Compatibilism and Moral Responsibility

Antony Eagle

University of Adelaide
<antony.eagle@adelaide.edu.au>

Metaphysics » Lecture 11

Contents

Compatibilism and the Paradigm Case Argument

Three Arguments for Compatibilism

1. The **Paradigm Case Argument**: common usage describes many actions as 'free', and would do so even if determinism were correct; meaning is fixed by common usage; so the meaning of 'free' is some feature of actions that is consistent with determinism. Even if we found out determinism was true, still, the acts we label *free* would still be rightly so-called (van Inwagen 1983: §4.2).
2. The ***Mind* Argument**: free will and indeterminism are incompatible; free will exists; therefore, free will is consistent with determinism.
3. The **Conditional Argument**: ability ascriptions like *can* ϕ are disguised conditionals, perhaps of the form *if A has tried to* ϕ*, A would have* ϕ*-ed*; such a conditional may be true even if its antecedent is determined false; hence unexercised abilities are consistent with determinism (van Inwagen 1983: §4.3).

The Paradigm Case Argument

There are various words and phrases we use in ascribing free action to people: … 'acted freely' and 'did it of his own free will'… We learn these phrases by watching people apply them in concrete situations in everyday life, just as we learn, for example, colour words. These concrete situations serve as *paradigms* for the application of these words: the words mean *things of that sort*. Therefore they must apply … at least to the paradigmatic objects or situations. Careful investigation, philosophical or scientific, of these situations may indeed yield information about what freedom of choice really consists in, but it cannot show us that there is no such thing as freedom of choice. This is strictly parallel to the following proposition: careful investigation, philosophical or scientific, may show us what colour really consists in, but it cannot show us that there is no such thing as colour. (van Inwagen 1983: 107; cf. Flew 1955)

Supplementing the Paradigm Case Argument

- This is an argument that **free will exists**; how is it an argument for compatibilism?
  - Couldn't we use the Consequence Argument to establish the incompatibility of free will and determinism, then use the paradigm case argument to show that determinism is false?
- So something more is needed to get from the premise that we correctly call some actions *free* to the compatibility of freedom and determinism.
  - One such premise would be that determinism is true; but that would seem unnecesarily strong, since some compatibilists think the issue of determinism is **irrelevant** to free will, and so won't want determinism as a premise; and certainly no compatibilist is dialectially entitled to the premise, given that their libertarian opponents reject it.
- Taking up this irrelevance idea: perhaps the needed premise is that esoteric scientific hypotheses, like determinism, could not possibly undermine our ordinary usage, based as it is on regular observation:

> only … superficial [i.e., ordinarily observable] features of people and their acts … can be relevant to determining whether to apply the term 'free': one must be able *effectively* to compare some present act in which one is interested with the paradigms of free action … in order to see whether the present act is sufficiently similar … to be correctly called 'free'. (van Inwagen 1983: 110)

Meaning and Use

- This supplementary premise is motivated by the slogan that **meaning is use**: that the meanings of our expressions are fixed by the global pattern of **usage**; that differences in meaning can only arise when usage is different; and that usage can only be different when there is some difference language users can detect and respond to.
- That *green* means what it does is fixed by the fact that most English speakers, most of the time, use it to label surfaces within the same fairly natural range of visible surface reflectances, and don't use it to label other surfaces.
- Likewise for *free*. Common usage applies *free* to some actions, in contrast to those which are coerced, involuntary, etc. Those uses depend on speakers deploying those terms on the basis of overt properties of what they are talking about.
- We might thus attempt to run a paradigm case argument with this key premise:
  (Same Usage ⇒ Same Meaning) If there are two observably indistinguishable possible situations in which the total pattern of linguistic usage of terms in language ℒ is the same, then the meanings of terms in ℒ are the same.
- This premise, coupled with the possibility of a deterministic world in which our actual usage is determined to occur, will entail that in that world determinism is true and there are actions that are correctly in world called *free*. Hence compatibilism is correct.

Semantic Externalism

- The possibility premise establishes that in a deterministic world, speakers could establish a convention in which *free* applies to some actions.
- But we need (same usage ⇒ same meaning) to establish that their word *free* has the same meaning as our word. And that premise seems to be subject to counterexample.
- Consider **Twin Earth**. On Twin Earth, there is no H_2O, but there is an indistinguishable clear drinkable liquid called XYZ (Putnam 1973: 701). Speakers on Twin Earth use *water* to refer to XYZ, with the same pattern of usage as our uses of *water* to refer to H_2O.
- Once the facts about XYZ are revealed, it is correct to say
  (1) On Twin Earth the word *water* means XYZ.
  (2) On Earth the word *water* means H_2O.
- Since XYZ ≠ H_2O, these two uses have different meanings, despite the same use (remember that use is based on observable features, and Twin Earth **looks just like** Earth from the point of view of an external observer of both).
- It is usage **plus environment** which makes for meaning. The internal concepts that implement our usage aren't the meanings, it is the product of that usage in that environment which fixes meaning.

