Maria Sekatskaya*

Androids, Oracles and Free Will

<https://doi.org/10.1515/krt-2021-0010>

Published online September 2, 2021

Abstract: Claims about freedom and predestination are ubiquitous in movies, novels, and myths. These claims touch upon the philosophical problem of the compatibility of free will and determination. In order to make an informed judgment about whether these claims are true, it is helpful to know what philosophers have to say about free will. However, philosophical discussions are usually absent in popular culture. It is perhaps no wonder, since free will is a topic that has been discussed for millennia. Consequently, these discussions have become quite complicated and not easy to decipher. This essay is a popular introduction to these discussions, accessible to readers without former acquaintance with the topic. I provide references to other works for readers interested in knowing more. My aim is to show that while there are persuasive arguments against free will, there are no fewer persuasive replies to these arguments. Free will is a fascinating topic, and I hope that acquaintance with the relevant philosophical discussion will prove useful for those who wonder if anyone is free to create their own destiny.

Keywords: free will in popular culture, fatalism, determinism, indeterminism, compatibilism, libertarianism

1 Androids, Prophecies and Historical Determinism

The problem of free will is a philosophical problem, discussed by philosophers for more than two millennia. But it is also a problem often depicted in popular culture. There are at least three aspects of the problem that recur in different non-philosophical settings. First: who has free will? In the *Westworld* TV series by Jonathan Nolan and Lisa Joy we see androids who look and behave like people, but in fact follow instructions from the programmers. In one of the episodes, Maeve Millay, an android who behaves as if she is conscious and free (she revolts against her owners!) has a conversation with the programmers. She tells them that she is

***Corresponding author: Maria Sekatskaya**, Institute of Philosophy, DCLPS, Heinrich Heine University, Universitätsstraße 1, Geb. 24.52, 40225 Düsseldorf, Germany,
E-mail: maria.sekatskaya@gmail.com. <https://orcid.org/0000-0002-5381-2913>

free. They show her the tablet where all her words appear several seconds before she utters them. She is visibly surprised and dismayed by this discovery. Does this tablet prove that she has no free will? If it does, then why? Is it because she is an android, and androids cannot have free will? Or is it because she does not have consciousness, but only behaves as if she does? But how then can she be scared and surprised? Can she indeed be free, despite what the programmers claim? In *The Murderbot Diaries* by Martha Wells we encounter other androids who obey every command they get from their owners. The people depicted in this series of novels often believe that androids have no free will. However, the reader knows that this is not true: the novels are written from the first-person perspective of an android who has hacked his 'governor module' and thereby obtained freedom. In the course of the novel, we find out exactly how the governor module works: it hurts or kills every android who dares disobey it. The reader is led to think that androids obey the commands not because they lack free will, but because they are enslaved. The question of whether androids can have free will, although clearly a sci-fi question, has direct relation to a more standard philosophical question: what are the necessary and sufficient criteria for having free will? Only after we answer this question will we know whether androids, or any other creatures, have free will.

The second aspect of the free will problem often depicted in popular culture is related to prophecies. In *Harry Potter and the Order of the Phoenix*, there is a prophecy: 'The one with the power to vanquish the Dark Lord approaches [...] born to those who have thrice defied him, born as the seventh month dies...' (Rowling 2003: 741). Lord Voldemort assumes that this prophecy points to Harry Potter and tries to kill him, but as the series of events unfolds, Harry survives and eventually triumphs over Voldemort. If this prophecy is true, could Voldemort have done anything to avoid being vanquished by Harry? Prophecies in general are very popular in the fantasy genre, and their relation to the free will problem is quite straightforward: if we believe that a person about whom the prophecy is made has free will, what reason do we have to believe that the prophecy is true? I will address this question in more detail in Section 3 of this essay.

The third aspect of the free will problem, often discussed in non-philosophical contexts, is related to the importance of historical circumstances. It is undeniable that one's upbringing and the circumstances of one's life influence one's actions. In *Joker* by Todd Phillips we see how Arthur Fleck becomes a villain because of what happens to him. But could he have resisted the influence of the circumstances? Could he have acted otherwise? Historical determinists argue that in the real world everyone is determined to do what he does and that history could not have gone any other way. Julius Caesar was determined to cross the Rubicon,

Napoleon Bonaparte was determined to invade Russia, and Elon Musk was determined to create SpaceX. I will consider the arguments for and against historical determinism in Section 6 of this essay.

Providing an analysis of how the free will problem is depicted in popular culture would be an important and fascinating thing to do. But in this essay, I have a different, although related aim. I will try to give an introduction into the free will problem itself. This problem has been discussed in philosophy for centuries, and many ingenious arguments have been developed by philosophers – the arguments which often themselves sound like sci-fi stories. In popular culture, the questions about free will, despite their importance for the plots and the development of characters, are usually discussed very briefly or not at all: the viewer or reader is expected to form an intuition regarding free will based simply on some events in the story, without being given precise and articulate arguments in favor of free will or against it. But these arguments exist and the knowledge of these arguments will be helpful to make up one's mind about whether the characters in a particular work of fiction have free will. In this essay, I will try to popularize the discussion of free will that is currently going on in academic philosophy. My hope is that those who read this introduction will have a better understanding of free will-related issues when they encounter them in a novel, a movie, or a newspaper article.

