

The Consequence Argument Meets the Mentaculus

According to Kant

“it is as impossible for the subtlest philosophy as for the commonest reasoning to argue freedom away. Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the same human actions, for it cannot give up the idea of nature any more than that of freedom. Hence even if we should never be able to conceive how freedom is possible, at least this apparent contradiction must be convincingly eradicated. For if the thought of freedom contradicts itself or nature ... it would have to be surrendered in competition with natural necessity.” (Fundamental Principles of the Metaphysics of Morals, 75-6)

Kant feared that freedom may be in conflict with what physics says are the laws that govern the world. His fears are supported by well-known arguments that aim to show that if all events including our decisions and actions conform to deterministic dynamical laws there is no room for free will. Chief among these is the Consequence Argument. Since Peter van Inwagen’s famous formulation and defense of the argument a large literature has grown up in response. In this paper I will develop a new response that is based on a part of physics that supplements the deterministic dynamical laws in order to account for macroscopic and in particular thermodynamic phenomena.¹ If this response works, then it will turn out that physics (together with some philosophy) is able to rescue freedom from the threat that it and the Consequence Argument seemed to pose.

Determinism means that the fundamental dynamical laws and the entire fundamental state of the universe at t (and that it is the entire state) entail the states at all times. So, for example, classical mechanics understood as a complete account of the motions of fundamental material objects is, as Laplace famously observed, deterministic. In rough form the Consequence Argument goes like this:

- 1) Determinism: The laws and past entail the future
- 2) The past is not up to me
- 3) The fundamental laws are not up to me

Therefore

- 4) The future is not up to me

The upshot of the argument is that if determinism is true then the future is not up to me; since my exercise of free will requires that at least to some extent the future up to me I lack free will.

In classical mechanics the state of the world at t consists in the positions (or relative positions) and the momenta of all the particles and the laws are the dynamical laws of classical mechanics (e.g.

¹ The response to the Consequence Argument I will defend is similar to approaches by Hoefer (2002), Loewer (2007) Vihvelin (2017), Beebe (2013) and Dorr (2016).

Hamilton's equations). Since the position of my hands at t_1 supervenes on the state of the universe at that time it follows that the state of the universe at times before my birth entail the positions of my hands at t_1 . This seems to mean that the position of my hands at t are not up to me since it had already been "decided" before my birth.

In discussions of free will it is often pointed out that classical mechanics has been superseded by quantum mechanics as the fundamental theory of our world and since quantum mechanics predictions are probabilistic not deterministic the conflict between physics and freewill is moot. But this is not so. While quantum theory as presented in physics textbooks is a fantastically accurate instrument for making probabilistic predictions its ontology and the dynamics it obeys are notoriously obscure. The usual textbook account (the so-called Copenhagen interpretation) makes reference to "measurement" in formulating its laws without a clear account of which physical interactions qualify as measurements. There are proposals for formulations of quantum mechanics that avoid this "measurement problem" but the best of these are as deterministic as classical mechanics.² There are theories that are modifications of quantum mechanics -so called "spontaneous collapse theories" like GRW that do have indeterministic dynamics. While these theories are not subject to the Consequence argument they pose other challenges for free will and, in any case we don't want the existence of free will to be hostage to the interpretations of quantum mechanics.³ Here I will restrict attention to the argument that claims to show that free will is incompatible with determinism and pretend that classical mechanics is true.

It is generally accepted that freewill requires an ability to "do otherwise" and this requires that one can decide between alternative actions that are open or possible -in some sense of "possible"-at t . There are differing views about the pertinent sense of "possibility." If one thinks that that the relevant notion of possibility is physical possibility at t -i.e. compatibility with the state at t and the laws- then the incompatibility of free will and determinism is immediate. No further argument is needed. There is only one decision that is possible at t so I can't do otherwise. The point of the consequence argument is to show that even if one doesn't appeal to the requirement that alternative actions are physically possible at t determinism is incompatible with free will. Because of this I will understand the relevant notion of possibility to be weaker, for example metaphysical possibility, and take the argument as attempting to establish that even on this understanding determinism and free will are incompatible.