> Cut the pie any way you like, 'meanings' just ain't in the *head*! (Putnam 1973: 704)

Semantic Externalism and the Paradigm Case Argument

- Van Inwagen gives an example where the usage of *free* is the same, but – as in the Twin Earth case – there is some hidden environmental variability:

  (M) When any human being is born, the Martians implant in [their] brain a tiny device – one that is undetectable by any observational technique we have at our disposal… – which contains a 'program' for that person's entire life: whenever that person must make a decision, the device *causes* [them] to decide one way or the other according to the requirements of the [program] (van Inwagen 1983: 109).

- The world in which (M) is true is only unobservably different from the actual world, and the patterns of usage are the same. So in the (M)-world the term *free* has an extension.
- But in that world, the word *free* is only **superficially** like our word *free*. In fact, none of the actions called *free* in the (M)-world are in fact free by our standards.
- Since compatibilism about free will and Martian control **doesn't follow** from the existence of a non-empty extension for *free* in the (M)-world, **likewise**, compatibilism about about free will and determinism doesn't follow from the existence of a non-empty extension for *free* in some deterministic world.

The *Mind* Argument

Hume on Regularities in Action

> The internal principles and motives [of human nature] may operate in a uniform manner, notwithstanding these seeming irregularities; in the same manner as the winds, rain, clouds, and other variations of the weather are supposed to be governed by steady principles; though not easily discoverable by human sagacity and enquiry. (Hume 1777: §8, ¶15)
>
> [L]iberty, when applied to voluntary actions … cannot surely mean, that actions have so little connexion with motives, inclinations, and circumstances, that one does not follow with a certain degree of uniformity from the other, and that one affords no inference by which we can conclude the existence of the other. (Hume 1777: §8, ¶23)

- Some have used this observation as the basis of an argument for the incompatibility of free will and indeterminism (Hobart 1934; Smart 1961).
- I will offer a reconstruction of the argument.

The *Mind* Argument: Indeterminism Excludes Free Will

(3) If indeterminism is true, our bodily actions are not fully explained by our 'motives, inclinations, and circumstances'.

(4) If our bodily actions are not fully explained by our motives, inclinations and circumstances, then those bodily actions are not freely chosen by us.

(5) So, if indeterminism is true, our bodily actions are not freely chosen by us. (3, 4, logic)

- This argument is logically valid.
- Premise (3) is true, if a **full explanation** of p is at least explains the **contrastive** fact that p rather than $\neg p$.
  - Since we could have the same motives, inclinations, and circumstances in the absence of p if the link between those prior grounds and p is indeterministic, then indeterminism excludes a full explanation in this sense.
- So the focus should be on (4). Why think an **undetermined act** is problematic from the perspective of **free agency**?
- van Inwagen (1983: 128–52) considers three hypotheses; we'll take them in reverse order.

Hypothesis 3: We would lack control

- If my acts are the result of undetermined events, then I have no control over those undetermined events. But how can I **freely** do what I have no control over?
- **Reply:** this is to endorse the validity of rule (β); in which case, we must be endorsing the reasoning in the Third Formal Argument for incompatibilism.
- Of course, rule (β) is invalid, so this can't be a good ground to deny that probabilistically-caused acts can be free.

Hypothesis 2: It wouldn't be agency

> Anyone … can see right at the outset that a person whose 'acts' are the consequences of undetermined events … is not really an agent at all. Such a person is not acting, but is merely being pushed about or interfered with. (van Inwagen 1983: 134)

- Compare Hume (1777: §8, ¶23) again: we don't want the agent to have motives and inclinations and then turn out to be a **bystander** while they wait for the environment to determine whether the act eventuates.
- This idea is that my 'act' is just a lucky bodily movement; it's not even an act, so of course can't be a freely chosen act. (Compare: I can act to toss a coin, but not act to toss heads.)
- **Reply:** two ways an undetermined event can be a real act:
  1. **Agent causation** can accomodate an agent who causes the outcome while no event causes it (van Inwagen 1983: 135–37). (Yet suspicion of the coherence of agents causing things without participating in events that cause is a major objection to conventional libertarianism.)
  2. More metaphysically plausible: accept that **reasons are causes** (Davidson 1963), and deny that causes **necessitate**. In which case, an agent's reasons might cause their action without necessitating it; such an event would both be an act, subject to rational evaluation, and would be undetermined (van Inwagen 1983: 137–41).