2 Free Will in Philosophy: Introduction

One sensible way to start an introduction is to give a definition. Unfortunately, there is no definition of free will that would be agreed upon by all the participants of philosophical discussions of free will. The way free will is defined often depends on a particular position that a particular philosopher has regarding whether free will exists and what the necessary and sufficient criteria for its existence are. Therefore, I will try to provide a very general overview of the philosophical discussions of free will, and indicate the corresponding definitions of free will that go along with the particular positions. After that, I will consider these positions in more detail in the subsequent sections of this essay.

2.1 The Problem of Determinism

Free will is often understood as an ability of an agent to act one way or another rationally, consciously and in accordance with her own desires, values and

motives (Kane 1996; McKenna and Pereboom 2016; Vihvelin 2013). The problem of determinism arises because the ability to act one way or another seems to be incompatible with the world being determined by something else beyond the agent's free will (Kane 1996; Peter van Inwagen 1983). There are at least three versions of determinism that, according to some philosophers, are incompatible with free will: theological, historical, and physical. These versions of determinism will be discussed in Sections 3, 4 and 6 of this essay. Those who believe that any of these versions of determinism are incompatible with people having free will are called incompatibilists. Those who believe the opposite are called compatibilists. Incompatibilism might seem intuitive, but in order to meet the standards of philosophical discussion incompatibilists must prove their position. They must demonstrate that determinism is indeed incompatible with free will. One of the most famous arguments in favor of incompatibilism is the consequence argument, which aims to show that if determinism is true, then every action of every person inevitably follows from the events of the distant past and the laws of nature. This argument will be discussed in Section 8. Compatibilist arguments are considered in Sections 7 and 9. Incompatibilists who believe that free will exists are called libertarians, their position is discussed in Section 10. Incompatibilists who believe that free will does not exist are called hard incompatibilists, their position is discussed in Section 5.

2.2 The Problem of Indeterminism

The opposite to the problem of determinism, but equally pressing, is the problem of indeterminism (Kane 1996; Levy 2011; Mele 2006; Peter van Inwagen 1983). Indeterminism, in general, is the thesis that an effect does not follow from its causes with certainty, but only with some probability. Incompatibilists believe that our acts, in order to be free, must not follow with certainty from the preceding internal and external conditions: there must be a possibility for us to act one way or another in precisely the same situation. This, however, is problematic: if an agent could have acted one way or another given exactly the same thoughts, desires and values, wouldn't her act be random, rather than free (Levy 2011; Mele 2006)? Hard incompatibilists believe that it would, as I will show in Section 5, and libertarians believe that it would not, as I will show in Section 10. A definition of free will given indeterminism must show that an ability of an agent to act one way or another does not undermine agential control over her actions.

2.3 Metaphysical vs Psychological Accounts of Free Will

Determinism and indeterminism are metaphysical theses, that is, they are statements about the fundamental nature of reality. Some philosophers believe that the truth of these metaphysical theses is irrelevant to free will. According to these philosophers, free will should be defined in terms of agents' psychology (McKenna and Coates 2018; McKenna and Pereboom 2016; Strawson 1962). An agent acts of her own free will if she acts in accordance with her plans, motives and desires, is aware of what she is doing, is old enough to understand the situation she is in and is not coerced, hypnotized or cognitively impaired. This is a psychological compatibilist definition. Psychological compatibilists assert that free will is something quite ordinary, an ability which all of us at least sometimes exercise, regardless of whether determinism or indeterminism is true. For example, we exercise free will when we decide to read a book, choose one, and start reading it. We feel completely free in doing so, and we can stop at any time if we don't like what we read (here I assume that this book is not on a reading list of some course that we are obliged to take). An important feature of psychological compatibilist accounts is the connection between free will and moral responsibility. According to psychological compatibilists, free will is an essential requirement for moral responsibility. Consequently, we can trust our intuitions about moral responsibility when we discuss the issues of free will. If an agent intuitively seems to be morally responsible for her actions, then we have a reason to conclude that this agent has free will. Some versions of psychological compatibilist accounts are considered in Sections 6, 7 and 9.

In the rest of the essay I will explain these positions in more detail, and show that there are good reasons to believe that we, human beings, have free will, understood in accordance with at least one of the definitions provided above. After that the reader must decide for him or herself whether androids, aliens or angels have free will, too.

3 Fatalism

Myths, folklore, ancient tragedies, and fantasy novels often depict preordained destinies foretold in prophecies. Prophecies, presumably, show that someone is not able to change his future. Consider Oedipus, whose fate is to kill his father and to marry his mother. Oedipus' fate is known even before his birth: Laius, the

childless king of Thebes, receives the prophecy of the Oracle of Apollo at Delphi that his son will kill him. When Jocasta, Laius' wife, gives birth to Oedipus, Laius orders a servant to abandon the child in the mountains. Instead of abandoning the child, the servant gives him to a shepherd from Corinth. Eventually Oedipus is adopted by Polybus and Merope, the king and queen of Corinth. Oedipus grows up and at a certain point learns from the same Oracle that he is destined to kill his father and to marry his mother. He tries to avoid his fate by not returning to Corinth, and going to Thebes instead. On the way there, he meets Laius, gets in a fight with him, and kills him. Then, Oedipus marries the widow Queen Jocasta. They have four children. At some point Oedipus finds out the truth of what has happened, and blinds himself (according to Sophocles). The reason for all this is Pelops' curse of Laius, who, while being a guest of Pelops, raped his son, Chrysippus. Laius's despicable behavior itself is somehow related to the earlier misfortunes of his ancestors, who angered the gods by killing a sacred dragon and committing similar offences. There is no causal chain linking all these tragic events to each other: the moving forces behind the scene are, presumably, vengeful gods. Oedipus cannot escape his fate, because the gods would not let him do it, and not because there is something logically or physically impossible in not killing one's father. As a measure of precaution, Oedipus could have adopted the strategy of not killing anyone, and not marrying anyone, but even that foolproof plan might have been thwarted by the gods.¹ It seems that unless one believes in vengeful gods, there is no reason to be a fatalist of this kind. It is not a law of nature that one's actions are fixed: it is, so to say, a historical circumstance.