My initial formulation of the Consequence Argument assumed that if the past and laws are not up to me then neither is the future. This is an instance of the principle that if A is not up to me and B is not up to me and $A \& B$ entail C then C is not up to me. This principle has apparent counterexamples and the phrase "up to me" is not completely clear. For this reason, I will discuss a version of the argument based on the notion of "influence." Whether or not a situation s is "up to me" requires that my decisions influence whether or not s occurs. If I am deciding between s_1 and s_2 and alternative decisions d_1 and d_2 are open to me then they influence which situation obtains if the subjunctive (counterfactual) conditionals

$d_1 > s_1$ and $d_2 > s_2$ ($d_1 = I$ decide d_1 , $s_1 = s_1$ obtains etc.)

² Bohmian mechanics and the increasingly popular "many worlds" understanding of quantum theory both have dynamics that are deterministic.

³ There are arguments analogous to the Consequence Argument that are claimed to show that free will is also incompatible with probabilistic theories like GRW that cover all physical events (see Loewer 1997).

are true.

The Consequence Argument reformulated in terms of influence is

- 1) Determinism: The total state of the world at time t (and that it is the total state) and the laws imply the states of the world at all other times t' .
- 2) I have no influence over the past at time t . i.e. there are no decisions d_1, d_2 open to me at t where s_1, s_2 such that $d_1 > s_1$ and $d_2 > s_2$ and where s_1 and s_2 are incompatible situations that pertain to times prior to t
- 3) I have no influence over the laws at t . There are no decisions d_1, d_2 open to me at t such that $d_1 > L_1$ and $d_2 > L_2$ where L_1 and L_2 are incompatible laws

Therefore:

- 4) I have no influence over the future at t .

Since having influence is necessary for free will it follows that I lack free will.

This argument is valid.⁴ Premises 2 and 3 strike many as obviously true. Nevertheless, there have been attempts to show that one or the other of the premises are false. I will first discuss a couple of proposals for rejecting premise 3 and then develop and defend a proposal for rejecting premise 2.

One approach is to argue that on a Humean view about laws that the laws are in a certain sense up to us. Beebe and Mele call this position "Humean Compatibilism." According to David Lewis' version of Humeanism, the Best Systems Account, laws are certain generalizations entailed by the axiom system that best systematizes the totality of fundamental events. The fundamental events are the instantiations at space time points (or in space time regions) of fundamental categorical properties. According to Humean compatibilism since my decisions and actions are constituted by the fundamental events that the best system systematizes and since the laws depend on totality of fundamental events my decisions and actions are part of the determinants of the laws. In this sense the laws are "up to me."

According to Helen Beebe and Al Mele

"... Humean perspective, then, imposes a modus tollens on van Inwagen's quick argument quoted above—and, a fortiori, on the modal version of the consequence argument. Since, in his deterministic world, it is up to Fred whether or not he skips breakfast, and it is manifestly not up to Fred what went on in the distant past, it is up to Fred what the laws of nature are. We admit that this may sound peculiar to some readers, but to Pat's Humean ears it is merely a way of saying that laws do not have the kind of ontological clout that would enable them genuinely to

⁴ The argument is valid when formulated in Lewis or Stalnaker logics for counterfactuals.

impede exercises of free will. According to the Humean view we have been assuming, there is no feature of the world that deprives Fred of the ability to skip breakfast tomorrow. “

On Humean accounts laws don't govern our decisions or anything else. Humeans generally hold that the idea that laws govern is a relic left over from the 17th century theological origin of the concept of laws of nature. On David Lewis' account - the BSA- laws don't govern but describe. They earn the title of “law” by being implied by the best systematization of the entirety of fundamental events. If not governing is what is meant by laws not having “ontological clout” then Humeans should agree that laws lack ontological clout. In contrast, non Humean accounts encourage the feeling that laws and prior states since they produce or constrain our decisions, they deprive us of free will. Jenann Ismael expresses this