Hypothesis 1: It would be chancy or random

- One idea is that what fills the explanatory gap between motives and action is something merely **probabilistic**: at best we could say that a certain proportion of cases where my motives are thus-and-such lead to an appropriate action.
- If so, my undetermined act would be an act of **chance** or **randomness**, and not freely chosen – the fact that it happened in my brain doesn't make it my decision.
  - Compare again Hume: 'liberty, when opposed to necessity, … is the same thing with chance' (Hume 1777: §8, ¶25).
- **Reply**: suppose my act did happen by chance. That does not mean it **only** happened by chance. It could have been caused **too** – again, **by my reasons**, in a probabilistic way.
  - Compare: I plan to go on holiday next week. If I am struck by a (chancy) meteorite tomorrow, I won't end up going. I do end up going because the meteorite strike **fails to eventuate**. My holiday is freely chosen, though there is some chance involvement in the full explanation: so (4) is false.
- Van Inwagen's own reply (1983: 128) is a bit weird: that the only thing he knows *random* might mean is something to do with long run frequencies, and single acts don't have long run frequencies. But that is a **failure of imagination**, not an argument (Eagle 2021).

What Survives of the *Mind* Argument?

> It should be unsurprising that the existence of quantum indeterminism, by itself, is inadequate to make the problem [of agency] disappear. It will certainly be insufficient … merely to replace a microphysically deterministic vision of the universe with a microphysically indeterministic one. For the problem … is as much about the way in which we suppose the different levels of reality relate one to another, as it is about the idea that each momentary state of the universe inexorably necessitates the next. … [Indeterminism] will not by itself supply the answer to the question how agency is possible. An answer to *that* question will require also an understanding of what could lead us to want to say that an *organism* rather than merely some part of one … has brought something about. (Steward 2012: 11)

- Steward's point is that the mere presence of indeterminism can't remove agency, nor the mere presence of determinism induce it: we need to know **how** a bodily movement was brought about in order to understand it as an act, as the product of agency.
- This emphasises a **residual task** for the incompatibilist: explain just how it is that free action fits into the (deterministic or indeterministic) laws of nature, given that it is not impossible.
  - Van Inwagen confesses he has no theory of free action (van Inwagen 1983: 150), but that's surely a significant outstanding task.

Abilities, Dispositions, and Conditionals

The Conditional Analysis of *Can/Could have*

- Another sort of Humean compatibilism: analyse the ability-attributing *can*, or its past tense *could have*, by means of a **conditional**, something like this:

  (COND) *A can ϕ (could have ϕ-ed)* means *If A tries (had tried) to ϕ, they will succeed (would have succeeded)*. (cf. Thomason 2005)

- This is close to what Hume calls 'hypothetical liberty':

  > *a power of acting or not acting, according to the determinations of the will*; that is, if we choose to remain at rest, we may; if we choose to move, we also may. (Hume 1777: §8, ¶23)

- No problem with determinism if this is correct: for even if it is determined that *A* won't try to ϕ, still it could be that, in the nearest possibility where they did try, they would have succeeded (the laws and past would have been such as to permit them to try and to succeed).
- Yet *can/could have* don't act like conditionals on the standard Kratzerian approach (Kratzer 1977, 1981).
  - Certainly they don't seem to act much like a conditional with an antecedent about trying: the view seems to force us to offer a distinct semantics for *Roger can flee* than for *the river can flood*, despite similarities in syntax.

Problems for the Conditional Analysis

- Moreover the proposal (COND) appears to be subject to direct counterexample:

  > Smith could have eaten one of the red candies.
  >
  > This proposition is not equivalent to
  >
  > If Smith had chosen to eat one of the red candies, then Smith would have eaten one of the red candies.
  >
  > For suppose that Smith is pathologically afraid of the sight of blood, and the candies are the colour of blood. Then it may well be that Smith was unable to *choose* to eat one of the red candies. And, in that case, he could not have eaten one… Nevertheless … if he *had* chosen to eat one of the candies, he would have. (van Inwagen 1983: 115–16)

- An instance of a general problem: Smith's inability to eat the candy is here linked to his inability to bring himself to choose to do so. The conditional is true; but it doesn't support the ability precisely because the ability consists in part of the agent's liberty to **make a certain choices**. Since Smith can't decide to **try** (not because of determinism, but because of his own pathology), he lacks the ability.

Dispositions and Abilities

- That abilities cannot be analysed in terms of conditionals only prompts the question: is there an account of abilities that is conformable to compatibilism?
- Vihvelin (2013: ch. 6) offers an account that is inspired by the conditional analysis without being subject to its problems.
- The conditional account of ability can be derived from two prior theses:

  Dispositional Abilities The idea that abilities are **dispositional** – that they are **powers** to manifest a certain response given an appropriate stimulus, rather than **categorical** properties.

  SCA The idea that the correct **analysis** of dispositions involves conditionals. E.g., a glass is **fragile** – possessing the disposition to readily break when struck – iff something like the following conditional holds: *if the glass were struck, it could/would easily break*.

- Put these together and we get the conditional analysis of ability (Vihvelin 2013: 196–97).