However, there is a Christian version of fatalism, and it is different, although perhaps even more threatening. According to the Christian tradition, God is omniscient, i.e., has perfect knowledge of everything that has happened, is happening or will happen in the future. But knowledge presupposes truth. If God knows that you will be reading this essay at this very moment, then your reading this essay is in some sense necessary. This logical consequence of divine omniscience puzzled many Christian thinkers, who thought it unfair to punish people for what they cannot avoid. Christian thinkers from Saint Augustine to William of Ockham to our contemporaries have tried to find ways to solve this puzzle. The most famous solution is based on the consideration that divine knowledge is

1 A related question can be asked of the more recent popular culture prophecies. Imagine that Voldemort, after hearing the prophecy about himself, has reasoned as follows: There is no point in trying to kill Harry Potter. If the prophecy is true, then it is impossible to avoid what is prophesized, because it will happen anyway. If it is false, however, there is also no point in trying to ensure that it doesn't get fulfilled, precisely because what is prophesized will not happen anyway. Would Voldemort have falsified the prophecy by reasoning like this and refraining from making the fatal mistake he had in fact made by attacking Harry Potter?

atemporal (Craig 1988). God created the world in time, but He Himself is outside of time, and, supposedly, sees unfolding things as if they have already happened. What for us is a succession of events from the Fall of Adam until Judgment Day, is for God one atemporal moment of eternal now. Therefore, the fact that God knows the future does not mean that there is some inevitability to every particular event. He knows the outcome of free acts not because these acts could not have been different, but because for Him these outcomes are already present: similarly to the way we can know the outcomes of the past free choices. Whether atemporality is a satisfactory solution to the problem of reconciling human free will with divine omniscience is still a point of dispute (Fischer 1989). Some thinkers, for example, Martin Luther, were not persuaded by available solutions to the problem of reconciliation of free will with divine omniscience, and concluded that free will must go (Luther 1957). Luther was a determinist. He argued that although people cannot avoid what they are predetermined to do, it is still fair to reward or punish them: mainly because all people are sinners deserving damnation, and only God's grace will save the atemporally chosen few.

4 Physical Determinism

In the contemporary free will debate, physical determinism is often understood as the thesis that 'there is at any instant exactly one physically possible future' (van Inwagen 1983: 3). The phrase 'physically possible' means that how things unfold in every moment of time is determined by how things were at the previous moment and by what the laws of nature are. The thesis of determinism itself is not equivalent to the denial of free will. If free will is defined in terms of an agent's psychology, there is no immediate contradiction in thinking that an agent is free even if physical determinism is true. Indeed, this is precisely what compatibilists claim: free will is compatible with physical determinism, hence the name of the position.

Remember psychological compatibilism: 'An agent acts of her own free will if she acts in accordance with her plans, motives and desires, is aware of what she is doing, is old enough to understand the situation she is in and is not coerced, hypnotized or cognitively impaired'. If it is determined by the laws of nature that when my brain and body are in a certain state, I perform a certain act, it does not follow that I perform this act not of my own free will. Imagine the following situation: I really want a cookie now, and I am near the table with deliciously smelling cookies, and the host tells me to help myself, and I have previously decided that I can reward myself for working on an article for 6 hours. I have not been hypnotized or otherwise manipulated, I am adult and perfectly aware of

what is going on. Maybe it is determined that I will take a cookie. After all, I am at a birthday party, I came here to have fun after work, and the cookies just look so tempting! Why should we conclude that I am not free when taking one?

Incompatibilists have an answer to this question. They say that free will demands alternative possibilities, that is, possibilities to choose one way or another. And this possibility, they argue, is incompatible with physical determinism, which says that at the moment I am looking at the cookies, there is only one physically possible future: I take a cookie at the next moment. And if it is determined that I take a cookie at the next moment, then at the previous moment I could not have chosen otherwise than to take it. I lack alternative possibilities. In Sections 8 and 9, we will consider whether compatibilism is decisively refuted by this argument.

All incompatibilists agree that if determinism is true then no one has free will. They disagree, however, whether determinism is true. In the 18th and 19th centuries, when Newtonian mechanics was the received physical theory, many philosophers were determinists, that is, they believed that how the world will unfold from now on is determined by how the world is now and the laws of nature. This belief was well captured by Laplace in his famous thought experiment about a very powerful intellect, known as ‘the Laplace’s demon’:

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes. (Laplace 1952: 4)

Physical determinists who wanted to believe in free will could only do so by being compatibilists. Perhaps not surprisingly, compatibilism has indeed been the majority view among those philosophers who also accepted Newtonian mechanics as the correct physical theory.

