

The idea of “free will” corresponds to a concept of being able to make meaningful choices between some group of options. It is our ability to be creative, and unpredictable. It is our ability to shape the future, in large ways (which Colleges should I apply for?) and small (what should I eat for lunch?). Furthermore, it implies that whatever choice is made, we had the ability to choose something else instead. “When an agent acts freely (or with free will), she is able to do other than what she does then.”<sup>[1]</sup> Others were dissatisfied with this interpretation, and have reinterpreted the idea of “free will” to instead correspond to the ability to not be constrained from making choices. If you are tied up, and an apple is set in front of you, it doesn't make sense to argue about whether or not you can choose to eat the apple. As Thomas Hobbes wrote, “a free agent is that which, when all things are present which are needful to produce the effect, can nevertheless not produce it, implies a contradiction and is nonsense; being as much as to say the cause may be sufficient, that is necessary, and yet the effect shall not follow.”<sup>[2]</sup>

The problem of free will exists in several forms. Baron D'Holbach tells us “Man's life is a line that Nature commands him to describe upon the surface of the earth: without his ever being able to swerve from it even for an instant.”<sup>[3]</sup> His argument is that of “hard determinism”. It starts from a physicalist position; everything is made of physical matter. Physical matter follows natural laws. Natural laws are invariant. The laws governing thermodynamics, gravity, chemical reactions, electromagnetism, and other properties of the natural world don't change when we look around us, be that locally on our own planet, or deep into the dark of space. If everything is made of physical matter, that implies that people are made of physical matter, too, being a member of everything. If people are made of physical matter, people are governed by invariant natural laws. This governance includes the choosy bits of people, whichever components of a person result in decision-making capacity, since

those bits too must be physical matter. The state of the choosey bits of a person, being physical matter following invariant natural laws, must be part of an unbroken chain of cause and effect, winding all the way back to the dawn of the universe. This means that the future states of the choosey bits of a person are yet more links in this causal chain, following invariant natural laws. Our choice of lunch or college is simply the next inescapable step in the continuous sequence of cause-and-effect. Whatever we choose is what we would always choose. We can do no other.

A similar idea from another direction is an argument from theology. This argument presumes a god with omniscience. Omniscience can be thought of as knowing everything knowable, necessarily including future events. An omniscient god knows, by definition, everything that will happen. This very knowledge goes against the concept of free will, since what can it mean to make a choice when the outcome is known in advance. “Pre-determinism implies that all the information in the universe today was implicit in the earliest moments of the universe. It is information conserving. It is consistent with the theological idea of God's foreknowledge.”<sup>[4]</sup> This sort of theological pre-determinism has various implications for several aspects of theology itself, beyond the topic of free will, but they are beyond the scope of this paper.

In both cases, the hard determinism and the theological determinism, the consequences of the problem of free will can be easily stated: Either determinism is true, our actions are determined, be that by physics, or by God, or some other mechanism, and free will does not exist, or indeterminism is true, our actions are therefore random, and having responsibility for our actions no longer makes sense.<sup>[5]</sup>

One attempt at a solution is relying on advancements in physics, namely quantum mechanics. The Stanford Philosophy Encyclopedia says “QM is widely thought to be a strongly non-deterministic theory. Popular belief (even among most physicists) holds that phenomena such as radioactive decay, photon emission and absorption, and many others are such that only a *probabilistic* description of them can be given.”<sup>[6]</sup> This runs into a few different problems. Firstly, some interpretations of quantum mechanics result in likely candidates for deterministic physics. “The fundamental law at the heart of

non-relativistic QM is the Schrödinger equation. The evolution of a wave function describing a physical system under this equation is normally taken to be perfectly deterministic.”<sup>[7]</sup> If it holds that QM is deterministic in nature, then nothing is different from the original determinism claim of D'Holbach. Our choosey bits are still governed by natural laws, and those laws still imply a causal chain, and therefore a determined source other than ourselves, for our will.

