, economic, or whatever other scale on which our actions are determined), then he takes the deformed product of that misapplication back to agent scale without the imposed control on which the intuition depends. Intentional control, whether it be the evil scientist's control over the subject or the subject's own control of herself, is a feature only found at agent scale. Control is nearly synonymous with will here, but we are not accustomed to saying that one person wills the actions of another; we would say that one person controls the actions of another. If one controls herself, she is acting according to her will. Being controlled by someone else is what makes one unfree in a manipulation case. When Pereboom removes that other person (the scientist) in the fourth case, he implicitly appeals to non-agent determinants of our actions, which must be significantly smaller, bigger, more complex, or less complex than agents (otherwise they would be agents). Our intuition tells us we are not free if we are controlled. That intuition should not be transplanted into another scale where there is no such thing as control. Control entails a certain scale, and leaving that scale should entail leaving behind any intuitions based on a sense of control. There are completely different features, properties, and dynamics at different scales. In appealing to non-agent causes at their necessarily non-agent scales, Pereboom analyzes our wills at scales where there is no will; he misappropriates the sense of being controlled to scales where there is no control. He finds no free will where there is no will at all, and then he jumps back into the agent scale, where there is will, to declare us unfree. This move is unsound, to say the least. If the more ultimate causes of my will are not at agent scale, then they can't tell us about the nature of my will. It can't even be determined on the basis of those more ultimate causes that there is a will.

Pereboom's fourth case implicitly relies on more ultimate causes at scales where the question of free will is incoherent — unlike his first three cases, which do not rely on making the question incoherent by jumping scales. So the fourth case (reality) is not, in all relevant senses, just like the three manipulation cases. Pereboom's conclusion is refuted. And compatibilists now have a hint of what is needed to make their theories more complete: something to accommodate a concept of unfree will which can exist alongside free will, much like blueness exists alongside redness.

### *I Am the Source*

My point about things changing with scale is very similar to a point that many complexity theorists have made. But the mathematics of complexity theory are

truly beside the point here; there is no known mathematical model which can encapsulate a conscious wanting to do or not do certain actions. We know that we consciously want to do or not do certain actions, and that our actual actions correlate with these conscious wants. We call this will. Before questioning whether or not our will is free, we know that a description of humans would be incomplete without a description of will (which might be free or unfree). Also, we know that no brain cell has such will. A description of neurons is complete without a description of will. Thirdly, we know that we (who have will) are "made up" of neurons (which have no will). Therefore we know from a simple combination of first-hand introspection and common knowledge of physiology that the "same thing" (a person) has different properties and dynamics at different scales. Without doing any math, we can conclude that important features of the self emerge at the agent scale, but do not exist at lower scales. Will is a feature of the self; it is not a feature of neurons. Redness is a feature of apples; it is not a feature of atoms. So a definition of will is: "a conscious want to do or not do certain actions, where ensuing actions correlate with the conscious want." Will is a feature of the self, and "free" or "unfree" are descriptors of that feature. The question is whether or not the "free" descriptor is at least sometimes true of the feature of the self called "will."

Now that we are thinking of will at the right scale, we can ask whether or not our wills are free. So we need a definition of "free." There is a common definition which fits the source compatibilist conception: "not under the control or in the power of somebody else; able to do what you want" (Oxford, 2020, § 4). Under this definition, our actions would not be free if they were controlled by someone else, and our wills would not be free if our desires were controlled by someone else. The straightforward negation of freedom describes manipulation cases. But if our desires are not controlled by someone else, then under this definition, our wills are free. The source incompatibilist would mainly contend for a modified definition of "free": not under the control or in the power of somebody or *something* else.

Why should we reject this modified definition? We should reject the source incompatibilist definition because the "something else" is just the actions of little bits of matter below the scale of the self-aware agent. At the scale where these non-agent neurological functions take place, is the will controlled, or is it free? There is no will at that scale, so it can neither be controlled nor free. If we can find no will at the neurological scale, how can we truthfully describe it from there? Put another way, we know that the neurological scale fails to inform us about the reality of the agent scale, because we know that will is real (whether or not it's free), but we also know that no such thing can be found when we zoom in too far. So when we ask about whether or not will is free, we are asking a question which should be answered at the agent scale, since we know the neurological scale fails to inform us about will. Without knowledge taken from the agent scale, the neurological scale would lead us to conclude that there is no conscious want to do or not to do certain actions. But we know that there is.

So where should we look for truths about conscious want; to the source that would deny its existence? No.