Dispositions and Conditionals

- But it was long ago noted in the dispositions literature that SCA, the simple conditional analysis, fails (Johnston 1992; Martin 1994). Consider

  > a fragile glass that is carefully protected by packing material. It is claimed that the glass is disposed to break when struck [i.e., is fragile] but, if struck, it wouldn't break thanks to the work of the packing material (Choi and Fara 2021: §1.2)

- In this case, the disposition is present but **masked**; the SCA predicts the conditional will be true, but it is not.
- There are also cases of the reverse sort, **finks** and **mimics**, in which the right sort of conditional is true but there is no disposition:

  > When a styrofoam dish is struck, it makes a distinctive sound. When the Hater of Styrofoam hears this sound, he comes and tears the dish apart by brute force. So, when the Hater is within earshot, styrofoam dishes are disposed to end up broken if struck. However, there is a certain direct and standard process whereby fragile things … break when struck, and the styrofoam dishes in the story are not at all disposed to undergo that process. (Lewis 1997: 153)

  - The case of Smith may also involve mimicry: Smith is not disposed to eat the candies, but if you could somehow get him to chose to, by some deviant route, he would eat them.

Abilities Without Conditionals

- Vihvelin endorses the idea of Dispositional Abilities:

  > To have one of the narrow abilities in virtue of which we are agents with free will is to have some intrinsic disposition or bundle of intrinsic dispositions. (Vihvelin 2013: 175)

- But Vihvelin rejects the simple conditional analysis. How then should we understand dispositions, and the abilities that are a special case of dispositions?
- Vihvelin suggests that we should follow the try-conditional account but without a conditional:

  > For a highly interesting subset of our narrow abilities, to have the narrow ability to do X is to have an intrinsic disposition to do X in response to the stimulus of one's *trying to do X*. (Vihvelin 2013: 175)

- Fara (2005) analyses the disposition ascription in terms of **habitual** constructions (like the habitual *Peter sings in the shower*, which seems to mean something like *Normally/typically, Peter sings in the shower*):

  > 'N is disposed to M when C' is true iff N has an intrinsic property in virtue of which it [habitually] Ms when C' (Fara 2005: 70)

Masked Ability Compatibilism

- Putting this together, the idea is that one has an ability just in case in typical or normal circumstances, one will ϕ when one tries to.
- This dispositions ascription isn't a conditional. But, like a conditional, it can be true even if we are determined not to be in the stimulus circumstance: that is, even if we are determined not to try to ϕ, we might still have the general tendency to succeed in ϕ-ing when we try.
- Fara (2008: 860–63) argues that in general the function of determinism is to **mask** the exercise of our abilities. Like the fragile glass being prevented from showing its tendency to break, the background conditions might prevent my abilities from manifesting in action, while I nevertheless retain those abilities – including the ability to do otherwise.
  - Understood in terms of masked abilities, Fara (2008: 862) argues that van Inwagen's rule (β) has **further counterexamples**: if Anna has the ability to write her name, but a bolt of lightning strikes her pen as she picks it up, then (i) no one can render it false that the lightning strikes; (ii) no one can render it false that the lightning prevents the inscription, but (iii) Anna has the ability to write her name, and can render it false that her name is uninscribed.

Dispositions Directly Involved

- Van Inwagen's objection is that compatibilists only secure the compatibility of ability and determinism by appealing to the conditional analysis of dispositions.
- But in fact **every** account of dispositions in the literature says that intrinsic dispositions can exist even under determinism because the truth of a **global** thesis like determinism is independent of the local intrinsic basis of the agent's ability.
- Vihvelin argues that in fact the **ordinary** use of *ability to* ϕ is (more or less) to have an intrinsic disposition to do ϕ when you try, and to be free from circumstances which remove your disposition (Vihvelin 2013: 192–96) – so-called **impediments**.
  - This notes that there are some environmental circumstances which remove abilities: like having my hands tied together behind my back removes my ability to scratch my nose.
- But determinism is not an **ordinary impediment** to the existence of our dispositionally-grounded abilities, even if it blocks the stimulus for such abilities to be exercised.
  - But see Vetter (2014) for an alternative account that links dispositions directly to *can* claims; such an account may be more susceptible to incompatibilist arguments.

Responsibility and Alternate Possibilities

Free Will and its Significance

- The importance of free will to our lives is twofold.
  1. We all **believe** that we are free, and indeed our self-image as deliberative rational creatures seems to require that we believe we are free (van Inwagen 1983: 153–61).
     - The importance of belief in freedom is dramatised in Chiang (2005).
  2. Our being **morally responsible** for our actions **requires that those actions are done freely**.
     - If someone does not perform an act freely, then whatever the moral status of the act, it does not morally reflect on the unfree actor.
     - '*Ought*, as the saying goes, implies *can*' (van Inwagen 1983: 161). If one **ought** to refrain from ϕ-ing, then it had better be that one **can** refrain from ϕ-ing.
- Can we turn these observations – particularly the second – into an explicit argument for the claim that **if we are morally responsible, then we have free will**?