When we adopt a globalist perspective, our activities become part of the pattern of events that make up history. Since our activities partly determine the pattern, and the pattern determines the laws, our activities partly determine the laws. But then something weird happens. We invert the order of determination and reify the laws, so that now it looks like the laws are not simply descriptions of patterns that is partly constituted by our actions but are instead iron rails built into the spatial and temporal landscape that won't let us act in any way not in accord with them. (p. 111)⁵

Humean laws certainly seem to be friendlier than non-Humean laws to free will. But even if Humean laws don't have ontological clout that doesn't mean that they support the claim it is “up to me” what the laws are in the sense relevant to the Consequence Argument. Recall that sense is that a law is up to me only if it is counterfactually dependent on my decisions. But counterfactual dependence of the laws on my decisions doesn't follow from the laws supervening in part on those decisions. This is a good thing too since otherwise Humeans would be faced with dealing with the problem that Humean laws would seem to open the possibility that I could travel faster than the speed of light if I decide to. Fortunately for Humeans this is not a consequence of Humeanism about laws. It is completely compatible with Humeanism that laws are not up to us in the sense that our decisions can influence them. That is, there are accounts of counterfactuals on which if L is a law and A and -A are compatible with the laws then both $A > L$ and $-A > L$ are true whether the laws are Humean or non Humean.⁶ I will discuss such an account later.

Although David Lewis is the most prominent proponent of a Humean account of laws he does not rely on Humean compatibilism to respond to the Consequence Argument. His response is to reject 3 not on the basis of his Humeanism about laws but on the basis of his account of counterfactuals. On the usual possible world semantics for counterfactuals $A > B$ is true at @ iff B is true at all possible worlds most similar to @ at which A is true B is also true.⁷ The truth conditions of counterfactuals depend on how “similarity” is understood. Lewis proposed an account that counts conformity to the actual laws as important but allows for local violations of the laws of @ he calls “small miracles” in worlds that are

⁵ Ismael's account of free will in *How Physics Makes Us Free* is subtle and full of insights.

⁶ Humean accounts of laws entail that whether or not a proposition is a law supervenes on the totality of non-nomological facts and so they will endorse counterfactuals of the form if A were the case then L would not be a law. But this is different from if A were the case then L would not be true, and it is this that is relevant to the Consequence Argument.

⁷ More accurately $A > B$ is true at @ iff there is a world W at which A&B are true and there is no world W* as or more similar than W to @ at which A&-B is true.

maximally similar to @. This is due to the fact that while conforming to @'s laws is important so are the extents of regions of perfect match to the actual world and these criteria can conflict. According to Lewis' well-known account the considerations involved in determining world similarity are:

1. Avoid big, widespread, diverse violations of law. ("big miracles")
2. Maximize the spatio-temporal region throughout which perfect match of particular fact prevails and maximize the time period over which the worlds match exactly in matters of fact
3. Avoid even small, localized, simple violations of law. ("little miracles")
4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

On this account even when A is compatible with the laws L $A > L$ may be false.⁸ For example, the truth of the counterfactual "If Nixon had decided to press the button at time t there would have been a nuclear war" involves a "small miracle" in a world similar to the actual world that violates the actual deterministic laws. Lewis thinks that the most similar world W to the actual world @ at which Nixon decides to press the button is a world that exactly matches @ up until a short time before t when there is a very local violation of the actual deterministic laws (Lewis calls the violation a "miracle") allowing for Nixon's decision to occur and then proceeds in conformity with the actual world laws to a nuclear war. So even if the state of the universe long before Nixon was born and the actual deterministic laws entail that he won't decide to press the button Nixon's decisions influence whether there is a nuclear war.