If QM is indeterministic, though, we don't really gain much back. QM may be stochastic in nature, but it is still predictable. “[T]he ordinary deterministic evolution of the Schrödinger equation is all that ever happens. [...] [We] are not described by the entire quantum state; rather, we live in one branch of the wave function, which also includes numerous other branches where different outcomes were observed. This approach [...] restores determinism at the level of the fundamental equations, but sacrifices it for the observational predictions made by real observers.”<sup>[8]</sup>

What about a bit of wiggle room, some bit of space in the indeterminism of QM where free will can emerge, and we just haven't found the mechanism yet? “. Quantum mechanics doesn't say 'we don't know what's going to happen, but maybe our ineffable spirit energies are secretly making the choices'; it says 'the probability of an outcome is the modulus squared of the quantum amplitude,' full stop. Just because there are probabilities doesn't mean there is room for free will in that sense.”<sup>[9]</sup>

If we start from the premise of hard determinism, with our free will completely curtailed by the invariant laws of physics, quantum mechanics doesn't give our free will back. Either QM is deterministic or it is indeterministic, but either way we're back where we started with D'Holbach.

Let us flee, then, momentarily away from hard determinism and look again at the theology angle. What about the thorny issue where God's omniscience predetermines our choices, and rids us of free will in the process? I've come across three broad answers to this problem.

In the first, God is omniscient in a fashion such that we can't understand it and that it is compatible with free will, and also we have free will. This view can be seen from Maimonides, who wrote “One might ask: Since The Holy One, blessed be He, knows everything that will occur before it

comes to pass, does He or does He not know whether a person will be righteous or wicked? [...] [We] do not have the potential to conceive how The Holy One, blessed be He, knows all the creations and their deeds. However, this is known without any doubt: That man's actions are in his [own] hands and The Holy One, blessed be He, does not lead him [in a particular direction] or decree that he do anything.”<sup>[10]</sup> This isn't particularly satisfying, because it sidesteps the question, and hides it behind a stop sign called divine mystery.

In the second, God is not omniscient, leaving room for free will. “The sublime thought of God, he says, embraces all the cosmic laws which regulate the evolutions of nature, the general influences exercised by the celestial bodies on the sublunary world, and the specific essences with which matter is invested; **but sublunary events, the multifarious details of the phenomenal world, are hidden from His spirit.** [emphasis added] Not to know these details, however, is not imperfection, because in knowing the universal conditions of things, He knows that which is essential, and consequently good, in the individual.”<sup>[11]</sup> In this case, God can be surprised, which might raise other theological issues, but does seem to leave room for free will.

In the third, the bullet is bitten, God is fully omniscient, everything is predetermined, and free will doesn't exist. “Freedom is an illusion that arises from human ignorance of divine cause and necessity. All that individuals do is actually determined and predetermined. God both pre-knows and pre-causes all that occurs.”<sup>[12]</sup> I've begun to run out of time for research, at this point, and have not been able to find an adequate source for that point of view. As far as I can tell, it may have a representation in Calvinism, which matches the above source. It also matches my own background trivia knowledge, but I can't actually find anything in my limited searching to support it. As such, I mention the option as one available, but don't offer it with the same degree of confidence.

The theological views that I've come across vary wildly, and all point back to similar source material for their evidence. The different arguments seem equally valid, from my distant naive view,

and I am unable to come up with any thought experiments that would make the world around me different if one were true and the other not.

As such, I move on from the theological angle, and explore a compatibilist angle on this topic of free will from a reductionist point of view. Reductionism is the view that there aren't irreducible complex things. There aren't some things that run on one layer of reality, and other things with some deeper connection. “the reductionist thesis is that we use multi-level models for computational reasons, but physical reality has only a single level.”<sup>[13]</sup> The Stanford Encyclopedia of Philosophy has a nice definition for Compatibilism, which is “[T]he thesis that free will is compatible with determinism. Because free will is typically taken to be a necessary condition of moral responsibility, compatibilism is sometimes expressed as a thesis about the compatibility between moral responsibility and determinism.”<sup>[14]</sup>