An Argument from the Principle of Alternate Possibilities

PAP ‘A person is morally responsible for what he has done only if he could have done otherwise’ (van Inwagen 1983: 162).

(1) Someone could have done otherwise than they did only if they did it freely. (From van Inwagen’s definition of *free will* (1983: 8).)

(2) A person is morally responsible for what they have done only if they did it freely (1, PAP, logic)

(C) ‘If no one has free will, moral responsibility does not exist’ (van Inwagen 1983: 162). (2, contraposition, logic)

Moral Responsibility Sometimes, people are morally responsible for their actions. (premise)

Free Will Sometimes, people have free will. (2, MR, logic)

Counterexamples to the Principle of Alternate Possibilities

- Frankfurt (1969) offers a recipe for constructing **counterexamples** to (PAP).
- The general strategy is this: suppose *A* wishes to ϕ, and forms an intention to do so. *B* also wishes for *A* to ϕ, and surreptitiously installs a device in *A*'s brain so that, if *A*'s becomes unwilling to ϕ, *B* will cause *A* to ϕ by triggering the appropriate sequence of brain states by means of the device. As things turn out, though, *A* does not change their mind, *A* does ϕ, and *B* does nothing.
- In that case, *A* **is morally responsible**; the only difference from a clear case of moral responsibility is an irrelevant fact about a small device in *A*'s brain that never did anything.
- But *A* **could not have done otherwise**; no matter what, *A* will ϕ – whether in the 'normal' way, or via *B*'s intervention.

A Frankfurt Case

Suppose someone – Black, let us say – wants $Jones_4$ to perform a certain action. Black … waits until $Jones_4$ is about to make up his mind what to do, and he does nothing unless … $Jones_4$ is going to decide to do something other than what he [Black] wants him to do. If it does become clear that $Jones_4$ is going to decide to do something else, Black takes effective steps to ensure that $Jones_4$ … does do, what he wants him to do. Whatever $Jones_4$'s initial preferences and inclinations, then, Black will have his way.

Now suppose that Black never has to show his hand because $Jones_4$, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, $Jones_4$ will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. (Frankfurt 1969: 835–36; see also van Inwagen 1983: 162–63)

Causal Preemption and Counterfactual Analyses

- This is very similar to a class of counterexamples to **counterfactual theories of causation** (Lewis 1973).
- The simplest counterfactual analysis says: C caused E iff **if C had not occurred,** E **would not have occurred**.
- Let us say that E is **guaranteed** if E would have occurred no matter what, because if C hadn't occurred, another cause of E would have – a cause that the actual occurrence of C **preempts**.
- If E is guaranteed, the counterfactual *if C had not occurred, E would not have occurred* is false. The simple analysis concludes: C **didn't** cause E.
- But clearly there are preemption cases where C causes E.
  - Consider the case of *Student Assassin* and *Expert Assassin*: if Student doesn't kill Victim, the Expert Will. Accordingly, Victim would have died anyway. But that doesn't mean that it wasn't Student who killed him; after all, Expert never shot.

Causation and Responsibility

- Indeed, we can link preemption cases and Frankfurt cases.
- For if our action is guaranteed, then we could not have done otherwise, so we can reformulate the PAP as:

  PAP2 A person is morally responsible for an act only if what they did isn't guaranteed.

- PAP2 is false whenever there are Frankfurt cases: cases in which someone's decision caused their action, but their decision preempts an alternative cause that would have produced the same action, so the action is guaranteed.
- But, as in the scenarios Frankfurt presents, one can still be **morally and causally responsible** for a guaranteed outcome.

Frankfurt's Alternative

- Frankfurt notes than in the case of Jones$_4$, the fact they he couldn't have done otherwise 'played no role at all in leading him to act as he did' (Frankfurt 1969: 836), and this is important:

  > Suppose a person tells us that he did what he did because he was unable to do otherwise … We understand the person who offers the excuse to mean that he did what he did *only because* he was unable to do otherwise…. And we understand him to mean … that when he did what he did it was not because that was what he really wanted to do. The principle of alternate possibilities should thus be replaced … by the following principle: **a person is not morally responsible for what he has done if he did it only because he could not have done otherwise**. This principle does not appear to conflict with the view that moral responsibility is compatible with determinism. (Frankfurt 1969: 838, my emphasis)

- The argument for the existence of free will fails if (PAP) is false, and this alternative premise doesn't repair the damage – at least, not if we retain van Inwagen's idea that free will is linked intimately to the ability to do otherwise than we in fact do.