Lewis responded to the Consequence Argument as follows: Nixon's decision to press the button requires a small prior miracle prior to the decision but this doesn't mean that Nixon is unable to decide whether or not to press the button. Even though the prior state and the deterministic laws entail that Nixon won't press the button (thankfully he didn't) he was able to do otherwise (press the button) since this required only a small miracle. Since Nixon was the president, he was able to influence whether there would be a nuclear war. But this entails that Nixon also has a kind of influence over the laws and what happens a short time before the supposed button pressing since if he were to decide to press the button a miracle would have occurred a short time prior. Since Nixon has the ability to decide whether or not to press the button, he also has the ability to make a decision that requires a prior miracle. This may sound odd but Lewis points out that this ability doesn't mean that Nixon is able to break a law or to do something that causes a law-breaking event since the violation of law occurs prior to the decision. Although there is counterfactual dependence of the laws and the immediate past on Nixon's decisions it is paltry and useless. Nixon doesn't have the ability travel superluminally or revive the dead or engage in any act that is itself law-breaking or causes a law to be broken only that he has the ability to make

⁸ It might be thought that Lewis' account of similarity that allows for worlds maximally similar @ to include "miracles" is connected with his Humeanism regarding laws. Non Humeanism regarding laws suggests that worlds most similar to the actual world contain the same laws as the actual world and thus no violations of the actual worlds. But Humean views of laws and Lewis' account of counterfactuals are logically independent. A non-Humean can adopt Lewis account of similarity and a Humean can adopt an account of similarity on which laws are not violated in the most similar A-worlds as long as A is compatible with the laws.

decisions and perform actions such that if he had made and performed them a law would have previously been broken. Lewis says that this dependence of the laws on decisions may strike us as unintuitive, but it is not clear that is false.

Lewis's response to the Consequence Argument is that premise 3 and 2 are strictly false since we can influence the laws and the past in the ways just described but that it is not obvious that we don't possess this ability. And on his account of counterfactuals this is required for us to have significant influence over the future. Van Inwagen agrees that Lewis' response succeeds in showing that the Consequence Argument is not a "slam-dunk" but maintains that it is too high a cost to pay to defeat the Consequence Argument.

There are two problems with Lewis' response to the Consequence Argument; one small and the other devastating. The small one is that it seems to many that the principle that if L is a law and A is logically compatible with L then $A \gt L$ is obviously correct. Marc Lange even characterizes laws as propositions that satisfy this condition.⁹ Be that as it may Lewis' account of abilities in terms of counterfactuals does have an odd feature. Nixon has the ability to press the button. It follows that he has the ability to do something that if he were to do it would falsify the laws. Of course, he never exercises this ability. Someone who hates determinism has the ability to falsify it with every decision he faces. Suppose his hate is so great that he makes decisions randomly hoping that by chance he will succeed. It is astonishing that despite his ability Lewis' account implies that he can't succeed.

The devastating problem is that Lewis' account of counterfactuals is not successful. It fails to capture the temporal asymmetry of counterfactuals and influence. Counterfactuals are temporally asymmetric in that local counterfactual antecedents that depart slightly from actuality can and often do lead to consequents that greatly depart from the actual course of events in the future but not in the past. For example, if Nixon had pressed the button at time t the course of events after t would have been radically different but the course of events prior to t would have at least been pretty much as it actually was. For example, there would still have been a cold war, atomic bombs used dropped on Japan and so on. Lewis believed that the truth conditions embodied in 1-4 captured this asymmetry. But Adam Elga showed that due to the temporal symmetry of the fundamental dynamical laws this is wrong. On conditions 1-4 a world W^* at which Nixon presses the button and that satisfies the actual dynamical laws except for a small "miracle" after the button pressing is at least as similar to the actual world as the world W in which the miracle occurs prior to the button pressing. In W^* the past of the button pressing is vastly different from the actual world's past while their futures match. In W it is the past that matches while the futures differ.¹⁰ The consequence is that on Lewis' account since W^* is as similar to @ as W is "If Nixon had pressed the button the past would have been pretty much as it actually was while the future vastly different" turns out to be false.

⁹ Lange endorses the stronger principle that if A is compatible with L and L is a law then $A \gt L$ is a law. As mentioned in the previous footnote this is incompatible with Humean account of laws.

