Is there a defensible libertarian account of moral responsibility?

Why moral responsibility rests on libertarianism

'If determinism is true, then, I am not free to make my decisions…'

The thesis of determinism is that the Universe is causally connected by long chains of events, which unfold according to overarching laws. Determinism, I argue, is at odds with moral responsibility because life's course would be fixed by these laws. For free will and moral responsibility to be real human behaviour must feature some undetermined properties.

Future states of a deterministic universe could, theoretically, be predicted by an extremely sophisticated simulation—or, perhaps, by an all-knowing, omnipotent deity. My behaviour would result from an amalgamation of physical events. With sufficient data available someone very clever *could* be capable of predicting what my future actions would be by determining the thought processes which preceded them. If determinism is true, then, I am not free to make my decisions because they are fundamentally set for me (Fig. 1); so I cannot be morally responsible for them. Right down to the atomic level of my brain, I would not be the *true source* of my actions: the Universe would be. This is not to say that we are forever fated to certain outcomes, anchored to single possibilities to be prophesised: rather, that we are not responsible for events which *we* do not determine.

Philosophers who believe that we can possess free will even if determinism is true are called 'compatibilists'. 'Incompatibilists', like myself, disagree. Specifically, I argue that, if I cannot fundamentally produce the thoughts behind my actions, I have no capability to choose or will or intend for my life to be like this. On a weak premise, one might argue, because I can 'conceive' multiple courses of action, I am responsible for the action I undertake ('principle of alternate possibilities'); but the will behind my thoughts could be determined by nature itself or even by a manipulator (e.g. a god, sinister neuroscientist, or controlling partner).

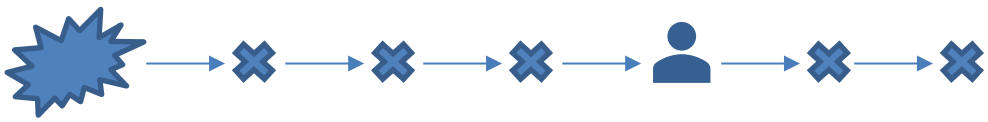

Fig. 1: An ultimately responsible universe entails an 'agent' in a single chain of causes and effects. An agent might be aware of complex thoughts and perform complex actions but there is nothing *intrinsic* to this complexity which constitutes *their* free will.

Indeterminism, on the other hand, describes a universe in which not all events are determined by antecedent causes. Predictions of outcomes could be made but they would not be founded on certainty, only on sets of likelihoods. In the eyes of physicists indeterminism is applicable at the atomic level: quantum mechanics tell us what *may* happen with assigned statistical probability. We can never, for example, know precisely when a nucleus will decay. Perhaps, then, the same is true for neural and mental events: our thoughts cannot be predicted, even when studied psychologically or through neuroimaging, because at any one time all possible avenues remain open. This uncontrollable randomness means we do not ultimately decide our thoughts and actions. How, then, could we conclude that we are morally responsible?

Libertarian philosopher Robert Kane, whose arguments are of focus in this essay, proposes a solution to this problem by factoring randomness *into* his moral equation (Kane 2002). I argue, conversely, that it is theoretically *possible* that we are in control of our actions: that nothing causes us to cause an effect.

In this case the thoughts behind our actions are *non*-random. This constitutes an 'agent-causal' theory, the only defensible libertarian account in my opinion, which is scrutinise in more detail later in the essay.

Kane's libertarianism

'A is ultimately responsible for being the source of an action if they hold "sufficient reason"…'

Libertarians theorise that moral responsibility is induced by certain undetermined actions. For many a moral agent must be the source, to some significant degree, of an action:

Sourcehood: For an agent, *A*, to be morally responsible for an action, *X*, *A* must be the source of *X*.

Robert Kane says this is true if *A* holds 'sufficient reason (condition, cause, or motive)' for an action's occurrence. In his words *A* is *ultimately responsible*. But multiple, undetermined possibilities of sufficient reason must be open to *A* to grant them 'plural voluntary control'; if there was only one possibility, the Universe's laws would be deterministic and *A*'s motivations would be redundant for reasons already discussed.