I take the existence of the will as a premise in this paper, trusting that most people can relate to the first-person experience of wanting to do something, deciding to do it, and then doing it. I take the claim that the will is contingent in some way on the actions of neurons (or molecules or quantum fluctuations) as a second premise. And thirdly, I use this observation as a premise: nothing we can see in neurons, molecules, or quantum fluctuations looks like wanting to do something, deciding to do it, and then doing it. Nothing about the micro scale components of the human body tells us that there is a will associated with it. In fact, if we took facts about the micro scale as applicable to the macro scale, the nonconscious character of each micro scale component would lead us to believe that there is no will associated with the body. But there is. And what do we find when we search through the micro scales for more ultimate causes of our willed actions? We find chemical events causing other chemical events; we find neural events causing other neural events; we never find a chemical or a neural event causing the will. We find many chemical or neural events taken together as a unified whole unit of analysis causing the will. But that unified whole unit of analysis is the human, the self, the agent. So my conclusion is that the self is the cause of the will and of willed actions. Attributing more ultimate causal status to events or objects other than the self entails attributing more ultimate status to causes at scales where we could never coherently say that the will has been caused.

As an illustration, suppose Benjamin Libet wants to measure a few neurons in my brain while I decide to press one of two buttons. When Libet measures the electrical activity in part of my brain, he says, "Your neurons showed a readiness potential to press each button before you pressed it, and before you reported deciding which button to press."

"Of course my action was contingent on many different events at many different scales. You found one of them. My decision was still a free act of my will, in that it followed from my own character and reasoning, which are intrinsic to my self. Your electrode didn't tell you about my character and reasoning; it didn't even tell you that I exist as a thinking experiencing being. What can it tell you about my features if it's incapable of even telling you that I exist?"

Derk Pereboom steps in: "But your will was not the ultimate cause of your actions. Your character and reasoning were also caused by cellular machinery."

I reply, "Not exactly. Microscopic cellular events cause other microscopic cellular events; none of which is equivalent to my character and reasoning. Only the entire package of cellular events taken as a whole was sufficient to cause my character and reasoning. And that entire unified emergent package is called my self. I caused my own character and reasoning, which caused my will, which caused my action. The agent scale character and reasoning were conditional on certain micro

scale events, but even that relationship is weak. If one of my neurons was replaced by a silicon transistor which was calibrated to sum synaptic inputs and generate an electrical output indistinguishable from the electrical output of the original neuron, my agent scale character and reasoning would be indistinguishable from the way they actually are. The agent scale character and reasoning are realizable by means of many different arrangements of micro scale events. So the micro scale events which you (hypothetical Pereboom) claim have caused my decision were neither necessary nor sufficient.”

Scale gives us a clear answer to the question, “Why is determinism any different from a manipulation case?” It is different in that normal deterministic (or natural law) causes of action other than agent-causes do not exist at agent scale. If they did exist at agent scale, they would be agent-causes. Non-agent-causes exist at non-agent scales, where will itself can not be found, and where free will was never supposed to apply.

This is a compatibilist argument, not quite a “libertarian” argument.<sup>3</sup> The will is caused by the self taken as a whole unit. The present state of the self is caused by past states of the self. There are many little micro-causes which, in complex combination, constitute the agent scale self-cause. But what I have argued is that these micro-causes are *deeply irrelevant* in comparison with the self-cause. The simple analogy helps in seeing what deep irrelevance is, and why it arises in issues of scale. An apple is made of 100% non-red parts; that is true. But it is not truly relevant to the question “Is the apple red?” It is deeply irrelevant. In this way, red apples are compatible with colorless atoms, but not fully independent of them. In the question of free will, there’s a similar deep irrelevance in the source incompatibilist’s more ultimate causes. It is true that my desires and actions depended in some way on micro scale events in the time before I was born and in the tissue of my own brain. But that dependency is not truly relevant to the question, “Is my will free?” The micro-causes exist, which would be denied by libertarians. But the deep irrelevance of the micro-causes makes them compatible with free will.

There is certainly more to develop in compatibilist theory, but now the strongest argument against compatibilism does not look so strong. When proper scale is considered, compatibilism shines. When compatibilists say that your choices and actions are caused by your reasons or desires, and that your desires are caused by your character, these are real causes, and they are the relevant kind of causes. As far as free will is concerned, they are the only kind of relevant causes. And isn’t there something ultimate about that which is the only relevant kind? We are as free as an apple is red.

---

<sup>3</sup>Free will libertarians say that the will is just not subject to deterministic cause and effect.