Frankfurt and Compatibilism

Classical Compatibilism and Semicompatibilism

- The traditional kind of compatibilist attempts to argue that agents have abilities to do otherwise than they do, even under determinism (Hume 1777; Lewis 1981; Fara 2008; Vihvelin 2013).
  - Such compatibilists **agree** with incompatibilists that moral responsibility requires alternate possibilities; they **disagree** only over the relevance of determinism to alternate possibilities.
- Frankfurt cases undermine a link between responsibility and alternate possibilities.
- One response to this is to focus directly on moral responsibility, rather than on free choice as a route to moral responsibility.
- This is the approach of **semicompatibilism**, the view that

  > determinism is compatible with moral responsibility, apart from whether causal determinism eliminates access to alternative possibilities. (Fischer 2012: 255)
- Semicompatibilism needn't take a stand on the traditional issue of free will.

Frankfurt Cases and Free Will

- Someone who accepts the semicompatibilist idea that determinism and responsibility are compatible, and also accepts van Inwagen's idea that freedom is necessary for responsibility, is led to endorse this conclusion: **free will cannot require the ability to do otherwise**.
  - This must involve a **redefinition** of *free will* from the stipulation that van Inwagen makes about it. But it is neither unmotivated nor dialetically inappropriate, because it follows from the Frankfurt cases, which are unambiguously about moral responsibility, and a principle connecting moral responsibility *in that sense* with free will.
- What would such a new species of compatibilism look like? We look first to Frankfurt's own discussion of his cases; if we can identify what Jones_4's responsibility depends on, then we might be able to identify the feature that grounds his freedom.
- Our intuitive response to the Frankfurt case is that Jones_4 is responsible because he did what he did because **he wanted to**. So, a first pass proposal:

  FC1 A's ϕ-ing is done freely iff A ϕ-ed because A wanted to ϕ.

Hierarchical Compatibilism (cf. McKenna and Coates 2021: §4.2)

- A counterexample to FC1: the **unwilling addict**. Consider the addict who

  > hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, helplessly violated by his own desires. (Frankfurt 1971: 12)

- The addict acts on his desires, but these desires at at odds with what he would **prefer his will to be**. In that sense, he acts 'helplessly', but not freely.
- If an agent's preferences about their will are aligned with what they will and how they act, then we aren't tempted to say the agent is the mere victim of their desires. In such a case, we think, the agent acts; they act from their desire; and their desire is one they fully endorse.
- This gives a **hierarchical** picture of the will: someone is free when their action-guiding desires ('volitions') align with the volitions they want to have (their **second-order** volitions).

Frankfurt's Compatibilism

> the statement that a person enjoys freedom of the will means (also roughly) that he [sic] is free to want what he wants to want. More precisely, it means that he is free to will what he wants to will, or to have the will he wants. Just as the question about the freedom of an agent's action has to do with whether it is the action he wants to perform, so the question about the freedom of his will has to do with whether it is the will he wants to have.
>
> It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will. And it is in the discrepancy between his will and his second-order volitions, or in his awareness that their coincidence is not his own doing but only a happy chance, that a person who does not have this freedom feels its lack. The unwilling addict's will is not free. This is shown by the fact that it is not the will he wants. (Frankfurt 1971: 15)

Frankfurtian Compatibilism **A's ϕ-ing is done freely iff A ϕ-ed because A wanted to ϕ and preferred that desire be active in guiding their action.**

PAP and Compatibilism

- The conception of freedom on offer in Frankfurtian Compatibilism concerns whether the agent endorses the **actual course of events**, not whether they controlled whether that or an alternative course of events came to pass (Sartorio 2016).
- Thus the adherent of Frankfurtian Compatibilism is an 'actual sequence' compatibilist: responsibility comes from features of how an action was produced or generated ('by' rather than 'through' the agent), whether or not the distant past guaranteed it would eventuate.
- Actual sequence compatibilism is typically motivated by Frankfurt cases: and many, compatibilist and incompatibilist alike, have thought Frankfurt's example **does not show** what he takes it to show.
  - For example, Vihvelin (2013: 91ff) argues that the correct (compatibilist) account of abilities entails that Jones_4 does have the ability to refrain from the action Black wants; it's just that, if he attempts to exercise this ability, Black's intervention will mask his ability.
  - For another example, note that Frankfurt's vignette appears to **presuppose** that there **are** alternate possibilities: viz., the possibility in which Black steps in and overrides Jones_4's intention: 'What was supposed to be a Frankfurt "no alternative possibilities story" not only does not entail determinism, but also, when understood in this way, is not even *consistent with* determinism' (Warfield 2007: 290).

Responding to Frankfurt Cases

Clarifying the PAP

- It is worth looking more closely at PAP. It's official formulation is 'A person is morally responsible for what he has done only if he could have done otherwise' (van Inwagen 1983: 162). This is a claim about the agent's **abilities**.
- Frankfurt's cases are those in which there is no possibility in which the agent **does otherwise**: either they do it of their own volition, or the intervention causes them to do it.
- These cases are a counterexample to PAP, in van Inwagen's formulation, only if something like the following principle **linking abilities and doings** is correct:

  Possible Exercise If S is able to ϕ, then there is a possible situation in which S exercises that ability to ϕ (cf. Spencer 2016: 465)

  - In a Frankfurt case, Jones cannot succeed in **acting** otherwise; Possible Exercise then entails that Jones lacks the **ability** to do otherwise; thus Jones could not have done otherwise, in the ability sense of *could have*; so if Jones is responsible, PAP is false.

