Things I Need to Know on Free Will[1]

What is free will?

- When an agent exercises free will over her choices and actions, her choices and actions are *up to her*. But up to her in what sense? Two common (and compatible) answers are:
  - Up to her in the sense that she is able to choose otherwise, or at minimum that she is able not to choose or act as she does, and
    - E.g. I choose not to lie even though I could and no one would find out
  - Up to her in the sense that she is the source of her action (O'Connor and Franklin/SEP 2021)
    - Incompatibilist libertarian example: I deliberately pushed a button to stop some electric saw running over a child's hand in the carpenter's workshop, not because someone bumped into me such that I accidentally fall atop the stop button and the child is saved

Why does free will matter in the first place?

- Ethics is about normative statements (what is good or bad, right or wrong), and also responsibility, praise, and blame (when is someone morally responsible)
- It may be that we are morally responsible for our actions iff we have free will and we performed those actions as an exercise of free will
- **The free will problem**: powerful arguments that seem to show that we cannot be morally responsible in the ultimate way that we suppose keep coming up against equally powerful psychological reasons why we continue to believe that we are ultimately morally responsible (Strawson/REP 2011)
- Both our autonomy and our accountability seem to hinge on free will

What is determinism and indeterminism?

- Determinism holds that the prior state of the world together with the laws of nature entail the present state of the world
- It follows that whatever I did, I could not have done otherwise in at least one sense, because what I did was entailed by the previous state of the world and the laws of nature
- But it does not follow that we don't have free will (at least on some views; for example, classical compatibilists say that there are other sense in which we satisfy the 'could have done otherwise' condition)
- Determinism cannot be falsified (although it has been argued that quantum theory falsifies it)
- Indeterminism is the denial of determinism

What is Hume's argument for indeterminism?

- If determinism is true, everything has to have an ultimate cause, and that ultimate cause has to be the Creator
- But people commit crimes. Either the Creator foresaw that people do so, and it was not their initiative; or those crimes aren't actually crimes
- Reductio ad absurdum: either of these conclusions are absurd, so determinism is false
- If determinism is false, the only alternative is a certain amount of randomness; but if our behaviour is random, then we are not responsible

What are the main views on the issue?

---

[1] Special thanks to the contributions of Martin Yip

View	Determinism	Free Will	Moral Responsibility	Associated People
Hard Determinism	YES	NO	NO	Galen Strawson Martha Klein Derek Pereboom
Soft Determinism (Compatibilism)	YES	YES	YES	Hume Frankfurt
Libertarianism	NO	YES	YES	Van Inwagen Robert Gale Kane
Scepticism (Pessimism)	Doesn't matter		NO	Peter Strawson

- Kant can be interpreted as a soft determinist (he acknowledges both the physical/natural world and the noumenal world; one can freely choose what one does in the noumenal world although what actually happens depends on a causal chain within the natural world) or a libertarian

What is compatibilism?

- Compatibilism holds that determinism is compatible with free will (so even if the world is deterministic, that does not preclude us from having free will)
- The compatibilist conception of freedom is that it is essentially just a matter of not being constrained or hindered in certain ways, e.g. being drugged or in chains, *given how one is* (e.g. genetic inheritance, upbringing)
  - E.g. someone who is brought up thinking aliens exist would be as free in making their decision as someone who is brought up not thinking aliens exist when it comes to the topic of space exploration, given that neither of them are drugged or influenced in that instance of making a decision to think in a particular way; not about choosing according to some set standard
  - BUT this seems to contrast with Kant's idea that one can only choose freely if they are rational, but compatibility allows for irrational people (e.g. a psychopath) to be free decision-makers too
- Thus, if what I did was determined by my thought processes and my desires and purposes, perhaps I am 'free' even if these factors are deterministic

What are the arguments for compatibilism?