Moral responsibility arises from *A*'s conflicting motives, providing a means to 'do otherwise' and generating a kind of neurological randomness that is reducible to *A*. Through successions of these undetermined chains of events (Fig. 2) *A* is ultimately responsible for *these* actions (of sufficient reason). This version of libertarianism constitutes an 'event-causal' theory of moral responsibility. Kane contends that moral responsibility can be attributed based on the choices we make. But not all choices: only 'self-forming' ones which contribute to shaping who we become.

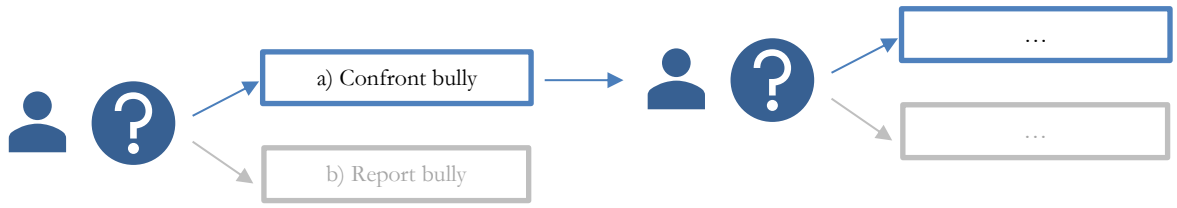

Fig. 2: *A*, according to Kane, is morally responsible for all potential outcomes in a given scenario on the premise that they have sufficient reasons for their occurrence. In this particular scenario *A* wants to deal retribution to a school student who has been bullying their sibling (a). However, *A* also wants to pursue a formal solution because *A* is concerned with their currently positive reputation at the school and how their grades will be affected if they are suspended (b). They are actuated into (a) by virtue of the randomness of the Universe but they are still morally responsible for it and would have also been morally responsible for (b) too.

A is morally responsible for holding sufficient reasons for these *voluntary* actions, despite not having the final say; they build their own *character* (*A*'s mental and moral properties) and *motives* through a culmination of these indeterministic events. At an instant just before *A* makes a choice it may seem to them that the idea presented itself but *their* decisions might have led to it (i.e. the source is not *A*): 'Often we act from a will already formed, but it is "our own free will," by virtue of the fact that we formed it by other choices or actions in the past for which we could have done otherwise' (Kane 2002, p408).

Consider *A* a shoplifter. They have *acquired* a pathological need for stealing. The process began when they were 18 years old, when they were low on money and needed food and clothes. They made a moral choice at the time to start stealing (in conflict with the moral choice to not start). Today the situation has spiralled out of control to become a compulsion. Yet, even though *A* feels out of control, *A* is morally responsible because they formed their free will from their prior life choices. *A* is *ultimately responsibility*, for

they are the source of motivations behind self-forming actions, notwithstanding the fact that their predispositions and other background conditions might take away *some* responsibility.

A key aspect of Kane's argument to consider is his discussion of alternate possibilities (the 'plurality conditions for free will'). Some incompatibilists take the availability of a plurality of indeterminate possibilities to be a sufficient condition of free will to permit the ascription of moral responsibility. But Kane used a few examples from the past to refute these (Austin 1961)—and I agree with him. Namely, the case of the person who pressed the 'Cream' button to add cream to their coffee; the golfer who attempted to putt the ball into the hole but missed; the assassin who shot the bodyguard instead of the prime minister—in each case there are multiple possibilities and the event is undetermined; however, what actually occurs is not within the agent's undetermined voluntary control, for each event comes down to a twitch. According to Kane, the agent is only morally responsible if the outcome, of which they causally contributed to, aligned with what they *voluntarily willed* or *intended* to happen. If the agent wanted to have cream, miss the putt, or hit the bodyguard, however, they would have held rationally sourced and intended reasons which stem from their character and motives for the events. Thus moral responsibility can be ascribed when there are multiple, undetermined alternate possibilities available to an agent *and* when they can be held ultimately responsible for their conduct.[1]

By having neurological access to a plurality of such alternatives an agent faces 'tension-creating conflicts'. In Kane's view, which is popular amongst incompatibilist philosophers, such neural battles and complexity are necessary for the self-formation of a rational human agent. The neural pathway that succeeds above the indeterministic noise is the self-forming one.