Masked Abilities Again

- Possible Exercise is very intuitive. But it seems that according to the masked abilities variety of compatibilism, Possible Exercise is false.
- For Jones might well be such that, intrinsically, he has the disposition to refrain from the act.
- Of course, if he attempted to refrain – to **manifest** that disposition – he would not succeed: Black would intervene at that point.
  - In fact, his failure is unavoidable – that is why there is no possible situation in which he acts differently.
- But that unavoidable failure doesn't, according to this species of compatibilism, **remove** Jones' ability: it merely **masks** it.
  - In the same way that carefully packing a glass doesn't remove its fragility, though it does ensure that the glass will not manifest its fragility by breaking.
- Therefore Possible Exercise is false: Jones has the ability to do otherwise, but there is no possible situation where he does so; and PAP may therefore survive.

The Dialectic

Where does this leave us in the debate over free will? Two important points:

1. The masked ability theory was presented as a form of compatibilism; but it is quite possible for an incompatibilist to think that **abilities are dispositional**, and that in Frankfurt cases those abilities would be merely masked by Black's intervention.

   › The incompatibilist has to think that determinism is the kind of thing that removes abilities, but they may well argue that Black's potential manipulation is not **relevantly similar** to the way in which determinism and the distant past foreclose Jones' alternatives before they even arise.

2. The masked ability compatibilist can accept PAP; but they don't have to. It's perfectly possible to say that the hierarchical view is – independent of the question of determinism – just a **better** account of responsibility. The fact that Jones is intrinsically disposed to do otherwise is not especially illuminating of the **source** of his responsibility, which is more clearly disclosed in the fact that he managed to act in line with his intention.

Patching the Argument

- Van Inwagen takes a different response to Frankfurt cases. He accepts that PAP might be false; but (in line with the Frankfurtian compatibilist) continues to accept that moral responsibility requires free will.

> Let us suppose that the Principle of Alternate Possibilities is indeed false. What follows? It does not follow that we might be morally responsible for our acts even if we lacked free will; it follows only that the usual argument for the proposition that moral responsibility entails free will has a false premiss. But might there not be other arguments for this conclusion? Might there not be other premisses from which this entailment could be derived? (van Inwagen 1983: 164)

Principle of Possible Action

(PPA) A person is morally responsible for failing to perform a given act only if he could have performed that act. (van Inwagen 1983: 165)

- The PPA is not susceptible to Frankfurt-style counterexample.
- Suppose I witness a bank robbery in progress, and fail to call the police then. It's also true that my phone – unbeknownst to me – won't work, because the robbers have sabotaged all the local phone towers.
- Am I responsible for that omission? Perhaps I am responsible for for not **trying** to call, but

  > Am I responsible for failing to call the police? Of course not. I couldn't have called them. (van Inwagen 1983: 166)

  - Actually, is this so obvious? The example doesn't quite seem to be parallel to a Frankfurt case – let the robbers sabotage the wires only if the agent tries to call.

Preventing Events

- Van Inwagen thinks that we also need to pay attention to the **events** and **states of affairs** involved in action.

  (PPP1) A person is morally responsible for a certain event-particular only if he could have prevented it. (van Inwagen 1983: §5.5).

- For van Inwagen 'no event could have had causes other than its actual causes' (van Inwagen 1983: 170). This is a **fine-grained** theory of event identity.
  - This makes preemption impossible — it is not possible for an event to be guaranteed unless its actual causes are likewise guaranteed. **This is not plausible**.
- Accordingly, there will be no Frankfurt-style counterexamples to (PPP1), because the cases where Black intervenes will feature slightly different causes, and hence slightly different events are involved: so Jones could have prevented *e*, even though an event very similar to *e* would occur regardless.

Preventing States of Affairs

(PPP2) A person is morally responsible for a certain state of affairs only if (that state of affairs obtains and) he could have prevented it from obtaining. (van Inwagen 1983: §5.6)

› Again, van Inwagen argues, there are no Frankfurt-style counterexamples here: if a certain state of affairs obtains 'no matter what', then the agent is not responsible for it:

> because this state of affairs is a *universal*, it can be reached by various causal roads, some of them differing radically from the road that is in fact taken; and, in the cases we have imagined, *every* causal road that *any* choice of the agent's might set him upon leads to this same state of affairs. This is why the agent … always turns out not to be responsible for the state of affairs he is unable to prevent. (van Inwagen 1983: 176)

A Patched Argument

> If (i) no one is morally responsible for having failed to perform any act, *and* (ii) no one is morally responsible for any event, *and* (iii) no one is morally responsible for any state of affairs, then there is no such thing as moral responsibility (van Inwagen 1983: 181)
>
> If (i) someone could have performed … some act he did not in fact perform, *or* (ii) someone could have prevented some event that in fact occurred, *or* (iii) someone could have prevented some state of affairs that in fact obtains, then the free will thesis is true. (van Inwagen 1983: 182)

- These principles, with (PPA), (PPP1), and (PPP2), entail that moral responsibility requires freedom. The first does not mention 'performed acts' – the subject of (PAP) – as a separate category; yet no performed act isn't ultimately realised in or grounded in one of the other categories.