- Frankfurt (1969) argues against the Principle of Alternate Possibilities (PAP), i.e. an agent is morally responsible for an action only if she could have done otherwise
- Frankfurt cases seem to show that we can have freedom without alternative possibilities
  - Suppose Jones is deciding between voting Democrat or Republican; he decides to vote Democrat. Unbeknownst to him, a neurosurgeon Black has installed a device in his brain that ensures that, if Jones' deliberation is such that he will vote Republican, it will intervene such that Jones will decide to vote Democrat. Since Jones in fact never wavers in his intention to vote Democrat, Black never intervenes
  - This suggests that there is a kind of freedom or control — corresponding to choosing and acting freely — that does not require alternative possibilities, and that this sort of control (and not the alternative-possibilities control) is the freedom-relevant condition necessary for moral responsibility (Fischer, in Copp (ed) 2007)

- One could argue that the counterfactual is irrelevant: the fact that Black's device would have intervened, had it needed to, does not in itself absolve Jones of moral responsibility. Indeed, the fact has no bearing on Jones' moral responsibility at all
  - The main basis for the judgment that Jones is responsible in a Frankfurt case is that, since Black did not *actually* intervene, Jones *acted freely*, indeed *willed freely*.
- The point is, if PAP is true, incompatibilism is true; but if PAP is false, this makes compatibilism more plausible, since it has been shown that determinism does not block free will *by way of ruling out alternate possibilities*
  - Note that Jones sees many epistemological possibilities (he thinks he could vote Republican or Democrat), even though there is only one metaphysical possibility (to vote Democrat)

What are the problems with compatibilism?

- **The prior sign dilemma:** The way Frankfurt's case works is that Black observes a 'prior sign' of what Jones intends to do at T1, and Jones acts at T2. Is the connection between the prior sign and the subsequent action deterministic or not?
  - If so, then Jones' action at T2 is deterministically brought about by factors beyond his control, so incompatibilists would think that he is not morally responsible
  - If not, then Jones appears to have free will at (or just prior to) T2: given the prior sign and the laws of nature, it does not follow that Jones will do what the sign suggests. So, he seems to have alternate possibilities (he retains the ability to do otherwise)
  - Therefore, it seems that either Jones is not morally responsible or there are alternative (metaphysical) possibilities for Jones; Frankfurt cases fail to provide a single context in which it is both true that Jones has no alternative possibilities (is unable to do otherwise) and is morally responsible for his decision
- **The trivialness argument:** The critic might argue that I am not free because I have no influence over my thought processes and my desires and purposes; so there is no other choice I could have made
  - But perhaps I am free in the sense that if I had wanted to do otherwise, I would have done otherwise
  - My lack of alternatives is due to my lack of a deviant desire rather than the lack of the possibility of my actions being otherwise
  - However, if the world is deterministic, then 'I had wanted to do otherwise' would be logically impossible, so the conditional would be trivially true
- **The low bar argument:** (Strawson/REP 2011) If free will is simply a matter of having genuine options and opportunities for action, and being able to choose between them according to what one wants or thinks best, then one might argue that on this account, dogs and other animals may be free agents too. But we don't think that dogs can be free or morally responsible in the way we can be. So the compatibilist needs to explain what the relevant difference is
  - We might think they're free (at least sometimes) in the sense which is necessary for moral responsibility, but they fail to satisfy some other necessary condition(s) for moral responsibility which we do satisfy
  - One idea is to say that it is our *capacity for self-conscious thought* that makes the crucial difference, because it makes it possible for us to be explicitly aware of ourselves as facing choices and engaging in processes of reasoning about what to do
  - Another idea is to say that we have the capacity to act for reasons which we explicitly take to be moral reasons (which dogs do not have)
  - Frankfurt (1971): As reflective beings, we are capable of a kind of freedom that is not available to the other creatures: we have the capacity to critically evaluate the motivational