The particular flavor of reductionist argument about free will that I've come across is structured approximately as follows. There are lots of different definitions about free will, and arguments about free will tend to get bogged down when people argue about “free will” but have differing interpretations in mind when they do. The classic example of “if a tree falls in a deserted forest, does it make a sound” epitomizes this. One person will say “no, for there is no one around to hear”, while another will say “yes, for the air vibrates with the sound of the fall and crashing of the tree”, and they will disagree. The thing is, the people agree about the actual experiences. The first person would agree that if a microphone resided in the forest, it would detect the vibrations; the second person would agree that there is nobody around to sense the crashing. The two people, arguing about what sound means, agree on all the facts, and expect the same things about experiences and microphones. There is no question to be answered, just a bit of confusion that dissolves away.<sup>[15]</sup>

So, the question of free will gets bogged down similarly, and maybe it dissolves away with a different angle of approach to the question.

The approach starts at reasoning about causality. The loose idea is that causality can be considered without having a time component. “Events do not happen in the *past* or the *present* or the *future*, they just *are*. But there may be a certain... *asymmetric locality of relatedness*... that preserves "cause" and "effect", and with it, "therefore". A point in configuration space would never be "past" or "present" or "future", nor would it have a "time" coordinate, but it might be "cause" or "effect" to another point in configuration space.”<sup>[16]</sup> Computing a universe with an initial condition of high entropy, and ending in a condition of low entropy would require computing lots of universes, and throwing most of them away. Computing a universe with an initial condition of low entropy, and ending in a condition of high entropy, however, could be done locally with a single computed universe. This points toward a direction for causality, regardless of time.

If causality is a thing that exists without necessarily having time, we're back to the future being determined, with D'Holbach once more. From the reductionist standpoint, though, it is a mistake to say that because the future is determined people cannot affect the future. People are not at the same level as physics, with both people and physics pointing toward the future, and physics leaving no room for people to steer. Physics surrounds both the people and the future alike. “Beliefs, desires, emotions, morals, goals, imaginations, anticipations, sensory perceptions, fleeting wishes, ideals, temptations... You might call this the "surface layer" of the mind, the parts-of-self that people can see even without science. If I say, "It is not *you* who determines the future, it is your *desires, plans, and actions* that determine the future", you can readily see the part-whole relations. It is immediately visible, like fingers making up a hand. There are other part-whole relations all the way down to physics, but they are not immediately visible.”<sup>[17]</sup>

From this view, free will isn't something that is squashed by physics, it is something that requires physics. How can thoughts, decisions, and reasoning exist in complete chaos? Something structured has to exist to house the mind-stuff. The reductionist view of “no irreducible complex stuff” applies to thoughts, decisions, and reasoning. Those are too complex to be irreducible primitives, so

they're probably made up of more primitive things, like the different heuristics and methods that the brain uses to process stuff. These smaller things themselves are made up of stuff, and so on and so on, down to atoms, quarks, and elementary fields.

If the future is determined from the past, this requires computing the present along the way. If you were to do so, fully computing the universe from some past point, moving forward, you'd hit the present and keep on going, but none of the causal relationships within your computation change by doing so. “The future is determined from the past” is just as true as “The future is determined from the present”.

Possibility, like probability, is in the mind, and not some feature of the external world. With probability, this is pretty straightforward to see. If person A walks into a room and a coin is flipped 5 times, with the result that each time it is flipped it comes up heads, this person will estimate the chances of the next flip being heads, using Bayesian probability, as leaning towards coming up heads again. If a second person, B, walks into the same room, between the 5<sup>th</sup> and 6<sup>th</sup> flip, and do their own estimation with out input from A on past results, they'll estimate something closer to even odds that the coin will come up heads. Neither of their estimates are a property of the coin itself, the probabilities are properties of the people.

Possibility is similar. “What concrete state of the world - which quarks in which positions - corresponds to "There are three apples on the table, and there *could* be four apples on the table"? Having trouble answering that? Next, say how that world-state is different from "There are three apples on the table, and there *couldn't* be four apples on the table." And then it's even more trouble, if you try to describe could-ness in a world in which there are no agents, just apples and tables.”<sup>[18]</sup> This example is what it feels like to have found a chain of actions which, if acted upon, takes you from the current state of the world to a “could” state.