References

References

Chiang, Ted (2005) 'What's Expected of Us', *Nature* **436**: 150–50. doi:10.1038/436150a.

Choi, Sungho and Michael Fara (2021) 'Dispositions', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/dispositions/.

Davidson, Donald (1963) 'Actions, Reasons, and Causes', *The Journal of Philosophy* **60**: 685–700. doi:10.2307/2023177.

Eagle, Antony (2021) 'Chance versus Randomness', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. http://plato.stanford.edu/archives/spr2021/entries/chance-randomness/.

Fara, Michael (2005) 'Dispositions and Habituals', *Noûs* **39**: 43–82. doi:10.1111/j.0029-4624.2005.00493.x.

Fara, Michael (2008) 'Masked Abilities and Compatibilism', *Mind* **117**: 843–65. doi:10.1093/mind/fzn078.

Fischer, John Martin (2012) 'Frankfurt-Type Examples and Semicompatibilism: New Work', in *The Oxford Handbook of Free Will*: 242–65. Oxford University Press.

References (cont.)

Flew, Antony (1955) 'Divine Omnipotence and Human Freedom', in Antony Flew Alastair Macintyre, ed., *New Essays in Philosophical Theology*: 144–69. http://commonsenseatheism.com/uploads/Flew%20-%20Divine%20Omnipotence%20and%20Human%20Freedom.pdf.

Frankfurt, Harry G (1969) 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy* **66**: 829–39.

Frankfurt, Harry G (1971) 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy* **68**: 5–20. doi:10.2307/2024717.

Hobart, R E (1934) 'Free Will as Involving Determination and Inconceivable Without It', *Mind, New Series* **43**: 1–27. doi:10.1093/mind/XLIII.169.1.

Hume, David (1777/2022) *An Enquiry Concerning Human Understanding*, Amyas Merivale and Peter Millican, eds. https://davidhume.org/texts/e/.

Johnston, Mark (1992) 'How to Speak of the Colors', *Philosophical Studies* **68**: 221–63.

Kratzer, Angelika (1977) 'What "Must" and "Can" Must and Can Mean', *Linguistics and Philosophy* **1**: 337–55. doi:10.1007/BF00353453.

References (cont.)

Kratzer, Angelika (1981/2012) 'The Notional Category of Modality', in *Modals and Conditionals: New and Revised Perspectives*: 27–69. Oxford University Press.

Lewis, David (1973) 'Causation', *Journal of Philosophy* **70**: 556–67. doi:10.2307/2025310.

Lewis, David (1981) 'Causal Decision Theory', *Australasian Journal of Philosophy* **59**: 5–30. doi:10.1080/00048408112340011.

Lewis, David (1997) 'Finkish Dispositions', *The Philosophical Quarterly*.

Martin, C B (1994) 'Dispositions and Conditionals', *The Philosophical Quarterly* **44**: 1–8.

McKenna, Michael and D Justin Coates (2021) 'Compatibilism', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/compatibilism/.

Putnam, Hilary (1973) 'Meaning and Reference', *The Journal of Philosophy* **70**: 699–711. doi:10.2307/2025079?ref=search-gateway:09e42f33b1b5c89125844f838efc62da.

Sartorio, Carolina (2016) *Causation and Free Will*. Oxford University Press.

Smart, J J C (1961) 'Free-Will, Praise and Blame', *Mind, New Series* **70**: 291–306.

Spencer, Jack (2016) 'Able to Do the Impossible', *Mind* **126**: 466–97. doi:10.1093/mind/fzv183.

Steward, Helen (2012) *A Metaphysics for Freedom*. Oxford University Press.

References (cont.)

Thomason, Richmond H (2005) 'Ability, Action and Context'. http://www.eecs.umich.edu/~rthomaso/documents/action/ability.pdf.

van Inwagen, Peter (1983) *An Essay on Free Will*. Clarendon Press.

Vetter, Barbara (2014) 'Dispositions Without Conditionals', *Mind* **123**: 129–56. doi:10.1093/mind/fzu032.

Vihvelin, Kadri (2013) *Causes, Laws, and Free Will*. Oxford University Press.

Warfield, Ted A (2007) 'Metaphysical Compatibilism's Appropriation of Frankfurt', in Dean W Zimmerman, ed., *Oxford Studies in Metaphysics*, vol. **3**: 283–95. Oxford University Press.