forces to which we are subject. This capacity is manifested in 'higher-order volitions': desires that some first-order desires be (or not be) effective in action
  - E.g. I can restrain myself from being noisy even though I want to blast rock ballads on the karaoke machine but I do not because I have a higher-order concern for my neighbours' welfare whereas a dog does not have such cares or if they did they cannot go against it because they just bark noisily all the time
- **The moral responsibility argument:** (Strawson/REP 2011) An incompatibilist notion of free will is essential in order to make sense of the idea that we are genuinely morally responsible
  - But the compatibilist would argue that since 'ultimate' moral responsibility is impossible, we should rest content with the compatibilist account, being the best we can do
  - The pessimist would argue that neither account works: indeterminism doesn't help. So, no punishment or reward is ever truly just or fair, when it comes to moral matters

What is incompatibilism?
- Incompatibilism holds that if the world is deterministic, we do not have free will
- The incompatibilist conception of free will is that we are only free if we are free in an absolute sense – free from external impediments and coercion, but also from causal processes that happen internally, within the confines of our brains and bodies (but most libertarian views would hold that we don't need to be completely uninfluenced by our internal processes to be free)

What are the arguments for incompatibilism?
- Van Inwagen's (1983) Consequence Argument: If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us

1. If determinism is true, everything, including human actions, is causally necessitated by the prior state of the universe in accordance with the laws of nature
2. If human actions are causally necessitated by the past together with the laws of nature, then we cannot ever do otherwise than what we do, unless we can falsify the laws of nature or falsify the description of the past
3. We cannot falsify the laws of nature or the description of the past [so we cannot do otherwise]
4. If we cannot act otherwise than we do, then we lack free will
5. Hence, if determinism is true, we lack free will
6. [Extension:] If we lack free will, we cannot appropriately hold one another responsible
7. Hence, if determinism is true, we cannot appropriately hold one another responsible

- It follows that there is no causal possibility that somebody could have done otherwise than they actually did

- One sort of argument that motivates incompatibilism argues that determinism would make it impossible for us to *cause and control our actions in the right kind of way* [focus on notions of self, causation, and responsibility] (Vihvelin/SEP 2018)
  - If whatever we choose depends on the past chains of causation then my choice could not have been made freely, but unless my choice was made freely, I cannot be held morally accountable for my actions
- Another sort of argument that motivates incompatibilism argues that determinism would deprive us of the *power or ability to do or choose otherwise* [focus on notion of choice] (ibid)

- E.g. if, finally, I have made up my mind and I will pull the lever to save five people from getting run over by a trolley at the expense of one life, but some random person unties the five from the tracks, it turns out that I kill one person without a good reason, and if determinism was true and that person was bound to turn up, then still my decision does not matter because under determinism I could not have chosen to do otherwise either

What are the problems with incompatibilism?

- Perhaps we are subject to moral assessment because our desires, purposes, and mental characteristics can be judged morally; but if these are deterministic, it seems pointless and inappropriate to judge them just as it is pointless to judge people by their genetic makeup
  - Maybe agency has intrinsic value even if it is predetermined
  - What is important could turn out to be the fact that I did something regardless of whether I had chosen to do it (consequentialist)
- According to incompatibilism, the notion of personhood dissolves, and no boundaries exist between 'me' and the rest of the physical world. Thus, I am not really a bearer of moral agency [and so it would seem pointless and inappropriate to attribute moral agency to people?]
  - [But then the incompatibilist could perhaps say that's precisely the point? To show that we shouldn't attribute moral agency and make moral assessments]
  - Perhaps "moral agency" is just some social construct designed to impart blame or praise on people devised for functionalist reasons

What are the arguments for scepticism (pessimism)?

- Galen Strawson's (2011) Basic Argument:

1. Nothing can be *causa sui* – nothing can be the cause of itself
2. In order to be truly morally responsible for one's actions one would have to be *causa sui* (be a cause of itself)
3. Therefore nothing can be truly morally responsible

- It follows that even the way we are is a matter of luck, since one's ability and/or intention to change the way one is is predetermined
- One cannot possibly be *truly* or *ultimately* morally responsible for what one does if everything one does is ultimately a deterministic outcome of events that took place before one was born; or (more generally) a deterministic outcome of events for whose occurrence one is in no way ultimately responsible

What are the problems with scepticism (pessimism)?

