Do We Have Free Will?

Antony Eagle

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

Metaphysics » Lecture 12

Contents

The Traditional Problem

The Traditional Problem

The strongest argument for the existence of free will…: moral responsibility requires free will and we *are* responsible for at least some of the things we have brought about. … Suppose there were arguments for determinism that were rationally more compelling than this argument for free will … then we should be rationally compelled to reject the thesis of the reality of moral responsibility. Such a conclusion might lead us to reopen the question of [compatibilism]. For … however plausible the premisses of our various arguments for incompatibilism may be, their denials are not nearly as implausible as the thesis that there is no such thing as moral responsibility. But [we should reopen the question] only if we can discover arguments for determinism whose premisses are more plausible than the premisses of our arguments for incompatibilism. (van Inwagen 1983: 188–89)

Evaluating the Traditional Problem

- According to van Inwagen, the dialectical situation **facing the incompatibilist** is this. We have arguments for
  1. The thesis that *there is moral responsibility only if there is free will*; and
  2. The thesis that *free will and indeterminism are incompatible*.
- If these arguments work, then incompatiblists are able to **accept** at most one of determinism and moral responsibility. And the only grounds for such a choice are **reasons** that favour one over the other.
- Van Inwagen thinks we have **stronger reason** to accept that there is moral responsibility than to accept that determinism is true; accordingly, he rejects determinism in favour of free will.
  - Note we also have the paradigm case argument, which supposedly establishes the existence of free will on the basis that our practice grounding the extension of the term *free will* is in good order. As van Inwagen (1983: 204) points out, this may not satisfy; for if it turns out that the Martian control hypothesis has always been correct, our actions would be correctly labelled *free*, but they would not be free in the way we hope or expect.

Compatibilism and the Traditional Problem

- The question of determinism isn't an existential threat to compatibilist free will, so this traditional problem isn't as pressing for the compatibilist.
- But compatibilists admit that there are prior circumstances which are incompatible with moral responsibility – circumstances where we are **coerced**, or caused to do something by external factors.
- So there is still an important question of articulating what features of the prior circumstances could **absolve** us of responsibility.
- Determinism itself isn't one of those factors, if compatibilism (of either sort) is correct. But it might have an indirect bearing on **how** our responsibility can be undermined.
  - For example, if determinism is true, then it is possible to set up prior causes that determine the absence of certain dispositions, or determine the absence of suitable higher-order volitions – in which case, determinism will ensure that certain of our actions lack the features needed for compatibilist freedom.
- That said, the focus in what follows will be squarely on the dialectical position for incompatibilists.

Arguments for Determinism

Arguments for Determinism: Modal Metaphysics

- One strand of argument for determinism is entirely *a priori*: it uses Leibniz' **Principle of Sufficient Reason** to show determinism.

  (PSR) For every [contingent] state of affairs that obtains, there is a sufficient reason for its obtaining. (Leibniz 1686: 321–22; van Inwagen 1983: 202)

- If determinism were false, then two possible worlds would exist, with a **shared past** and **divergent futures**. Consider the point in time immediately prior to the divergence; does any sufficient reason obtain at that point for one future rather than the other? It cannot, otherwise only one of those futures would be possible (the other being excluded by that factor). Hence the PSR is false.
- By contraposition, if the PSR is true, determinism is true.
- This argument isn't watertight (what if the needed sufficient reason comes after the time of divergence?)

The PSR and modal triviality

PSR must be rejected, for it has an absurd consequence: the collapse of all modal distinctions. (van Inwagen 1983: 202–3)

- We need two supplementary principles:
  (1) 'If x is a sufficient reason for y, then x must entail y'.
  (2) 'No *contingent* states of affairs may be its own sufficient reason'.
- Consider P, 'the conjunction of all contingently true propositions'. If PSR is true, P has a sufficient reason, S.
  - If S is **necessary**, it cannot entail P.
  - So, by principle (1), S must be **contingent**, hence a conjunct of P; which cannot be a sufficient reason for itself, by principle (2).
- So there can be no such S, **contrary to the PSR**. The only way out is to deny that there is any such P – but then, implausibly, **every** truth would be necessary (van Inwagen 1983: 203–4).