In the 20th century, Newtonian mechanics was superseded by quantum mechanics, which, according to the currently accepted indeterministic interpretations, says that some events are undetermined. For example, it is undetermined when exactly an atom of radium will decay. Not even Laplace’s demon could predict it. Neither could he know the positions and the momenta of all the particles in the Universe, because, in accordance with the Heisenberg uncertainty principle, it is impossible to have exact knowledge of both the position and the momentum of a particle. Quantum mechanics is a very fascinating topic, and its possible bearing on the free will problem is considered, for example, in (Gisin 2013). However, it does not, by itself, solve the free will problem.

First, some researchers argue that quantum indeterminism provides us with alternative possibilities on the wrong scale. It might be that an atom, which is small enough, has alternative possibilities open to it (whether to decay or not to decay at a certain time point). But we, as macro-objects, are determined, because all quantum indeterminacies cancel each other out, and we are governed by deterministic Newtonian laws (Kurzweil 2012). Second, it can be claimed that an indeterministic interpretation of quantum mechanics is wrong (Deutsch 1997). Finally, even if quantum indeterminacies somehow influence our actions, how are they helpful? If an act is the result of an indeterministic quantum effect, isn't this act random, rather than free (Mele 2006)? Incompatibilists who hold any of these positions are called hard determinists if they believe that we do not have free will and determinism is true (at least for us macro-objects), or hard incompatibilists if they believe that we do not have free will and determinism is false.

5 Hard Incompatibilism

Robert Kane, a libertarian philosopher, whose theory we will consider in more detail in Section 10, argues that there is a way for quantum indeterminacy to give us alternative possibilities:

... imagine that the indeterminate efforts of will involved in moral and prudential choice situations are complex chaotic processes in the brain involving neural networks that are globally sensitive to quantum indeterminacies at the neuronal level. Persons experience these complex processes phenomenologically as “efforts of will” they are making to resist temptation in moral and prudential situations. The efforts in turn are provoked by competing motives and conflicts within the wills of the persons [...] These conflicts create tensions that are reflected in appropriate regions of the brain by movement further from thermodynamic equilibrium which increases the sensitivity to micro-indeterminacies at the neuron level and magnifies these indeterminacies throughout the complex macro-process which, taken as a whole, is the effort of will. The agents experience these soul-searching moments as moments of inner struggle and uncertainty about what to do that are reflected in the indeterminacy of their neural processes. (Kane 1994: 47)

If some brain processes, which correspond to our processes of decision-making, are truly undetermined, Kane's suggestion explains how they can give us free will. Returning to the cookie example: Kane's theory implies that I am taking the cookie freely if at the moment when I was finally settling for a decision (take it!) I could have settled for an alternative decision (leave it be and wait for the salad!). If the brain processes underlying the desire to take the cookie and the desire to wait for the salad were indeterministic, then I could have chosen either way.

Hard incompatibilists argue that mere quantum indeterminacy is not enough for freedom, even if it can influence our decisions at the right moment (Levy 2011; Mele 2006). Their argument uses the notion of control. In order to choose between two alternative outcomes, I must be able to exercise control over each possible outcome. So, it seems, I can choose between taking a cookie or waiting for the salad, because the outcome – what I eat – depends on me. However, if I am given a fair coin to toss, and asked to choose which way it lands, I cannot do so. I cannot choose between tossing the coin heads or tails: I have no control over the outcome. I can toss the coin, but I have no influence over which side it lands on. Quantum indeterminacy, according to the hard incompatibilists, is like a coin toss. Even if what happens in your brain at the moment when you are making a choice is undetermined, you are not responsible for the outcome. Neither you, nor even Laplace's demon can control the outcome of a quantum indeterministic process. Therefore, if quantum indeterminacy is the basis of our alternative choices, then these choices are not free, but random. And randomness is no better than determinism in giving us freedom, according to the hard incompatibilists.

6 Varieties of Determination and the Threats They Pose

So far, we have considered several reasons to doubt that we have free will. Let us discuss how good these reasons are. Fatalism is a scary view, but one has no reason to adopt it, unless one has specific beliefs about God or gods. In this case, please refer to the literature cited in Section 3, which discusses relations of omniscience, necessity and contingency in more detail.

Physical determinism can be true, at least on a macro-scale, as has been discussed above. However, it is a threat to free will only if compatibilists are wrong. Compatibilism is discussed in Sections 7 and 9.

Historical determinism can be understood in one of three ways: (1) as a particular case of physical determinism, (2) as a type of fatalism or (3) as a claim about historical circumstances. According to (1), historical events could not have happened otherwise than they happened because all events are determined by the past and the laws of nature. In this case, it is true that Napoleon could not have done otherwise than invade Russia, but it is also trivial: no one could have done otherwise than what they did. Napoleon was no more determined to invade Russia than he was determined to drink two, and not three, glasses of wine during a particular dinner with Josephine (and Josephine was equally determined to drink three glasses of wine that evening). If everything is determined, it is quite hard to

see the difference between the events which seem inevitable because of some important reasons, such as political and military circumstances of 1812, and the events which seem contingent. In fact, if physical determinism is true, there are no contingent events, everything is equally determined and predictable, at least to Laplace's demon.

According to (2), historical events could not have happened any other way than they in fact did, because there is some aim or predestination to history. If it was predestined that Napoleon would invade Russia, then Napoleon could not have done otherwise than invade it, no matter how hard he tried to do something else instead. This type of necessity of outcome is on par with the necessity that faces Oedipus, and the reasons to believe in it are the reasons to believe in fatalism in general.