I could eat pizza by spending 20\$ if I wanted. This “if I wanted” hides a preference ordering, and to get a preference ordering, I need to consider the consequences, the results. If I spend 20\$ I'll get

to eat pizza. If I save 20\$, I'll be able to make rent next month. Only after I've considered the consequences can I decide whether or not to spend or save. This leans on physics being required for free will, again. "You can only decide by going through an intermediate state where you do not yet *know* what you will decide. But the map is not the territory. It is not required that the laws of physics be random about that which you do not know. Indeed, if you were to decide randomly, then you could scarcely be said to be in "control". To determine your decision, you need to be in a lawful world."<sup>[19]</sup>

This reductionist view of free will also speaks to the moral responsibility side of the problem. Moral responsibility lies with the person, because the person is making the choices. It doesn't matter if I'm created by aliens, or a product of evolution, or popped into existence thirty-seconds ago with the world a simulation running in some 5th-dimension dollar store hardware. If any of those create me, but are then hands-off, the choice is mine, and the responsibility as well.

I don't know what my future decisions will be, I have to consider them first. This consideration leads to weighting the consequences, and deciding on a preferred outcome. Responsibility lies in aligning my preferences with my desires and goals in my mind, all of which are a part of physics. "If you are going to talk about moral responsibility, it ought not to be affected by anything that plays no role in your brain. It shouldn't matter whether I came into existence as a result of natural selection, or whether an alien built me up from scratch five minutes ago, *presuming that the result is physically identical*. I, at least, regard myself as having moral responsibility. I am responsible here and now; not knowing my future decisions, I must determine them. What difference does the alien in my past make, if the past is screened off from my present?"<sup>[20]</sup>

This view is satisfying to me, and I think that makes it harder for me to critique it successfully. Moral responsibility (not what is right or wrong, but whether I'm responsible either way) stemming from the consequences of a choice to turn a hypothetical among hypotheticals into an action among counterfactuals seems to align nicely with what it feels like in my head when I think "moral responsibility". Free will as a consequence of physics, our goals, desires, and preferences being made



of physics, not trodden over by physics, stemming out of ideas of computation, and causality, it scratches an itch in a very satisfying way. The problems of D'Holbach and theology seem adequately explained away here. Nothing is papered over, the question is dissolved.

I'm now here, at the end of my paper, dipping into a-page-too-far, past the required upper limit. So, I conclude with this: The problem of free will is old, and there is a tremendous amount of writing about it from many many angles. I've skimmed dozens more articles, blogs, encyclopedias and books than I've managed to cite as references here, trying to wrap my head around the myriad views on the matter. All I've ended up with, it seems, is more questions than answers, and an appreciation for the amount of thought and effort that goes into philosophical work. That seems like a decent outcome, even if the intermediate product, this paper, is a little aimless as a result.

References:

- [1] [https://books.google.ca/books?id=7V6n6-vobcIC&printsec=frontcover&dq=libertarian+accounts+of+free+will&source=bl&ots=MRp0nVdZyv&sig=v3OCzCbi\\_g1iGtaczI2SM9ultHo&hl=en&sa=X&ei=1kpUGiqlowCxc-AkAo#v=onepage&q&f=false](https://books.google.ca/books?id=7V6n6-vobcIC&printsec=frontcover&dq=libertarian+accounts+of+free+will&source=bl&ots=MRp0nVdZyv&sig=v3OCzCbi_g1iGtaczI2SM9ultHo&hl=en&sa=X&ei=1kpUGiqlowCxc-AkAo#v=onepage&q&f=false) “Libertarian Accounts of Free Will”, Randolph Clarke. Chapter 1, “Incompatibilism.”
- [2] [http://www.informationphilosopher.com/solutions/philosophers/hobbes/of\\_liberty\\_and\\_necessity.html](http://www.informationphilosopher.com/solutions/philosophers/hobbes/of_liberty_and_necessity.html) “Of Liberty and Necessity”, Thomas Hobbes. Section 32.
- [3] <http://www.philosophy-index.com/d-holbach/system-nature/i-xi.php> “The System of Nature”, Baron D'Holbach. Volume 1, chapter 9, “Of the System of Man's free-agency.”
- [4] <http://www.informationphilosopher.com/freedom/pre-determinism.html> “Pre-Determinism”, second paragraph.
- [5] <http://www.informationphilosopher.com/freedom/> Under the heading “The Standard Argument Against Free Will”
- [6] <https://plato.stanford.edu/entries/determinism-causal/#QuaMec> “Causal Determinism”, Carl Hoefer. Section 4.4, paragraph 1. “Quantum Mechanics”
- [7] <https://plato.stanford.edu/entries/determinism-causal/#QuaMec> “Causal Determinism”, Carl Hoefer. Section 4.4, paragraph 2. “Quantum Mechanics”
- [8] <http://blogs.discovermagazine.com/cosmicvariance/2011/12/05/on-determinism/> “On Determinism”, Sean Carroll. 5<sup>th</sup> bullet point, last paragraph.
- [9] <http://blogs.discovermagazine.com/cosmicvariance/2011/12/05/on-determinism/> “On Determinism”, Sean Carroll. End of 5<sup>th</sup> paragraph, starting from after the bullet list.
- [10] [http://www.chabad.org/library/article\\_cdo/aid/911903/jewish/Teshuvah-Chapter-Five.htm](http://www.chabad.org/library/article_cdo/aid/911903/jewish/Teshuvah-Chapter-Five.htm) “Mishneh Torah”, Maimonides. Sefer Mada, Teshuvah, chapter 5, section 5.

- [11] <http://www.jewishencyclopedia.com/articles/9813-levi-ben-gershon#anchor8> “Levi Ben Gershon”, M. Seligsohn, Kaufmann Kohler, Isaac Broydé. “God's Omniscience”, 3<sup>rd</sup> paragraph.
- [12] [https://www.fairmormon.org/answers/Mormonism\\_and\\_the\\_nature\\_of\\_God/Foreknowledge](https://www.fairmormon.org/answers/Mormonism_and_the_nature_of_God/Foreknowledge) “Does Jesus already know who will be returning to God's presence?”, FairMormon volunteers.
- [13] <https://intelligence.org/rationality-ai-zombies/> “Rationality: From AI to Zombies”, Eliezer Yudkowsky. “Hands vs. Fingers”, page 983, second paragraph.
- [14] <https://plato.stanford.edu/entries/compatibilism/> “Stanford Encyclopedia of Philosophy”, Michael McKenna, D. Justin Coates. “Compatibilism”, opening paragraph.
- [15] <https://intelligence.org/rationality-ai-zombies/> “Rationality: From AI to Zombies”, Eliezer Yudkowsky. “Disputing Definitions”, page 716.
- [16] [http://lesswrong.com/lw/qv/timeless\\_causality/](http://lesswrong.com/lw/qv/timeless_causality/) “Less Wrong”, Eliezer Yudkowsky, “Timeless Causality”, 5 paragraphs from the bottom.
- [17] [http://lesswrong.com/lw/r0/thou\\_art\\_physics/](http://lesswrong.com/lw/r0/thou_art_physics/) “Less Wrong”, Eliezer Yudkowsky, “Thou Art Physics”, approximately halfway down the post.
- [18] [http://lesswrong.com/lw/rb/possibility\\_and\\_couldness/](http://lesswrong.com/lw/rb/possibility_and_couldness/) “Less Wrong”, Eliezer Yudkowsky, “Possibility and Couldness”, approximately three-quarters down the post.
- [19] [http://lesswrong.com/lw/rb/possibility\\_and\\_couldness/](http://lesswrong.com/lw/rb/possibility_and_couldness/) “Less Wrong”, Eliezer Yudkowsky, “Possibility and Couldness”, 4 paragraphs from the bottom of the post.
- [20] [http://lesswrong.com/lw/ra/causality\\_and\\_moral\\_responsibility/](http://lesswrong.com/lw/ra/causality_and_moral_responsibility/) “Less Wrong”, Eliezer Yudkowsky, “Causality and Moral Responsibility”, 4 paragraphs from the bottom of the post.