- One could accept (1) but deny (2) [this is what the compatibilist might do]; perhaps the fact of self-conscious awareness confers responsibility
  - The positive argument: we can give a satisfactory account of the (admittedly elusive) notion of self-determination without insisting that self-determination requires us to be the first causes of our choices (e.g. Frankfurt 1971)
  - The negative argument: if we accept (1), we are committed to the conclusion that free will and moral responsibility are impossible, regardless of whether determinism is true or false [which is just pessimism; so this is a problem for incompatibilists, but not pessimists]
  - The pessimist might say that that fact may be a source of belief in ultimate moral responsibility, but not something that could constitute that responsibility

What is the libertarian dilemma?

- If determinism is true, then we are not free agents, according to libertarianism. But if determinism is not true, we are not free and responsible agents of undetermined events, because those are a matter of 'chance'
- So, there are two components to libertarianism
  - Negative component: indeterminism (denial of determinism)
  - Positive component: a non-deterministic form of explanation distinctive of free agency
- The libertarian needs to explain why the falsity of determinism is any better than the truth of determinism when it comes to establishing our free agency and moral responsibility. Why am I not merely lucky? How can the fact that my effort of will is indeterministic in such a way that its outcome is indeterminate make me truly responsible for it, or even help to make me truly responsible for it?

What are the arguments for libertarianism?

- Kane (1999) argues against the Luck Principle (LP): If an action is *undetermined* at a time *t*, then its happening rather than not happening at *t* would be a matter of *chance* or *luck*, and so it could not be a *free* and *responsible* action (LP is used to argue against incompatibilist notions of free will)
- **1. Rejection of LP:** One cannot infer from "indeterminism's being involved in something's happening" to "its happening merely as a matter of chance or luck", or from "it was undetermined" to "he was not responsible"
  - The inference might run as follows (**the luck argument**)
    - In the actual world, person *P* does *A* at *t*. On the assumption that the act is undetermined at *t*, we may imagine that:
    - In a nearby-possible world which is the same as the actual world up to *t*, *P** (*P*'s counterpart with the same past) does otherwise (does *B*) at *t*
    - But then (since their pasts are the same), there is nothing about the agents' powers, capacities, states of mind, characters, dispositions, motives, and so on prior to *t* which explains the difference in choices in the two possible worlds
    - It is therefore a matter of luck or chance that *P* does *A* and *P** does *B* at *t*
    - *P* is therefore not responsible (praiseworthy or blameworthy, as the case may be) for *A* at *t* (and presumably *P** is also not responsible for *B*)
  - Consider a husband in an argument with his wife trying to break the table. Whether the table is broken is indeterministic, e.g. his arm might twitch and reduce the force on the table. But if the husband succeeds, he would be hard pressed to say that he was not responsible just because it was a matter of chance (it was not certain) that the table has been broken
    - Husband and Husband* made the same efforts (as well as having the same capacities, motives, and characters) up to the very moment of breaking of the table. Yet it does not follow that the husband is not responsible when he succeeds
  - Suppose you are trying to think through a mathematical problem and there is some indeterminacy in your neural processes complicating the task – a kind of chaotic background. It would be like trying to concentrate and solve a problem with background noise or distraction. Whether you are going to succeed in solving the mathematical problem is uncertain and undetermined because of the distracting neural noise. Yet if you concentrate and solve the problem nonetheless, I think we can say that you did it and are responsible for doing it even though it was undetermined whether you would succeed. The indeterministic noise would have been an obstacle to your solving the problem which you nevertheless overcame by your effort

- Schrödinger has a cat named Kitty. He puts it into a box with radioactive material along with a Geiger counter, a vial of poison and a hammer. When the radioactive substance decays, the Geiger counter detects it and triggers the hammer to release the poison which would kill poor Kitty. Since the decay of the radioactive substance is random, Kitty could be dead or alive inside the box. Whether or not Kitty is dead is irrelevant; Schrödinger is still morally blameworthy for deliberately endangering Kitty.