*The Ability to Do Otherwise*

So far, I've focussed on the sourcehood notion of free will; the notion that we have free will if we are the true source of our decisions. But there's another way that people frame the question of whether or not we have free will: the ability to do otherwise. They ask, "When I made a choice, did I have the ability to choose something other than I did?" Determinism seems *prima facie* incompatible with the ability to do otherwise. Merely framing the question of free will in the language of alternative possibilities is enough to rule out compatibilism for many. Psychologist Jonathan Schooler writes:

In principle, the compatibilist perspective is ideal as it allows us to have our material cake while freely choosing to eat it, too. The problem is that I, for one, simply cannot get my head around compatibilism.... I just don't understand the following: If any non-coerced reasoned decision that I am about to make must necessarily be carried out in a specific manner based on a preexisting causal chain, then how can I be free to choose otherwise? And if I really have no option but to do exactly what I end up doing, how can it be said that my choice was free? (Schooler et al., 2008, p. 195)

Schooler is making the same point made in the philosophy literature most famously by Peter van Inwagen. "It's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us" (van Inwagen, 1983, p. 56). In the philosophy discourse, this issue of "the power to do otherwise" has been largely treated as a separate question from the issue of sourcehood (Vihvelin, 2018, § 2). But it turns out that a properly scale-sensitive conception of sourcehood greatly helps in thinking about the power to do otherwise.

The question of the ability to do otherwise spirals through a self-referential cycle until it collapses into the question of sourcehood. Here's an example of that cycle:

"I could have done otherwise if I wanted to."

"But you couldn't have wanted to."

"I could have wanted to if my character was different."

"Your character couldn't have been different."

"My present character is a result of my past choices, which could have been different if I wanted them to be different."

"But you couldn't have wanted them to be different because your original character resulted from causal events which happened before you existed at all."

"Well now you're appealing to "more ultimate causes," which you'll find at all the wrong scales, just like the source incompatibilist. Now we're speaking of scales and times at which I could not even do what was actually done."



The question, "Could I have done otherwise?" is about "I," and since scale affects the intelligibility of "I," the question must not be applied to scales where "I" is unintelligible. That is, the question should not be applied to all of the little events in the causal chain before I was born, nor the little causes in the parts of my brain. Yet the feeling of unfreedom evoked by the question, "Could I have done otherwise?" relies completely on doing exactly that; shifting the analysis to a point in time or scale where the concept of self becomes unintelligible, where there is no "I." The following example reveals how and why this is done, and why it is wrong.

Suppose that yesterday, I apparently chose between action A and action B. And suppose that before I was born, my genome might have ended up in configuration  $\alpha$  or configuration  $\beta$ . Configuration  $\alpha$  is part of the causal chain resulting in action A. Configuration  $\beta$  is part of the causal chain resulting in action B. I chose action A yesterday, and I say that I could have chosen action B if I wanted to. Schooler tells me that it was already determined by  $\alpha$  that I could not want to choose B. And since I could not choose  $\beta$ , this rules out the possibility of B. And since I had no option except A, my choice doesn't seem free. But if my argument about sourcehood is right, I am the source of action A, despite not being the source of configuration  $\alpha$ . That means I could not bring  $\alpha$  to be, but I could bring A to be. My inability to produce  $\alpha$  does not rule out the possibility of me producing A. So why should my inability to produce  $\beta$  rule out the possibility of me producing B?

At this point one might respond that it's not my inability per se that rules out B; it's the actuality of  $\alpha$  that rules out B and rules out my ability. But the only reason that the actuality of  $\alpha$  matters here is so that  $\alpha$  can function as an example of something I can't change. If we agreed that I could change  $\alpha$ , then its actuality would be irrelevant. We take  $\alpha$ 's actuality as fixed solely to establish that I can't change  $\alpha$ . I can slip away from the incompatibilist as long as I can say things like "I could have if I wanted to." So the whole point of bringing up the time before my birth is so that the incompatibilist can point to things like  $\alpha$ , which I could not have done differently. But how do we really know that I can't change things like  $\alpha$ ? Because I can't be a cause of things that happened before my existence. That is, I have an inability to produce  $\alpha$ . So the actuality of  $\alpha$  only matters in this question because of my inability to produce  $\alpha$ . My inability to produce  $\beta$  is supposed to rule out the possibility of me producing B. But my inability to produce  $\alpha$  does not rule out the possibility of me producing A. So why should my inability to produce  $\beta$  rule out the possibility of me producing B? It does not. So the tactic of referencing the causal chain before my birth neutralizes itself.