Scientific Arguments for Determinism

> There would seem to be two ways in which scientific evidence could convince us we are determined; first, we might believe this … on the basis of our most general physical theories (which apply to all physical systems and hence to us); secondly, we might believe this on the basis of the empirical study of man [sic]. But our most general physical theories are no longer deterministic. And the empirical study of man has a long way to go before it will be in a position to tell us anything about whether we are or are not determined. (van Inwagen 1983: 201–2)

- We turn to quantum mechanics below.
- First: do the human sciences, particularly psychology, give us grounds to think we are determined? Recall in this connection Hume:

  > the conjunction between motives and voluntary actions is as regular and uniform, as that between the cause and effect in any part of nature. (Hume 1777: §8.16)

Arguments for Determinism: Cognitive Science

- A prominent model in contemporary cognitive science is the **computational theory of mind** (Rescorla 2020).
- Putnam (1961) and Fodor (1975) both defend the idea that thinking is the **syntactic manipulation of mental representations**.
- If so, putative quantum indeterminism is neither here nor there: for the **operations of a computer are deterministic** (there are so-called 'non-deterministic Turing machines', but they cannot compute anything not able to be computed by a regular Turing machine – though they might be able to do it faster).
- So the computational theory of mind suggests that our mental lives – including our desires and decisions – arise from deterministic symbolic processes on mental representations; so we are 'for all practical purposes' determined (van Inwagen 1983: 198).
  - The compatibilist might be particularly interested in this point too – if it turned out that our mental deliberations were just the playing out of a deterministic computational process whose end-point was fixed prior to deliberation beginning, that might pose a threat even to compatibilist freedom.

Responding to the CTM

- Van Inwagen replies by saying 'we are too complicated for our internal workings to be surveyed like a watch', and hence, no evidence sufficient to establish that we are for all practical purposes determined 'ever has been collected, and, even if human beings are determined, it looks like being a very long time before anyone ever does collect it' (van Inwagen 1983: 201).
- The commitment to the CTM is **not empirical**; it is, explicitly in Fodor, a result in 'speculative psychology' (Fodor 1975: viii) – the working out of the general groundwork that any successful psychology must presuppose.
- I know little enough about these debates to adjudicate; but it does seem that the successes of the computational paradigm in **cognitive science** lend it more support than van Inwagen appears willing to grant.
- Yet how compelling have the successes of the CTM been in accounting for decision-making, as opposed to reasoning and inference?

Quantum Mechanics

Arguments for Determinism: Quantum Mechanics

The standard interpretation of quantum mechanics is indeterministic. (van Inwagen 1983: 191)

- Perhaps – it depends on what the 'standard interpretation' **is**: an issue over which there is a huge amount of controversy.
- An old, but very gentle introduction for philosophers is David Albert's *Quantum Mechanics and Experience* (1992). There are now a number of other excellent texts that are especially clear on the philosophical issues, though some have a narrow focus (Barrett 1999; Ney and Albert 2013; Lewis 2016; Maudlin 2019). As usual, the various entries on quantum mechanics in the *SEP* can generally be recommended – see especially the 'Related Entries' listed at the end of Ismael (2021).

Quantum Mechanics, Without Much of the Mathematics (Albert 1992: 30–43; Ismael 2021: §3)

- A **quantum system** has a **quantum state** at each moment, represented by a **wave-function** or ψ**-function**.
  - Technically, the state is a unit vector in a certain kind of vector space, called a Hilbert space. Vectors can be multiplied by numbers ('scalars'), resulting generally in a vector of a different length but pointing in the same direction.
- An **operator** $\mathcal{O}$ maps any state of the system $|A\rangle$ to yield a resulting vector: $\mathcal{O}|A\rangle = |A'\rangle$ (which may not be a unit vector, so may not be a quantum state).
- An **eigenstate** of an operator $\mathcal{O}$ is a state $|A\rangle$ such that $\mathcal{O}|A\rangle = a|A\rangle$.
  - Operators whose **eigenvalues** a are always real numbers correspond to **physical quantities**; if $\mathcal{O}|A\rangle = a|A\rangle$, then the **result** of an $\mathcal{O}$-measurement on a system in state $|A\rangle$ is a.
- The state evolves over time according to the **Schrödinger equation**. This evolution is **deterministic**: if the system is in state $|v\rangle$, for any other time t', the Schrödinger equation entails the existence of a unique **state** $|v'\rangle = \mathcal{U}_{t'}|v\rangle$.
- **And that's all**.