¹⁰ W^* is constructed by temporally reversing the actual world from its end until a time shortly before the time of the button pressing when a "miracle" leads to the button being pressed after which W^* continues in accord with the actual laws. W^* differs from W and the actual world in that its entropy increases in both temporal directions from the time of the button pressing.

So Lewis' account fails to capture the temporal asymmetry. On his account it can turn out that I have no influence over the future but large influence over the past. A quick fix is to build into the account a distinction between past and future by stipulating that match or similarity with respect to the history prior to t is much more important than match after t .¹¹ Lewis hoped to explain times' arrow (the distinction between past and future) in terms of his account of counterfactuals but this would presuppose the arrow rather than explain it. In the following I will describe an account of counterfactuals that captures the temporal asymmetry without presupposing a temporal direction and show how it can defeat the Consequence Argument.

Lewis's strategy for defanging the Consequence Argument was to attempt to show that on his account of counterfactuals decisions can significantly influence the future but their influence over the laws and past is paltry and useless. His idea is that while laws (and to an extent the past) counterfactually depend on decisions this doesn't provide us with an ability to perform a law breaking (or past altering) action only with an ability to perform an action that requires a prior "miracle." Lewis' account fails because his account of counterfactuals fails. But his strategy of rejecting one of the premises of the argument and showing how to make this acceptable is on the right track. A better account of counterfactuals is needed. Fortunately, physics provides the materials for such an account.

The part of physics that rescues freewill from the Consequence Argument is statistical mechanics. Boltzmannian statistical mechanics was developed to explain thermodynamic regularities and in particular, the second law of thermodynamics. The second law says that the entropies of the universe and suitably isolated subsystems increase over time until equilibrium (maximum entropy) is attained. The entropy of a system's microstate is measured by the size (relative to a standard measure) of the set of microstates that realize it. The second law, unlike the fundamental dynamical laws, is temporally asymmetric; entropy increases from past to future. Boltzmann's initial attempt to account for it involved observing that "most" (on the standard measure) micro states that realize the microstate of a system not yet at equilibrium- say a half melted ice cube sitting in a tub of warm water at time t - evolve in conformity to the micro dynamical laws to microstates that realize higher entropy macrostates – the ice cube becoming more melted- until it reaches equilibrium- the ice cube fully melted. On the assumption of a uniform probability distribution over microstates it follows that it is very likely that entropy of a system not at equilibrium will increase. Good! But Boltzmann soon realized that the because of the temporal symmetry of the fundamental dynamical laws it also follows that it is also very likely that the system evolved from a state of higher entropy i.e. that the ice cube was more melted prior to t . Bad! It was pointed out by Reichenbach, Feynman and others that this result can be avoided by positing that the entropy of a system to which statistical mechanics is applied starts out low. Applied to the entire universe the proposal is that is entropy at the earliest time was very small. David Albert

¹¹ If one does this and also adopts NP, then the result is an account of counterfactuals very close to the account that Jonathan Bennett develops in *A Philosophical Guide to Conditionals* Kadri Vihvelin employs it in her defense of compatibilism.

calls this assumption “the past hypothesis” (PH). He and I have proposed a fundamental theory that includes the PH and the probability distribution as laws that we call “the Mentaculus.” The Mentaculus has three ingredients:

1. The fundamental dynamical laws
2. The past hypothesis
3. An objective uniform probability distribution over the micro histories compatible with the PH.