Kane needs to do more

'…he does little to propose a solution which bridges the gap between this occasional 'randomness' and moral responsibility…'

Kane, however, fails to substantiate exactly how we are *ultimately* responsibly. This is the beginning of my argument against his theory.

Kane's 'soft' version of libertarianism does not elucidate why chance is so important. The possibility of multiple, conflicting motives does not mean that I make my choice of independent, free volition. I accept that some of my thoughts and actions *could* be undetermined—say, by virtue of some quantum neurological randomness in my brain—however, from this premise I cannot claim that I have ownership over these thoughts; that my will is autonomous. Gary Watson, too, posed the problem to Kane (Watson 1999, p355): '[S]oft libertarian views cannot give a proper account of the *significance* of indeterminacy.' (I refer to Watson's views again later.)

Kane asks us to endorse a negative condition of free will and moral responsibility (that there is an *absence* of external causation), leaving a vacuum of positive explanation as to why we should. This is something that Kane tries to deal with as he, again, invokes free action through the potential indeterminacy of motives, intentions, and so forth. My primary issue with these arguments is that he himself asserts that an agent's will cannot be set by anyone or anything else ('will-setting'). They must be ultimately responsible for their will and their undetermined actions: 'agents … must be responsible for their wills having been set that way—not God or fate or society of behavioral engineers or nature or upbringing' (Kane 2002, p412). Yet I cannot see how humans can 'set' their own wills. His statements echo what his theory is fundamentally lacking: reason to believe in the importance of randomness. Soft conditions say nothing of what pertains to me fundamentally. If, on a different timeline—e.g. 'God' reinitiated the Big Bang or because there is a parallel universe—the same agent faces the same task (e.g. *A* deliberated how to react to the bully), another

[1]The agent ought to know of any uncertainty, though. There was a chance of missing the prime minister, for example; so they are responsible for intending to kill *someone*.

outcome might manifest this time ('Challenge from Chance'). This casts doubt over the importance of self-forming actions since conflicting desires could amount to statistical expressions of the universe, not evidence of agency.

Another sticking point for Kane's theory—indeed, all compatibilist views—exists in the form of manipulation or *prevailing* determinism. If we had no say in 99.9 % of our actions (Berofsky 2006), indeterminacy would not protect our moral judgement from being undermined most of the time. Kane might argue that this external influence just changes how moral responsibility is distributed, not whether there is an ultimate source present; but the significance of his theory is again diminished either way.

Kane must also address the possibility that the Universe appears to be deterministic on a macro scale *all of the time*. Indeterminism has not been shown to pervade human cause-and-effect behaviour as we can only apply it on the atomic scale. The onus is still on incompatibilists, including Kane and myself, therefore, to demonstrate that it is true on a scale relevant to moral responsibility. Without a gauge on the extent of indeterminism, it is difficult to treat this negative-condition response as significant—we cannot scrutinise it—so, at this stage, the indeterminacy of our choices is philosophically contingent with respect to moral responsibility.

In breaking the causal chain between desire and choice by plugging in randomness we do not complete a theory of moral responsibility. Kane himself *recognises* this in 'The Intelligibility Question' as he warns of a 'vicious regress' of moral responsibility if we cannot delineate the sources of our wills as the 'power to be ultimate creators'. But, given his negative condition, he does little to *propose* a solution which bridges the gap between this occasional 'randomness' and moral responsibility which can definitely be made attributable to an agent.

Hard incompatibilism defeats Kane's libertarianism

'For an agent, A, to be morally responsible for an action, X, A must be the source of X.'

Thus my next logical next step is to offer a denial of free will and moral responsibility altogether, presupposing 'hard incompatibilism'. The implications of this are that we do not think or enact on our own accord and praise and blame hold no real meaning.