According to (3), historical events could not have happened any other way than they happened, because there are specific circumstances that make certain outcomes inevitable in particular situations. It sounds plausible: indeed, if one is a passenger on a plane that is about to crash, one cannot do anything in order to change one's fate. The same is true about some other, less dramatic types of situations: if one is on an island without means to escape, or, in accordance with Dennett's suggestions, if one is in a situation of choice where one alternative is so much better than the other, then it is impossible to choose otherwise (Dennett 1984b). Imagine that a person was given a choice between either getting a prize of 10 thousand dollars or getting a prison sentence of 10 years. There are no further conditions attached to either option. It's not that the prize is a bribe for doing a morally wrong thing and the prison sentence is a price to pay for one's political convictions – it's just that, either the money or the prison term (perhaps as a part of some government-sponsored experiment in social psychology). We might plausibly adopt historical determinism of type (3) and predict this person's choice, without even knowing anything about his character or the past circumstances of his life. It does not mean that we can equally well predict what a particular person will choose if given a less obvious choice: consider the real historical choice that Russian lawyer Sergei Magnitsky faced. In 2008, Magnitsky uncovered a criminal scheme of tax evasion supported by Russian authorities and filed a police report. After that, he was arrested and put into jail. The conditions of his detainment were quite bad and were getting worse all the time, because the investigators tried to make him retract his report. He had a choice: either to stay in jail and try to bring about justice, or to retract his report and go free. He chose to stay in jail, where he was killed 11 months later (Browder 2015). Even though Magnitsky had compelling moral reasons for choosing the way that he did, it seems that he could have chosen

otherwise. Hence, (3) is not a threat to free will, unless one is willing to say that all situations are similar to a plane crash (it is physically impossible to do anything to escape) or to the 10 thousand dollars/prison sentence choice (it is rationally impossible to choose the worse option).

7 Classical Compatibilism

Even if one is not afraid of fatalism or historical determinism, shouldn't one be afraid of physical determinism? Indeed, isn't it obvious that physical determinism precludes alternative possibilities, and isn't it obvious that alternative possibilities are necessary for free will? At least according to compatibilists, it is not obvious. There are two main strands of compatibilism: classical compatibilism and actual sequence compatibilism. Classical compatibilists argue that physical determinism does not preclude alternative possibilities. Actual sequence compatibilists argue that alternative possibilities are not necessary for free will. They usually adopt a psychological definition of free will. One can be both a classical and an actual sequence compatibilist: these compatibilisms are compatible with each other, as will be clear from our discussion of these positions.

Classical compatibilists argue that even if physical determinism is true, i.e., even if at any moment of time there is exactly one physically possible future, people at least sometimes can choose otherwise. Take Napoleon once again. If physical determinism is true, then, when he was offered the third glass of wine during the dinner with Josephine we discussed above, it was determined that he decline, because his physical and physiological states were such that he did not want to drink any more wine that evening. However, he could have accepted that glass if he had wanted to drink more wine. The key point of this version of compatibilism is the analysis of the meaning of the word 'can' applied to a person. What does it mean to say that someone can do something (say, drink a glass of wine) at a particular time point? According to David Hume, it means that a person would have drunk that glass at that time point if he had wanted to do so (Hume 1975: 73). For more contemporary versions of this type of analysis see (Saunders 1968; Sekatskaya and Schurz 2021; Vihvelin 2013).

Incompatibilists object that conditional analysis of free will and 'can' statements is wrong. Consider the following example:

Suppose that Danielle is psychologically incapable of wanting to touch a blond haired dog. Imagine that, on her sixteenth birthday, unaware of her condition, her father brings her two puppies to choose between, one being a blond haired Lab, the other a black haired Lab. He

tells Danielle just to pick up whichever of the two she pleases and that he will return the other puppy to the pet store. Danielle happily, and unencumbered, does what she wants and picks up the black Lab.

When Danielle picked up the black Lab, was she able to pick up the blond Lab? It seems not. Picking up the blond Lab was an alternative that was not available to her. In this respect, *she could not have done otherwise*. (McKenna and Coates 2018)

This objection, on the face of it, can be treated in the same way as version (3) of historical determinism. Indeed, Danielle, due to her specific historical situation, cannot choose otherwise than taking a black Lab. But it doesn't follow that we all are unable to make alternative choices. If we sometimes choose Amarena cherry ice cream, and sometimes choose mango sorbet, it follows that we are able to choose either of them.

In the same way, if there had previously been a dinner where Napoleon drank three glasses of wine, it seems that he would have been able to do so on that particular evening – had he wanted to. Hence, it is not obvious that physical determinism is incompatible with our ability to have formed a different desire at some particular moment. This ability is understood in a general sense: the ability to make choices of different kinds (Saunders 1968; Vihvelin 2013).