- **2. Self-forming actions (SFAs):** When we act from a will already formed (as we frequently do), it is "our own free will" by virtue of the fact that we formed it (at least in part) by earlier choices or actions which were not determined and for which we could have done otherwise voluntarily, not merely as a fluke or accident. I call these earlier undetermined actions SFAs
  - If there were no such undetermined SFAs in our lifetimes, there would have been nothing we could have ever voluntarily done to make ourselves different than we are–a condition that I think is inconsistent with our having the kind of responsibility for being what we are which genuine free will requires
  - [SFAs] occur at times in life when we are torn between competing visions of what we should do or become [...] The uncertainty and inner tension we feel at such soul-searching moments of self-formation is reflected in the indeterminacy of our neural processes themselves
  - When we do decide under such conditions of uncertainty, the outcome is not determined because of the preceding indeterminacy and yet it can be willed (and hence rational and voluntary) either way owing to the fact that in such self-formation, the agents' prior wills are divided by conflicting motives
- Imagine that a businesswoman is trying or making an effort to solve two cognitive problems at once, or to complete two competing (deliberative) tasks at once
  - With respect to each task, as with the mathematical problem, she is being thwarted in her attempt to do what she is trying to do by indeterminism. But in her case, the indeterminism does not have a mere external source; it is coming from her own will, from her desire to do the opposite
  - I argue that, if she nevertheless succeeds, then she can be held responsible because, like them, she will have succeeded in doing *what she was trying to do*. And the interesting thing is that this will be true of her, *whichever choice is made*, because she was trying to make both choices and one is going to succeed
  - **Objection:** One may object that the businesswoman makes one choice rather than the other *by chance*, since it was undetermined right up to the last moment which choice she would make. [So effort comes first and "chance takes over" at last.] But this is wrong. You cannot separate the indeterminism from the effort to overcome temptation in such a way that *first* the effort occurs *followed* by chance or luck (or vice versa). One must think of the effort and the indeterminism as fused; the effort is indeterminate and the indeterminism is a property of the effort [So there is no point at which the effort stops and chance "takes over"]
  - Since both of them are simultaneously trying to do *both* of the things they may do (choose to help or go on, overcome the temptation to arrive late or not), they will do either with intent or on purpose, as a result of wanting and trying to do it–that is, intentionally and voluntarily. Thus, their "failing" to do one of the options will not be a mistake or accident, but a voluntary and intentional doing *of the other*
  - **Objection:** Perhaps we are begging the question in assuming that the outcomes of the efforts of the businesswoman and her counterpart were *choices* at all. If they were not choices to begin with, they could not have been voluntary choices. One might argue this on the grounds that (A) "If an event is undetermined, it must be something that merely happens and cannot be somebody's choice"; and (B) "If an event is undetermined, it must be

something that merely happens, it cannot be something an agent does (it cannot be an action).” But (A) and (B) imply respectively (A') "If an event is a choice, it must be determined" ("All choices are determined") and (B') "If an event is an action, it must be determined" ("All actions are determined”)
    - To choose is consciously and deliberately to form an intention to do something; and she did that, despite the indeterminism in her neural processes (as did businesswoman* when she chose to go on to her meeting)
- **3. Plural voluntary control:** It is one thing to say that she chose and another to say she chose *freely* and *responsibly*. This would require that she not only chose, but had voluntary control over her choice either way
- Indeterminism, wherever it appears, does seem to diminish rather than enhance agents’ voluntary control. We must also grant that indeterminism, wherever it occurs, functions as a hindrance or obstacle to our purposes that must be overcome by effort
- Despite the businesswoman’s diminished control over each option considered separately, due to a conflict in her will, she nonetheless has what I call *plural voluntary control* over the two options considered as a set. Having plural voluntary control over a set of options means being able to bring about *whichever* of the options you will or most want, *when* you will to for the reasons you will to do so, without being coerced or compelled in doing so
- Every SFA “is the initiation of a ‘value experiment’ whose justification lies in the *future* and is not fully explained by the *past*”
- In sum, *you can choose responsibly for prior reasons that were not conclusive or decisive prior to your choosing for them*