The Measurement Problem (Albert 1992: 73–79)

- The problem is, that doesn't **look** like all.
- Suppose we have a system with two possible measurement values for a given operator $\mathcal{O}$, a and b. That is, there are states $|A\rangle$ and $|B\rangle$, such that $\mathcal{O}|A\rangle = a|A\rangle$ and $\mathcal{O}|B\rangle = b|B\rangle$.
- We can have the following: we start in state $|v\rangle$, and, under the standard dynamics, end up in a **superposition**: a state

$$\mathcal{U}|v\rangle = |v'\rangle = \frac{1}{\sqrt{2}}|A\rangle + \frac{1}{\sqrt{2}}|B\rangle.$$

- But what we see experimentally is **not** a superposition.
- Rather, when a system is in this state $|v'\rangle$, and we perform an $\mathcal{O}$-measurement, the system ends up **either** in $|A\rangle$, with measurement outcome a, with probability $\frac{1}{2}$; **or** in $|B\rangle$/measurement outcome b with probability $\frac{1}{2}$.
- This probabilistic, indeterministic, evolution of the state $|v'\rangle$ into either $|A\rangle$ or $|B\rangle$ with some probability, rather than some distinctive evidence of a superposition of states, is the so-called **collapse of the wave function**.

What to Say About the Measurement Problem?

1. **Collapse** is a second law of **dynamical evolution**: normally the system follows the Schrödinger equation, except in measurement, where what happens is that the system 'jumps' indeterministically to an eigenstate of the particular quantity to be measured (Albert 1992: 80–116).
2. The Schrödinger dynamics are by themselves sufficient; but for some reason it looks like there is collapse. The most famous example of such a view is the **many worlds** theory or Everettian theory (Barrett 1999; Wallace 2012).
3. The Schrödinger dynamics are insufficient, but we don't have collapse: rather, **something else** supplements the theory.

i. One famous example is **Bohm's theory** (Albert 1992: 134–79), the theory that states **position** is always determinate, but has the cost of **non-locality** of cause and effect.
ii. Another example are **modal interpretations** (van Fraassen 1991): among which, apparently, is van Inwagen's own view.

Van Inwagen on Quantum Mechanics

Though quantum mechanics may be a deterministic theory, it seems to me to be wholly implausible to suppose that its truth entails determinism. This would be a plausible supposition if *all* the properties of a physical system at a moment were determined by its quantum-mechanical state at that moment. This, however, does not seem to be the case, since the properties a macroscopic object can actually be *observed* to possess at a given moment do not in general follow from the quantum-mechanical state of that object … at that moment. Rather, there is in general only a *statistical* correlation between an object's being in a certain quantum-mechanical state … and its possessing a determinate observable or measurable property. (van Inwagen 1983: 193)

Autonomy of the Observable

- Van Inwagen thus **rejects** this thesis, the **eigenvalue-eigenstate link**:

  (EEL) A system in state $|S\rangle$ determinately has physical property P only if $|S\rangle$ is an eigenstate of $\mathcal{O}_P$, the operator corresponding to P.

- Van Inwagen thinks that a system can be in a determinate macroscopic state **without** being in an eigenstate of that property, so can have a macro property even though there is only a 'statistical correlation' between the macro property and the microstate.
- Here is one thesis van Inwagen seems to be attracted to which violates (EEL):

  > we have properties that do not supervene upon the properties of the atoms that we consist of. (van Inwagen 1983: 217)

Quantum Mechanics and Determinism

- Wave-function-only theories, like the many-worlds view, are deterministic, as is Bohm's theory.
- Collapse theories are indeterministic, as is van Inwagen's own view – since the prior quantum state does not determine the posterior total state (though the dynamics of the quantum state are governed by the Schrödinger equation).