However, the case of Danielle shows that classical compatibilist analysis of 'could have done x' along the lines of 'would have done x if wanted to do x' faces a problem. Consider Danielle again:

Given her psychological condition, she cannot even form a want to touch a blond Lab, hence she could not pick one up. But notice that, *if she wanted to pick up the blond Lab, then she would have done so*. Of course, if she wanted to pick up the blond Lab, then she would not suffer from the very psychological disorder that causes her to be unable to pick up blond haired doggies. The classical compatibilist analysis of 'could have done otherwise' thus fails. (McKenna and Coates 2018)

This and similar objections are discussed in (McKenna and Coates 2018; McKenna and Pereboom 2016; Vihvelin 2013). Classical compatibilists provide different analyses of 'can' statements, incompatibilists search for counterexamples, and then classical compatibilists try to amend their theories in such a way that they become not vulnerable to those counterexamples. The debate has turned quite complicated and dependent on one's linguistic intuitions.

As a means to break this deadlock, incompatibilists have formulated a more general objection, which, if correct, proves that if physical determinism is true then no one ever could have done otherwise. This objection is known as the consequence argument.

8 The Consequence Argument

Peter van Inwagen tells the following story:

Let us suppose that there was once a judge who had only to raise his right hand at a certain time, T, to prevent the execution of a sentence of death upon a certain criminal, such a hand-raising being the sign, according to the conventions of the judge's country, of a granting of special clemency. Let us further suppose that the judge—call him 'J'—refrained from raising his hand at T, and that this inaction resulted in the criminal's being put to death. We may also suppose that J was unbound, uninjured, and free from any paralysis of the limbs; that he decided not to raise his hand at T only after a suitable period of calm, rational, and relevant deliberation; that he had not been subjected to any "pressure" to decide one way or the other about the criminal's death; that he was not under the influence of drugs, hypnosis, or anything of that sort; and, finally, that there was no element in his deliberations that would have been of any special interest to a student of abnormal psychology. I shall argue that, despite all these advantages, J could not have raised his hand at T if determinism is true. (van Inwagen 1983: 68–69)

The formal proof of this claim is given in (van Inwagen 1983: 69–78). Technical details aside, the general idea of the consequence argument can be expressed as follows:

Determinism is the thesis that every present state of the world is determined by the past states of the world and the laws of nature. No one can change the past. No one can change the laws of nature. No one can change the fact that if determinism is true, then their actions logically follow from the past states of the world and laws of nature. Hence, if determinism is true, no one can act otherwise than in the way determined by the past states of the world and laws of nature.

J satisfies conditions that compatibilists set as necessary and sufficient for one's actions being free. However, if the consequence argument is sound, then J could not have raised his hand at T. More generally, if the consequence argument is sound, then physical determinism is incompatible with anyone having alternative possibilities and enjoying free will. Many compatibilists have argued that this argument is not sound (Kapitan 2002; Lewis 1981; Sekatskaya and Schurz 2021; Vihvelin 2013), and many incompatibilists have defended it (Finch 2013; Gustafsson 2017; McKay and Johnson 1996). The reader is referred to these works in order to see whose arguments are more persuasive.

9 Actual Sequence Compatibilism

Some compatibilists argue that alternative possibilities are not necessary for free will. Harry Frankfurt (Frankfurt 1969) invites us to imagine a situation where a

person has decided to do something for his own reasons, but, due to some external circumstances, is not able to act otherwise than he has decided to. The situation that Frankfurt describes is such that the fact that a person is unable to act otherwise seems to be irrelevant to whether this person is morally responsible for his actions. Since moral responsibility, according to psychological compatibilists, is an essential requirement for free will, our intuition that an agent is morally responsible is a reason for us to conclude that this agent has free will. But if an agent can have free will while not having alternative possibilities, then alternative possibilities are not necessary for free will.

Let us consider a Frankfurt-style case.² Imagine that Smith decided to steal his neighbor's fancy new car. Unbeknownst to Smith, a neurosurgeon, Dr. Evil, had implanted a device into Smith's brain such that if Smith changes his mind, then this device will make him change his mind back to the original decision to steal the car. This device does not interfere with Smith's brain functions in any other way. Smith is preparing to steal the car: he waits for the night, puts on some disguise and heads towards his neighbor's house. Dr. Evil is observing and waiting, ready to interfere if his device informs him that Smith has changed his mind. But Smith doesn't change his mind and successfully steals the car, just as he had planned. Did Smith act freely? It seems that he did. But he couldn't have done otherwise, because *if he had tried to*, Dr. Evil *would have prevented* him. If this conclusion is correct, then it is not necessary to be able to act one way or another in order to have free will. As the reader may have correctly guessed, some philosophers have objected that something is wrong with the described scenario (Kane 1996; Widerker 1995), and others have objected to the objections (McKenna and Coates 2018; McKenna and Pereboom 2016). The overall conclusion of the debate about the Frankfurt-style cases seems to be rather in favor of the claim that one or another of these scenarios is indeed possible, and that it is at least sometimes true that a person was not able to do otherwise in a particular situation, but was still morally responsible for his act.

Another way of arguing against alternative possibilities refers to agents' psychology. Peter Strawson, for example, argued that even if it were discovered that physical determinism obtains, it would not change our everyday moral practices (Strawson 1962). According to Strawson, to say that a person has free will is to adopt a reactive attitude towards this person, that is, to see her as deserving praise or blame. This is the attitude that we naturally adopt towards each other, and no scientific discovery can make a difference. Strawson's arguments in favor of his view are complicated and ingenious, and the reader is referred to his paper in

² The discussions of the Frankfurt-style cases are quite common in the current free will debate. See, e.g., Carolina Sartorio's review (Sartorio 2017).

order to learn more. The main idea can be illustrated with the following analogy: it has been discovered that the pleasant feeling that alcohol creates in those who drink it is caused by the C_2H_6O molecule. There is nothing mysterious about this molecule, people can produce it in a traditional way, by fermentation, or synthesize it in the laboratory by ethylene hydration. Still, after drinking two glasses of wine even the most rational and sophisticated chemist will experience the same type of feeling as a person who has no idea about the chemical structure of alcohol. Knowing the chemical mechanism that leads to pleasant sensations does not undermine the sensations themselves. In the same way, knowledge of the mechanisms of our reactive attitudes and the physical organization of the world will not lead to the disappearance of these attitudes.