What are the problems with libertarianism?

- The complexities and uncertainties of Kane’s conjectures raise a worry of a different kind. On this account, it is very unclear how we could ever be as confident as we seem to be in our ascriptions of responsibility (Watson 2004)
- If one accepts the view [i.e. libertarianism], one will have to grant that it is impossible to know whether any human being is ever morally responsible. For moral responsibility now depends on the falsity of determinism, and determinism is unfalsifiable (Strawson 1994)

What is required for someone to be morally responsible?

- An agent S is morally accountable for performing an action X.
    - S deserves praise of X goes beyond what is reasonably expected of S
    - S deserves blame if X is morally wrong
- An agent’s moral responsibility consists in her being an appropriate candidate for ascriptions of certain ethical predicates, such as ‘good‘ ,’bad‘ ,’courageous‘ ,’charitable’, etc; an agent is morally responsible insofar as she has a “moral ledger” (Fischer, in Copp (ed) 2007)
- Strawson (1962): praising and blaming an agent consists in experiencing reactive attitudes or emotions directed toward the agent; in the case of praise, these emotions are such as gratitude, approbation, and pride; in the case of blame, these emotions are such as resentment, indignation, and guilt
    - These reactive attitudes or emotions need not directly translate to actions; one can experience an emotion of resentment but quickly forgive the other person
    - E.g. one may have negative emotions when they find out they are serially lied to but those attitudes readily disappear when they find that they have only been lied to for the purposes of keeping their birthday surprise a secret from them

- Strawson (1994): true moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven
- Pereboom (2014): *basic desert* means an agent deserves to be praised or blamed for their actions just because they have performed an action; on this view, it does not matter whether the agent has an understanding of the moral status of their actions
  - E.g. if I trip over and push a grandma onto the streets where she then gets hit by a car, I am blameworthy whether or not I had the intentions for the grandma to get hurt (consequentialist approach)
  - E.g. if I give to charity because I have entered a social contract which requires me to do so (and I only follow because I believe participating in that contract would give me the most goods), I am praiseworthy, and whether or not I actually intended to help that charity is unimportant (contractualist approach)
  - Problem: basic desert is wrong because we should not designate praise or blame unless their actions are up to them – i.e. that they have free will
    - E.g. if the universe is set such that I would always trip and cause the grandma injury and there was no way I could have done otherwise, it seems like I should not be blamed for her getting hurt
    - BUT free will seems to be different from having the correct intentions, which is required for moral praise
      - E.g. Even if I freely give to charity but only because it is personally optimal for me to do so as part of a social contract I have entered on grounds of private benefit, it seems like I do not deserve moral praise because I lack the correct intentions
    - Free will seems like a necessary but insufficient condition for moral blame or praise

What are the implications of this debate?

- **Moral deliberation:** We might think that moral deliberation is still meaningful, even if determinism is true (and therefore my actions are causally determined)
  - Recall the distinction between epistemic possibilities and metaphysical possibilities: agents deliberate based on the former, which may differ from the latter
- So, as far as normative ethics is concerned, the deliberative function of ethical theories do not lose force insofar as we still need them for our reasoning
- **Moral judgment:** If ethical theories make moral judgments about actions, and those actions were not products of free will, then these moral judgments lose prescriptive force; they are merely descriptive – like scientific theories explaining the state of the world