Albert has argued that the Mentaculus entails not only that the entropies of the universe and its suitably isolated subsystems increase from its earliest times but also that it plays a central role in grounding two other important temporal asymmetries- the asymmetry of records and the asymmetry of influence. I showed how the Mentaculus can be used to characterize counterfactuals involving decisions and argued that the PH and the probability posit are best understood as laws along the lines of Lewis’ best system account. I won’t repeat the arguments for these claims here. However, I do want to emphasize that despite its name the PH does not presuppose a past future distinction. It claims that at one point in the universe (around the time of the Big Bang) the entropy of the universe was very small. It doesn’t say that this time is in our past. If Albert and my arguments succeed in showing that the Mentaculus explains the temporal asymmetries of influence, records, counterfactuals, and so on then it will *explain* time’s arrow not presuppose it.¹²

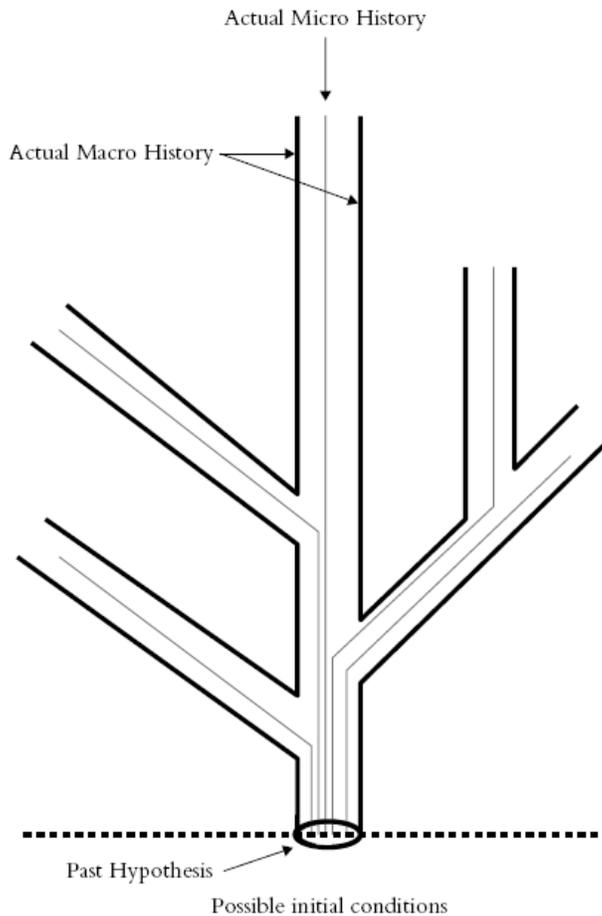
For the remainder of this paper I will show how the Mentaculus account of counterfactuals can be used to provide a response to the Consequence Argument and a metaphysical framework for a compatibilist account of free will.

The distinction between macro and microstates is central to understanding how the Mentaculus grounds temporal asymmetries. The microstate of the world at a time t is the complete fundamental physical state at t . In classical mechanics this consists of the positions and momenta at t of all the fundamental material particles. The notion of macro state is more flexible. The thermodynamic macro state of a system S consists of the values of thermodynamic quantities- temperature, energy, pressure, average frequency of radiation, mass density etc.- in small volumes of S . More generally, the macroscopic state of the universe relative to a collection Σ of macroscopic properties is the complete description in terms of the properties in Σ . A macro state is realizable by a continuous infinity of microstates. Its entropy is the measure on the standard measure of the set of microstates that realize it. If S is a suitably isolated system whose macro state is M then the vast majority of microstates compatible with M evolve to macro states whose entropies are greater. The same holds for every small convex set of infinitely many microstates.¹³

¹² For elaborations and defenses of the claims see Albert (2000, 2012) and Loewer (2001, 2007, 2019).

¹³ The reason is that in the neighborhood of every abnormal entropy decreasing microstate almost all (on the standard measure) states are normal entropy increasing state. It is also the case that every convex set of microstates contain some abnormal entropy decreasing states.

The Mentaculus implies that while the evolution of the microstate of an isolated system is deterministic the evolution of the macro realized by that microstate is indeterministic where the probabilities of evolutions are the statistical mechanical probabilities. So, there are micro trajectories that are macroscopically identical until a time t when they macroscopically diverge. The result is a ranching forks structure for the evolution of macro histories that looks like this.



There are infinitely many micro-histories emanating from the PH that are macroscopically indiscernible until time t and then macroscopically diverge. There will also be microstates that are macroscopically distinct for a period and then converge. As long as the time is not too long after to the time of the PH divergences greatly predominate over convergences. The overall picture is one in which while determinism obtains on the microscopic level indeterminism reigns at the macroscopic level.