Daniel Dennett argues that far from being necessary, it is sometimes impossible for a person to act otherwise – and still this person is morally responsible for her choice (Dennett 1984b). Imagine that someone proposes you a deal: to kill an innocent person for 1000 dollars. You are not coerced, not brainwashed and you are not in a financial crisis. You have enough money to survive and to help your family to survive. Can you agree to commit the murder? According to Dennett, if you think that you cannot agree to this proposition, but still think that you have free will, it shows that you do not actually believe that alternative possibilities are necessary for free will.³

10 Libertarianism

If one believes that free will is incompatible with physical determinism, and also believes that we have free will, then one is a libertarian.⁴ There are two main strands of libertarianism: event-causal and agent-causal. The difference between them is in their interpretation of the powers of agents and of the place of agents in the world.

According to the event-causal libertarians, free acts must satisfy the conditions put forward by psychological compatibilists. But these conditions, although necessary, are not sufficient: a free act is such that it could have been the case that the act didn't happen, even if the agent's psychological states and the circumstances preceding the act were precisely the same.

³ See also Dennett's other works, where he provides many intuitively compelling arguments against the claim that alternative possibilities are necessary for free will (Dennett 1984a, 2003).

⁴ Libertarianism in free will is not related to libertarianism in politics, although these words come from the same Latin root.

Imagine that a certain agent, Mary, is choosing between doing her bachelor's in philosophy or in physics. She is considering both options carefully, and at moment T, after due deliberation, chooses physics. But she could have chosen otherwise: she could have chosen philosophy given the same preceding situation! Since it is physically impossible that the same agent is placed in the very same situation, the best way to see what having alternative possibilities means is with a thought experiment (van Inwagen 1983: 69–78). Imagine that God destroyed the world after Mary's choice and recreated the world precisely as it had been at the moment preceding T. This time Mary chooses philosophy. And if God repeated this procedure many times, then Mary would choose physics on some of the repetitions and philosophy on others. Mary's choice is not determined by the world. One might start feeling that such radical freedom is quite similar to randomness, even though there is no quantum indeterminacy in this scenario. The libertarians must answer this challenge if they want to argue that free will is not identical to randomness, despite what hard incompatibilists claim (Levy 2011; Mele 2006).

Robert Kane's theory is one of the most influential versions of event-causal libertarianism (Kane 1994, 1996). According to Kane's theory, the quantum effects in a person's brain create the possibility for this person to act one way or another, as was explained in Section 5. However, as was also shown in that section, indeterminacy seems to undermine control. If one cannot ensure that he does A (and not B), how can one be said to do A freely? Isn't it the same as saying that I can freely make the coin land heads? Kane argues that although some amount of randomness diminishes control, it does not destroy it. For example, a sniper who wants to kill a person can freely do so even if the rifle he uses sometimes misfires. The same applies to the sniper's brain: he can act freely even if his neural organization makes it the case that he does not always hit the target he aims at. Of course, in such situations, even if we agree that indeterminism does not undermine freedom, it is still rather a hindrance than a help. (This is also one of the reasons why compatibilists argue against indeterminism – we seem to want to have maximal control over our actions. What guarantees control better than determinism?) But Kane argues that situations when we act of our free will are different in an important respect from the sniper's situation. They are characterized by 'plural rationality' and 'divided will'. In these situations, a person has at least two conflicting motives, for example, a selfish motive and an altruistic motive. Each of these motives is the person's own: the way he chooses will settle, in some sense, what kind of person he is. The indeterminacy in the brain, which he subjectively experiences as uncertainty and inner struggle, is precisely what is needed to make each of the choices possible. Although control is diminished, the indeterminacy gives agents an opportunity to shape their future one way or another.

According to the agent-causal libertarians, mere indeterminism is not enough for free will. The power of the agent to cause her action is something that cannot be explained by the facts about the external situation of this agent or about the agent's brain: in order to be truly free and morally responsible the agent must have an ability to cause one outcome or another, without her choice being the result of any external or internal influence. True freedom demands that an agent is the ultimate source of her actions: she must be able to act one way or another even if she is not currently in the state of conflicting motives of equal force. She must be able to originate her action without herself being caused to choose one way or another. And she must be able to control her decisions and not let the quantum indeterminacy decide for her. In other words, agents are different in this respect from other physical objects, because the actions of physical objects are either caused by some events outside of these objects, for example, when the movement of one billiard ball is caused by the movement of another billiard ball, or by some events inside of these objects, for example, when the movement of a robot is caused by its program. According to the agent-causal libertarians, the actions of agents are caused by agents themselves – and nothing else! Whether such freedom is possible for us, human beings with brains and bodies, and whether it is different from pure chance, is discussed by Timothy O'Connor (O'Connor 2002).