I have already put forward a case in which I challenge the coherence of Kane's theory of moral responsibility if we cannot be the *true sources* of our wills. His conditions of ultimate responsibility, I posit, are arbitrary because indeterminacy might be intrinsic to the Universe, not rooted to agenthood. Perhaps 'God', a separate agent, infused indeterministic decision-making into human psychology. Perhaps the laws of nature, set at the beginning of time and transpiring into evolution later, conditioned us to be this way. Either way, from governed rules I cannot see how an agent appears out of nothing. I put forward that we are not the sources of any soft, undetermined properties, as Kane stipulates (e.g. motives); hence hard incompatibilism, not libertarianism, at this point, seems inevitable.

To flesh out this point consider a human-like robot equipped with artificial intelligence. When faced with a moral dilemma it is sophisticatedly programmed to make a complex decision with an element of randomness. Is the robot morally responsible if it slowly built 'character' through the decisions it makes? We might claim: 'Yes! The robot is responsible for successive self-forming actions which are free from external causes and which reflect its emerging character and motives.' This conclusion, however, might appear absurd, for we programmed the robot. We know its *ultimate* functioning did not originally belong to it. Its actions are not causally determined, yes, but the randomness which dots its thinking cannot be reduced to an agent. Further, if anything, it is the programmer who is responsible, for they determined which paths the robot 'chooses' from. We might, alternatively, claim: 'No! The robot cannot be responsible, for there is nothing voluntary or reflective or purposeful in its actions as it was programmed to have this

indeterminacy.' This is more difficult to refute since it is a subject, not a controller, of its external environment. Now, replace 'robot' with 'human' and 'programmed' with 'conditioned by nature or "God"' (in Kane's arguments I cannot fathom a stance that distinguishes 'self-forming actions' of a human and a robot) and we see that humans are neither *sources* of free will nor do they perform actions that are truly voluntary because they are products, conditioned to be this way, undetermined or not, rendering the soft-libertarian case for free will and moral responsibility unconvincing and deficient.

It could be asserted that 'programmer' *and* the 'programmed' can each be ascribed partial responsibility. However, I would challenge this, for each entity's character and motives is contrived to the same, limited degree in the absence of an *ultimate source pertaining to the agent*. Gary Watson reckoned this to be an issue of 'ultimacy and uniqueness', claiming that Kane's 'tracing strategy is simply too weak to yield all we want' (Watson 1999, p359–360). An agent can hold multiple motives which are undetermined at any one time but who or what put the motives there in the first place? If we trace a human's or a robot's indeterministic self-forming actions back to the beginning, a source would be required (the Big Bang?) unless time and causation are spurious conceptions. When conflicting options are presented in an agent's head (or a robot's circuitry) there is little to conclude as 'self-forming' on the basis of disconnecting uncertainty surrounding a decision because, again, it does not reduce it to *them*. Thus, with a soft criterion met, we are still a step away from finding a source of an action and fail to evade the 'vicious regress' Kane warned of. I, like Watson, argue that any account of moral responsibility must be 'hard' to be at least theoretically qualifiable.[2]

To succinctly illustrate these points consider an empowered, female feminist, who would ordinarily be praised for their views. But is this fair, given that they themselves are conditioned by their environment? They grew up with an overbearingly, abusive, and protective brother and a suffocating mother who tried to force them into being a 'pretty girl' and rebelled equipped with notions of free will and justice. Should we totally or partially praise/blame the brother and mother for their maligned actions accordingly? What if these two were conditioned too? And so a regress is initiated. I dispute that there is such thing as 'power of originator' because 'sufficient reason' is notional. It can end but only when we can trace it to a truly ultimate cause. Until then there is nothing in the Universe that proves that we are capable of purposeful determination, meaning 'self-forming action' and 'sufficient reason' are fundamentally meaningless terms., for these principles are not *necessarily* attached to us.

Causation, in which we could play not part, might be nothing more than a conjunction of determined events and/or successions of undetermined events, on the part of the Universe or something greater than it: humans might only bear witness to events. When we recognise regular successions of events we project our own narratives—wills, intentions, motivations—onto the Universe. We might not be the sources of *anything* as things unravel, in orderly or disorderly fashion. This would spell the end of any notion of moral responsibility.