I will suppose, as seems plausible, that alternative decisions correspond to very small differences in brain states -say the firing of a few neurons- and that the macroscopic state at time t outside of the agent's skull is compatible with the agent choosing either d_1 or d_2 a moment after t .¹⁴ Two deterministic microhistories h_1 and h_2 may be macroscopically indiscernible at time t when they

¹⁴ By "a moment after t " I mean a moment between the time of the PH but right before the decision. This doesn't presuppose the temporal direction of influence but is part of its explanation.

diverge with one realizing decision d_1 and the other decision d_2 a moment after t . Subsequent to t the macroscopic histories M_1 realized by h_1 and M_2 realized by h_2 may be very different. For example, M_1 may include Nixon pressing the button and all that follows and M_2 is the actual macro history in which he didn't press the button. The macroscopic state M at t doesn't determine what decision an agent will make a moment after t although the microscopic state that realizes M does determine the subsequent positions of Nixon's finger. Christian List claims that indeterminism at the agent level by itself is sufficient to defuse the consequence argument. But I don't see that this can be so since if I can't influence the laws and can't influence the past and determinism is true it will follow that I can't influence the position of my finger.¹⁵ However, indeterminism at the macro level does provide a sense of possibility on which alternative decisions are possible for me.

I propose we evaluate decision counterfactuals as follows

$d_1 > P(A)=x$ is true iff the $P(A/d_1 \& M(t))=x$ where $M(t)$ is the macroscopic state outside of the agent's skull a moment prior to d_1

For example, if d_1 is Nixon's pressing the button the probability of there being a nuclear war soon after d_1 given the macroscopic state a moment prior to d_1 will be very high. If d_2 is his decision of not pressing the button then the probability of there being a nuclear war soon after d_2 is very low. So Nixon has influence over whether there is a nuclear war. What about events prior to the decision? If E is a macroscopic event then except under very unusual circumstances $P(E/M(t) \& d_1) = P(E/M(t) \& d_2)$.¹⁶ In other words E is probabilistically independent of d_1, d_2 . And if $M(t)$ contains records of E then this probability is near 1. If the macroscopic state at the time just prior to Nixon's decision contains records of his crimes that led to his impeachment inquiry, then whatever he decides the probability of his having committed those crimes at the time of his decision remains near 1. His decision has no influence on them.¹⁷ On this account of influence Nixon does have influence of the world's micro history. Since determinism obtains the alternative decisions occur on different micro histories, m_1, m_2 of the world so the counterfactuals $d_1 > m_1$ and $d_2 > m_2$ are true. In other words, Nixon's decision influences the micro history of the world. But all this means is that had Nixon decided to press the button the micro history of the world would have been different all the way back to the big bang. This influence is useless and paltry. As we observed it doesn't enable Nixon to influence any past macroscopic event.¹⁸

¹⁵ See List (2018). List seems aware of this worry but tries to wriggle out of it by claiming that it is some kind of category mistake to combine agential and fundamental physical levels. But I don't see how this can show that the consequence argument is unsound.

¹⁶ The special circumstances are ones in which there are no macroscopic records of E so that d_1, d_2 serve both to influence and be the only records of whether or not E occurred. Mathias Frisch discusses a case like this, but it is so unusual that it has no effect on my response to the Consequence Argument.

¹⁷ If he decides to press the button and a nuclear war ensues that destroys the earth than that might destroy all macroscopic records of his crimes. If so then his decision influences the future probability of his having committed those crimes but the probability relevant to influence is the probability at the time of making the decision.

¹⁸ This account applies to decision counterfactuals. One way of extending it to counterfactuals with a non-decision antecedent A is to find times between the time of the PH and the time of the antecedent and close to the latter at which a decision like event makes A likely. This approach to counterfactuals is close to interventionist accounts of causation.