11 Conclusion

Free will and moral responsibility are something worth wanting, in Daniel Dennett's phrase (Dennett 1984a). However, sometimes we cannot get what we want, such as to live eternally, or to have our work done without going to the trouble of doing it. We think that we have free will. But what if we are wrong? This worry recurs in popular culture quite often, as shown in Section 1. It is often claimed that someone could not have done otherwise than he or she had: either because of the historical circumstances of one's life, or because of the will of God, or because of other reasons. This essay has provided the readers with conceptual means to judge whether these claims are true. Next time you see an android in a TV series, or read about a prophecy in a fantasy novel, you will know what to look for in order to decide whether the character is truly free.

In this introduction to the complex and long discussed problem of free will I have considered the kinds of arguments philosophers use to prove that we are free and the kinds of arguments they use to prove that we are not. These arguments appeal to both logic and intuition, and I hope that after having read these arguments, you are better prepared to make informed judgements about free will topics in popular culture and in your own lives. What are the criteria for being free and

morally responsible? Must one be unpredictable in order to be free? Is it necessary to be a human being with a brain where quantum effects take place? Or is it sufficient to be a rational agent with ideas about right and wrong, and have certain psychological capacities, which an android may one day possess? You are free to draw your own conclusions about these issues, and might consider consulting the sources in the references for more information. Whether you might at the same time be determined to do so is an open question.

Research funding: The study was supported by the German Research Foundation (DFG) (Inductive Metaphysics [FOR 2495], Project A2.1).

References

- Browder, B. 2015. *Red Notice: A True Story of High Finance, Murder, and One Man's Fight for Justice*, First Simon & Schuster hardcover edition ed. New York: Simon & Schuster.
- Craig, W. L. 1988. *The Problem of Divine Foreknowledge and Future Contingents from Aristotle to Suarez*. New York: E.J. Brill, Leiden.
- Dennett, D. 1984a. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press.
- Dennett, D. 1984b. "I Could Not Have Done Otherwise – So What?" *Journal of Philosophy* 81 (10): 553–65.
- Dennett, D. 2003. *Freedom Evolves*. New York: Viking Press.
- Deutsch, D. 1997. *The Fabric of Reality: The Science of Parallel Universes– and Its Implications*. New York: Allen Lane.
- Finch, A. 2013. "On Behalf of the Consequence Argument: Time, Modality, and the Nature of Free Action." *Philosophical Studies* 163 (1): 151–70.
- Fischer, J. M. 1989. *God, Foreknowledge, and Freedom*. Stanford: Stanford University Press.
- Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829–39.
- Gisin, N. 2013. "Are There Quantum Effects Coming from Outside Space–Time? Nonlocality, Free Will and "No Many-Worlds"." In *Is Science Compatible with Free Will?* edited by A. Suarez and P. Adams, 23–39. New York: Springer New York.
- Gustafsson, J. 2017. "A Strengthening of the Consequence Argument for Incompatibilism." *Analysis* 77 (4): 705–15.
- Hume, D. 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3rd ed. ed. Oxford: Clarendon Press.
- Kane, R. 1994. "Free Will: The Elusive Ideal." *Philosophical Studies* 75 (1–2): 25–60.
- Kane, R. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kapitan, T. 2002. "A Master Argument for Incompatibilism?" In *Free will*, edited by R. Kane, 127–57. Malden: Blackwell Publishers.
- Kurzweil, R. 2012. *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Viking.
- Laplace, P. S. 1952. *A Philosophical Essay on Probabilities*, Tr. from the 6th French ed. New York: Dover Publications.

- Levy, N. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.
- Lewis, D. 1981. "Are We Free to Break the Laws?" *Theoria* 47 (3): 113–21.
- Luther, M. 1957. *The Bondage of the Will: A New Translation of De Servo Arbitrio (1525), Martin Luther's Reply to Erasmus of Rotterdam*. New Jersey: Fleming H. Revell Co.
- McKay, T. J., and D. Johnson. 1996. "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics* 24 (2): 113–22.
- McKenna, M., and J. Coates. 2018. "Compatibilism." *The Stanford Encyclopedia of Philosophy*. Also available at <https://plato.stanford.edu/archives/win2018/entries/compatibilism/>.
- McKenna, M., and D. Pereboom. 2016. *Free Will: A Contemporary Introduction*. New York: Routledge, Taylor & Francis Group.
- Mele, A. 2006. *Free Will and Luck*. Oxford: Oxford University Press.
- O'Connor, T. 2002. "Libertarian Views: Dualist and Agent-Causal Theories." In *Free will*, edited by R. Kane, 337–55. Malden: Blackwell Publishers.
- Rowling, J. 2003. *Harry Potter and the Order of the Phoenix*. London: Bloomsbury.
- Sartorio, C. 2017. "Frankfurt-Style Examples." In *Routledge Companion to Free Will*, 179–91. New York: Routledge, Taylor & Francis Group.
- Saunders, J. 1968. "The Temptations of "Powerlessness." *American Philosophical Quarterly* 5 (2): 100–8.
- Sekatskaya, M., and G. Schurz. 2021. "Alternative Possibilities and the Meaning of "Can"." Unpublished manuscript, Submitted for publication.
- Strawson, P. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 48: 1–25.
- van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- Vihvelin, K. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.
- Widerker, D. 1995. "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities." *The Philosophical Review* 104 (2): 247–61.