An agent who causes is morally responsible

'*Any theory that is adequate to explain the significance of indeterminacy must have a non-conjunctive structure … the will [must] be determined by the agent.*' *(Watson 1999, p356)*

My position on moral responsibility is that we cannot claim its truth unless humans can be positively shown to exhibit non-conjunctive, non-random powers, leaving hard incompatibilism the position with the most intellectual integrity, given the information available to us. There is

[2] Watson, however, endorses Frankfurt's view of hard compatibilist moral responsibility through identification and second-order volitions (Frankfurt 1969). I repudiate this position on the premises that nothing—desire, motive, idea, etc.—can cause itself deterministically, while Frankfurt's arguments do not offer a solution if human behaviour at the neurological level is indeed indeterministic.

no strong libertarian case which compels me to believe *we* determine anything at all—that is, unless humans are moral agents capable of *overriding* antecedent causal chains.

My stance is akin but not identical to Derk Pereboom's (Pereboom 2004), who claims that, in lieu of a convincing argument, we cannot coherently ascribe any form of moral responsibility. While Pereboom employs this 'piecemeal' approach to defeat cases for moral responsibility of all kinds, my stance focuses on the origins of responsibility. In my view it is crucial that moral choices are genuinely up to the agent in line with the principle of sourcehood (as opposed to the principle of alternate possibilities made by 'leeway incompatibilists', for example). Such a position cannot be posited for purely deterministic worlds.

If an agent, *themselves*, can determine their actions, they can be held morally responsible, for they determine events voluntarily according to their free will. Unlike in Kane's libertarianism, it is not enough to just break causal chains: we must start new ones (Fig. 3).

Fig. 3: An agent capable of performing undetermined action emerges from a world or unknown origins with genuine, non-conjunctive actions. The onus is now on the agent-causal theorist to formulate how.

The standard scientific worldview belongs to fundamentalism: the world, theoretically, can be explained through building blocks—descriptions of fundamental laws—which can be exposed by experimentally probing the Universe. What comes out are explanations—equations and descriptions—which reduce nature's complexity to simpler terms.

However, gaps in the reductionist chains have been hypothesised by philosophers of science such as Nancy Cartwright (Cartwright 1999). Cartwright expounds theories that ultimately cannot be described in the terms set out by the fundamental laws. For example, she described a banknote which did not fall according to Newton's second law, $F = ma$, because this law was not a fundamental, accessible explanation. So we take fundamental explanations (say, probably quantum mechanics) to be true only on good faith.

Her problem, then, is with fundamentalism. Fundamental laws are 'descriptions of what regularly happens, not regular associations or singular causings that occur with regularity' (Cartwright 1999, p4). These laws, by the very way they are constructed, do not describe anything outside their own explanations and conditions. Ontologically, the world could be underpinned by a 'dappled' explanatory patchwork, not a rigid framework of reductionist explanation that seeks to explain all behaviour with an all-encompassing pyramid. This is because no single grand unified theory can handle *all* physical situations from its explanatory base.

What does this mean for agents and moral responsibility? An agent could cause without being caused by something we can express in simpler scientific terms. This, in my mind, is the only defensible case of libertarianism. A possible hole in explanatory power at a level in the pyramid of Fig. 4 could lead to novel fundamental properties in biological organisms, such as us, in the form of agenthood ('emergentism').

Exactly what would constitute an agent is the remaining question (bundles of desires and motivations?). Pereboom scrutinises substance-causation as a potential contender, where an agent has free will and can cause, as a whole substance, and whose agency is irreducible to events occurring *within* them. However, because we are generally aware of our motivations but unaware of our neural and psychological causes, the notion of free will could stem from causally determined events too. This position, therefore, is ambiguous and uncompelling. Pereboom puts the most weight behind a form of Stoic theory, which postulates a rational, ruling soul ('hegemonikon') which has executive independence in all of its motivational states. For me, Stoicism is a metaphysically unfeasible theory without expansion on what the *hegemonikon* is.