> Although it is not clear what the quantum theory implies about determinism, it is clear that the implications are potentially profound. Bringing the implications into sharper focus requires a simultaneous focusing of a host of other foundational issues, most especially concerning the nature of quantum magnitudes and the nature of the quantum measurement process. By now it is no surprise that pressing the question of determinism has helped to unearth the deepest and most difficult problems that challenge our understanding of the theory. (Earman 1986: 233)

The Argument for Free Will

The Argument from Responsibility

We all believe that responsibility exists. … if we examine our convictions honestly and seriously and carefully, we shall discover that we cannot believe that this assent is merely something forced upon us by our nature and the nature of human social life…. I think that we shall discover that we cannot but view our belief in moral responsibility as a justified belief, a belief that is simply not open to reasonable doubt. …

It is as adequate a defense of the free-will thesis as has ever been given for any philosophical position to say, 'Without free will, we should never be morally responsible for anything; and we are sometimes morally responsible'. (van Inwagen 1983: 209)

The Objection From Overstepping

Surely this cannot succeed! If this argument works, then armchair reflection on responsibility can secure the falsity of a purely scientific hypothesis about motion; this latter is surely an empirical thesis, and cannot be resolved by armchair techniques – the philosopher has overstepped their bounds.

- Response: The armchair thesis has scientific consequences only when combined with doctrines about the **completeness of physics** and the truth of **reductionism**. These doctrines are partly about physical particles, but too partly about **people** and other **ordinary objects**. What we already know about such things is obviously relevant to the truth of these reductionist doctrines. So the philosopher is well within their rights to use what they **know** about the familiar to argue for the falsity of certain ways of linking the familiar with the unfamiliar world of fundamental physics.

Backup Positions

- We know what van Inwagen's position is. But we also know something about **counterfactual van Inwagen**.
- If determinism does come back into scientific favour,

  I believe I would say that (β) was, after all, invalid. [And so he would reject the argument for incompatibilism.] … I have defended (β) entirely on *a priori* grounds. But it would not surprise me too much to find that this proposition, which at present seems to me to be a truth of reason, had been refuted by the progress of science. Such refutations have happened many times. And it does not follow from the fact that they have happened that there is anything wrong with accepting on *a priori* grounds a principle that later turns out to be empirically refutable. One must simply realize that *a priori* convictions are as corrigible as any others. (van Inwagen 1983: 219–21)

Last Words

The more precisely science locates man in nature the more difficult it becomes to sustain a sense of autonomy for human actions. As autonomy shrinks so does our sense of uniqueness and worth as well as the basis for a moral perspective on human action. As I have tried to indicate, this difficulty would arise even if the ultimate laws of nature proved to be non-deterministic, but since determinism poses the difficulty in its sharpest form it is appropriate to continue to speak of the determinism-free will problem. As a practical 'solution' I recommend the ostrich tactic: don't think too closely or too long on the issues raised here, and in daily life continue with the presumption that the 'I' that chooses and the self to which we attach value judgments are autonomous. Let those who want to call themselves philosophers bear the risk to their mental health that comes from thinking too much about free will. (Earman 1986: 249–50)

References

References

Albert, David Z (1992) *Quantum Mechanics and Experience*. Harvard University Press.

Barrett, Jeffrey A (1999) *The Quantum Mechanics of Minds and Worlds*. Oxford University Press.

Earman, John (1986) *A Primer on Determinism*, vol. 32. D. Reidel.

Fodor, Jerry A (1975) *The Language of Thought*. Harvard University Press.

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

Ismael, Jenann (2021) 'Quantum Mechanics', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/qm/.

Leibniz, Gottfried Wilhelm (1686/1989) *Philosophical Essays*, Daniel Garber and Roger Ariew, eds. Hackett.

Lewis, Peter J (2016) *Quantum Ontology*. Oxford University Press.

Maudlin, Tim (2019) *Philosophy of Physics: Quantum Theory*. Princeton University Press.

References (cont.)

Ney, Alyssa and David Z Albert, eds. (2013) *The Wave Function*. Oxford University Press. http://www.worldcat.org/title/wave-function-essays-on-the-metaphysics-of-quantum-mechanics/oclc/793497136.

Putnam, Hilary (1961/1980) 'Brains and Behavior', in Ned Block, ed., *Readings in Philosophy of Psychology*, vol. **1**: 24–36. Harvard University Press.

Rescorla, Michael (2020) 'The Computational Theory of Mind', in Edward N Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/computational-mind/.

van Fraassen, Bas C (1991) *Quantum Mechanics: An Empiricist View*. Oxford University Press.

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

Wallace, David (2012) *The Emergent Multiverse*. Oxford University Press. http://tinyurl.com/emergent-multiverse.