The account of decision counterfactuals in terms of conditional probabilities is very much like Lewis' account in terms of world similarity but differs from it in important respects. In evaluating $d1 > P(A) = x$ we consider the set of worlds that match the actual world a moment t prior to $d1$ outside of the agent's skull and at which $d1$ is true and then ask how likely A is according to the Mentaculus. The $d1$ worlds we are considering are identical with respect to macroscopic events to the actual world prior to t as long as there are records at t outside of the agent's skull of these events.¹⁹ However, if $d1$ is not an actual decision then these worlds will be microscopically different all the way back to the big bang. Macroscopic similarity over the period between $d1$ and the PH replaces Lewis' "perfect match". Unlike Lewis' account the dynamical laws are not violated in these $d1$ worlds. So, no "miracles." In fact, not only is the Mentaculus not violated in these worlds but it continues to be a law even on Humean accounts of laws. Finally, this account gets the temporal asymmetry of decision counterfactuals and does it without presupposing the past-future distinction.

My response to the Consequence argument is to reject premise 2 by showing that given the physics of our world (i.e. statistical mechanics) and a proper understanding of influence freely chosen decisions may influence the past. But that the influence over the past is useless and paltry since it doesn't enable us to exert any control over past events that are of any interest to us. My response is similar to Lewis' response but differs in that my account of counterfactuals involves no "miracles" and implies that counterfactuals are time asymmetric in the way required to endorse our having significant influence over the future but not the past. Like Lewis' account my response has what may strike one as a counterintuitive consequence. Where Lewis' account implies that an agent has the ability to make a decision which if she had made it a law would have previously been broken my account implies that she has the ability to make a decision which if she had made it the microscopic past would have been different from the actual past. Where Lewis' account implies that I have the ability to falsify it even though it is an ability I never exercise my account doesn't have this consequence. On my account I have no influence over the laws. It does say that if h is a specification of the actual micro history, I have the ability to falsify h though I never exercise that ability. But this doesn't seem nearly as troubling. Both Lewis' and my account have a counterintuitive ring but it is I think easier to come to appreciate that feeling that we can't influence the past is based more on mistaken views about time than the feeling that we can't influence the laws.

One might resist the idea that we have influence over the past even of the kind I am supposing because on confusing having influence with actually changing, The fact that I have influence over the past doesn't mean that I can change the past. Our picture of time is one in which we imagine that as time flows events are nailed down and once nailed down cannot be changed. This is encouraged by thinking about time in terms of the so called "A-series". But from a B-series perspective there is no question of changing either the past or the future. I can change the channel on my tv since if it is on channel 4 at one time, I can cause it to be on channel 7 at a subsequent time. But my influence over the past doesn't mean that if the past is one way at a certain time, I can make it different at a subsequent

¹⁹ If there is not a record of an actual macroscopic event E then the Mentaculus may assign a low probability of its occurring and so the relevant counterfactual $d1$ worlds are mostly ones in which E does not occur. But in this case whether or not $d1$ the probability of E is low so the decision has no influence on E 's occurring.

time. Past events are no more “nailed down” than are future events. Talk of “nailing down events”, if it makes any sense at all, refers to the fact that we can influence future events in ways that we cannot influence past events.²⁰

We live in a world in which the fundamental dynamical laws are (or well may be) deterministic. But there are further non-dynamical laws specifying a low entropy boundary condition and a probability distribution. The result yields indeterminism at the macro level and determinism at the micro level. This explains why our decisions appear to be undetermined. More significantly it shows how our decisions can influence the future in ways that are important to us while having no influence of interest over the past. My proposal is that this is what freedom consists in. If this is correct, then rather than physics undermining freedom it is part of its explanation.²¹

²⁰ See Carl Hoefer’s (2002) for an insightful discussion of how mistaken views about time underlie the alleged conflict between free will and determinism.

²¹ Of course physics is only a part of the explanation of free will since at most it makes free will physically possible. Free will involves making decisions on the basis of reasons and rational deliberation. Accounts of these are the business of psychology and rationality theory.