We currently lack sufficient warrant to craft any agent-causal theory; however, we should assign the same of lack of warrant to negative cases against and remain agnostic.

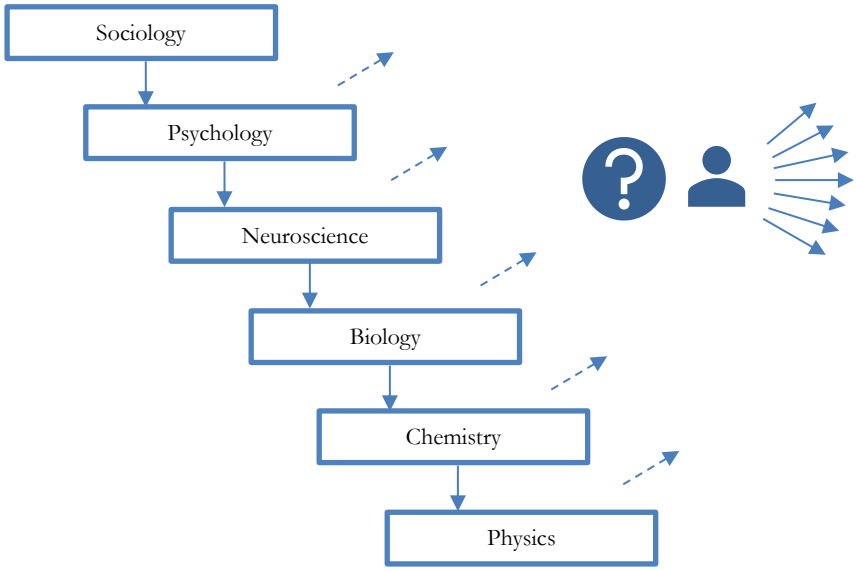

Fig. 4: Does agent-causation emerge (dashed lines) to supersede our current fundamentalist and reductionist picture? It is out of a hole in this patchwork that a liberation agent-causal theory can emerge: that we can genuinely be the sources of our undetermined wills in ways that cannot ever be illuminated in more-basic terms (e.g. by physics or neuroscience). Such causes of agency would have to non-reductively sit above fundamentalism's realm in discordance with dominant scientific thinking.

Kane 'disavows' agent-causation, suggesting it just as possible under a deterministic doctrine. I contest this: an agent could overrule deterministic states of the universe whereas his theory stakes little claim to agency because it is based on soft principles. Moreover, he says his event-causal theory amounts to an agent-causal theory anyway, claiming his agent does not 'disappear' because his agent *has* the capacity to produce, to cause, and to control (from their self-forming actions and so forth). In my opinion this does not take us any further than before; and, as such, I appeal to the explanatory power of agent-causation as a separate theory.

In conclusion

In this essay on libertarianism I have focused on Robert Kane's event-causal theory of free will due to its prominence in describing 'self-forming actions' to delineate moral responsibility. My view is that this is not defensible and that hard incompatibilism is a more-likely property of the Universe. I conclude that we need to transempirically *bring about* causation *directly* to be morally responsible agents. This is claimed to be metaphysically problematic and I am only tentatively committed to such an idea. Still, the search should go on. If we positively demonstrate the implausibility of agent-causation, hard incompatibilism reigns and there is no such thing as moral responsibility at all.

Bibliography

Austin, J L. 1961. *Philosophical Papers*, Oxford: Oxford University Press.

Berofsky, B. 2006. Global Control and Freedom, *Philosophical Studies*, 131(2), 419–445.

Cartwright, N. 1999. *The Dappled World. A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Frankfurt, H. 1969. Alternate Possibilities and Moral Responsibility, *The Journal of Philosophy*, 66(23), 829–839.

Kane, R. 2002. Some Neglected Pathways in the Free Will Labyrinth, *The Oxford Handbook of Free Will*, Oxford: Oxford University Press.

Pereboom, D. 2004. Is Our Conception of Agent-Causation Coherent?, *Philosophical Topics*, 32(1), 275–286.

Watson, G. 1999. Soft Libertarianism and Hard Compatibilism, *The Journal of Ethics*, 3(4), 351–365.