In order for the question, "Could I have done otherwise?" to elicit a convincing feeling of unfreedom, it needs to escape the self-referential cycle in which I might say "Yes, if I wanted to." It needs to reach a point where my option to recursively reference my self disappears. It needs to shift the analysis to a point in time or scale where the concept of self becomes unintelligible, where there is no "I." But the point is to figure out whether or not I have free will right now. I know that I

have will. Is it free? Before forming any answer, I know not to lift my answer from a scale or time where I do not exist or where will in general can not exist. This means we actually should keep the analysis framed in the time and scale where I can recursively reference myself. And the feeling of unfreedom is not so strong there.

Could I have done otherwise? If I am to answer the question at all, I prefer to interpret the question in a way that keeps it intelligible, in a way that the will can actually be examined. And the will can only be examined qua will at the time and scale where the will exists. Within that time and scale, what prevents me from doing otherwise? I do. My own character and preferences determine whether I choose one action or the other. In a sense, I *have* to do what my own character and preferences lead me to do, so there is a sense in which I have to do what I actually do, but what I actually do is literally up to me, in the sense that my character and preferences are me. That seems pretty free, and it seems pretty deterministic; it is compatibilistic. It's not a simple clean answer, but it's more coherent than the alternative on offer. Pointing this out should help dissipate the incompatibilist intuition that Schooler and others feel when they ask themselves whether they could have chosen something other than they did.

### Conclusions

Considering scale allows us to see that we are the real sources of our own actions, and that we could choose to do otherwise if we wanted to (or if we had reasons to). Derk Pereboom's argument, which was seen as the biggest obstacle to source compatibilism, has been shown to be faulty. George Howard's psychological source compatibilist position is vindicated. And the incompatibilist answer to the "ability to do otherwise" question is shown to be unintelligible. In all psychological and philosophical inquiries into free will, we should be sure that our unit of analysis puts us at the scale where there is will. Where there's a will, there's a way.

### References

- Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Jacob Filho, W., Lent, R., and Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532–541.
- Bryden, J., and Cohen, N. (2004). A simulation model of the locomotion controllers for the nematode *Caenorhabditis elegans*. In *From Animals to Animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior* (pp.183–192). Cambridge, Massachusetts: MIT Press.
- Coyne, J. (2019). Why we shouldn't bet on having free will — a reply to William Edwards. *Quillette*, July 17. Retrieved from <https://quillette.com/2019/07/17/why-we-shouldnt-bet-on-having-free-will-a-reply-to-william-edwards/>
- Dweck, C.S., and Molden, D.C. (2008). Self-Theories: The construction of free will. In J. Baer, J. C. Kaufman, and R. F. Baumeister (Eds.), *Are we free?: Psychology and free will* (pp. 44–64). New York: Oxford University Press.

- Fischer, J.M. (1994). *The metaphysics of free will: An essay on control*. Cambridge: Blackwell Publishers.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68, 5–20.
- Harris, S. (2012). *Free will*. New York: Free Press.
- Howard, G. S. (1994). And binding nature fast in fate, left free the human will. *Journal of Theoretical and Philosophical Psychology*, 14(1), 73–78.
- Lynn, C.W., and Bassett, D.S. (2019). The physics of brain network structure, function and control. *Nature Reviews Physics*, 1, 318–332.
- McKenna, M., and Coates, J. (2018). Compatibilism. *The Stanford encyclopedia of philosophy*. Retrieved from <https://plato.stanford.edu/archives/win2018/entries/compatibilism/>
- Nahmias, E., Morris, S., Nadelhoffer, T., and Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology*, 18, 561–584.
- Nichols, S. (2008). How can psychology contribute to the free will debate? In J. Baer, J. C. Kaufman, and R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 10–31). New York: Oxford University Press.
- Oxford. (2020). *Oxford learner's dictionaries*. Retrieved from [https://www.oxfordlearnersdictionaries.com/us/definition/english/free\\_1?q=free](https://www.oxfordlearnersdictionaries.com/us/definition/english/free_1?q=free)
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Schooler, J., Shariff, A. F., and Vohs, K. D. (2008). The hazards of claiming to have solved the hard problem of free will. In J. Baer, J. C. Kaufman, and R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 181–204). New York: Oxford University Press.
- Slife, B. D. (1994). Free will and time: That "stuck" feeling. *Journal of Theoretical and Philosophical Psychology*, 14, 1–12.
- van Inwagen, P. (1983). *An essay on free will*. Oxford: Clarendon Press.
- Vihvelin, K (2018). Arguments for incompatibilism. *The Stanford encyclopedia of philosophy*. Retrieved from <https://plato.stanford.edu/archives/fall2018/entries/incompatibilism-arguments/>