

The Consequences of Incompatibilism

Forthcoming in *The Routledge Handbook of Responsibility*, ed. M. Kiener.

There are several well-known pitfalls when it comes to the task of characterizing *incompatibilism* concerning responsibility and determinism, pitfalls that are primarily associated with further characterizing the two theses at stake in the alleged incompatibility. In particular: What is *determinism*? And what is *moral responsibility*? Both questions have been the subject of vigorous debate. On the first question, P.F. Strawson went so far as to begin his familiar essay on the compatibility problem – “Freedom and Resentment” – by claiming that he doesn’t know what the thesis of determinism is meant to be in the first place.¹ On this question, however, the standard – albeit perhaps problematic – characterization in the responsibility literature has been some variant of the following: determinism is the thesis that the past and the laws of nature entail one unique future.² On the second question, there is vastly more to say than I could say even in a short book. However, at least since Strawson’s same article, it has been common (although certainly not universal) to characterize moral responsibility in terms of the fairness or desert or propriety of the *reactive attitudes*, especially the key attitudes of resentment and indignation.

This essay is about certain ways of misconstruing incompatibilism. However, this essay is not about problems that arise in connection with understanding the two terms of the relevant incompatibility. In this essay, I address what I regard to be ways of misconstruing incompatibilism that arise in connection, so to speak, with the “incompatibilism” part of incompatibilism – that is, in connection with the *relationship* envisaged between the two given theses, whatever precisely they are. Most philosophers, of course, know the strict definition of incompatibilism: that there is no possible world in which anyone is both morally responsible (in the sense at stake) and determined (in the sense at stake). More generally: p and q are incompatible if and only if there is no possible world in which both p and q . If we limit our attention to the characterization of responsibility in terms of the fairness/desert/propriety of resentment and indignation, as I shall in this essay, the strict definition has it that there is no possible world in which anyone is both a fair/deserving/appropriate target of resentment and indignation, and yet also determined. Incompatibilism, however, is often presented directly in

¹ Strawson 1962.

² This sort of “entailment-based” characterization has been standard at least since van Inwagen 1983. However, see Helen Steward (2021) for objections to this characterization, and an alternative.

terms of its alleged *consequences*. My focus in this essay is on what I take to be an implicit argument operating in the background of much of the literature on responsibility, viz.:

The strict modal thesis. There is no world in which anyone is both determined and a deserving target of the reactive attitudes. *Therefore,*

The consequence. If we found out that determinism is true, we would have to give up the reactive attitudes.

More generally, my sense is that many philosophers regard this consequence as so obvious so as not to be worth mentioning – so obvious that we can simply and directly present incompatibilism as the second thesis, the thesis that if we found out that determinism is true, we would have to give up the reactive attitudes. For instance, Gary Watson writes that “incompatibilists insist that the truth of determinism would require us to take the objective attitude [i.e., give up resentment and indignation] universally.”³ My claim is that this is a serious mistake, and that there is an attractive, neglected form of incompatibilism that denies the consequence.

There are, I contend, two distinct ways in which the incompatibilist might deny the consequence, corresponding to which of the two premises in the following argument he or she denies:

- (1) If we found out determinism is true, we would find out that no one deserves the reactive attitudes.
- (2) If we found out that no one deserves the reactive attitudes, we would have to give up the reactive attitudes. So,
- (3) If we found out determinism is true, we would have to give up the reactive attitudes.

The *flip-flopping incompatibilist* denies (1), and what I propose to call the *innocent incompatibilist* denies (2). I investigate each option in turn, although I focus most of my attention on a denial of (2).

³ Watson 1987, in Watson 2004: 225. Cf. Fischer and Ravizza 1998: 253, who contrast their view with the incompatibilist view as follows: “We would *not* have to withhold our attributions of responsibility ... if we were convinced ... that the universe is – surprisingly – causally deterministic.”

Incompatibilities and Indicatives

I begin with a puzzle about the indicative conditional, which I shall represent throughout as ‘if p , q ’. Consider an example. Consider an atheist who accepts, on *a priori* grounds, the standard argument from evil, according to which the existence of God is incompatible with the existence of evil. That is, this atheist is a God/evil incompatibilist: there is no world in which both God exists and evil exists. Since this atheist takes it as obvious *a posteriori* that evil *does* exist, this atheist concludes, of course, that God does not. Now the puzzle. Is this atheist committed to the truth of the following indicative conditional?

(4) If God exists, then there is no evil.

Not obviously. Indeed, if anything, it is obvious that our atheist *will not* be willing to accept (4). That is, the atheist who accepts the standard argument from evil is not likely to think that if God actually does exist, then there is no evil. Instead, this atheist is likely to think that if God actually does exist, then her *a priori* argument from evil is somewhere mistaken, even if she can’t say where. In other words, the atheist is likely to accept that if God actually does exist – has existed this whole time – then *of course* there is still evil; it is just that God and evil are after all somehow compatible. So this is the puzzle. One can rationally accept that there is no possible world in which both p and q , and yet *not* accept (the indicative conditional) that if p , $\sim q$. This result is puzzling, but it is compelling – and it is even more compelling on a theory that links belief in an indicative conditional ‘if p , q ’ to a disposition to infer q on accepting p .⁴ In particular, the relevant atheist probably isn’t disposed to conclude that there is no evil on coming to accept that God actually exists, but instead to conclude that the argument from evil was somewhere mistaken. The general point here is the following. One can accept on *a priori* grounds that there is no world in which both p and q . However, it may nevertheless be the case that one is *not*

⁴ Cf. Ramsey (1931: 249) and the “Ramsey Test”, or Mellor (1993: 236): “‘If P, Q’ . . . expresses a disposition to infer Q from P. In other words, fully to accept a simple ‘If P, Q’ is to be disposed fully to believe Q if I fully believe P.” For a recent development and extension of a view of this kind, see Khoo 2022. Note: I am assuming that the indicative conditional is not just a material conditional “ $p \supset q$ ” (i.e. “ $\sim p \vee q$ ”). The atheist of course accepts “God doesn’t exist, or there is no evil” – the atheist accepts that God doesn’t exist.

disposed to reject q on finding out that the *actual world* is a p -world; instead, finding out that p may lead one to reject the incompatibility of p and q .

The upshot here is perhaps obvious. The incompatibilist about moral responsibility and determinism can certainly believe that there is no world in which anyone both deserves the reactive attitudes and is determined. And yet she may deny premise (1) above, that *if we found out that determinism is true, then we would find out that no one deserves the reactive attitudes*. Consider van Inwagen's well-known stance that if he found out determinism is true, then he would reject his argument that free will (understood as the freedom to do otherwise) is incompatible with determinism.⁵ Fischer has protested that van Inwagen's position is somehow unstable, but his objections can, I believe, be countered.⁶ More generally, however, the God/evil example shows, I believe, that there is *in principle* nothing problematic about believing that there is no world in which both p and q , and yet *not* accepting the indicative 'if p , $\sim q$ '. And this is what matters.⁷

But perhaps we should momentarily back up. The suggestion here is that some incompatibilists will accept that there is no world in which anyone is both responsible and determined, and yet will *not* accept that if determinism is actually true, no one is responsible. Now, which incompatibilists will take this sort of line? Plainly, it is only the incompatibilist who is also a *libertarian* – that is, who believes that we are in fact responsible. In point of fact, it is the libertarian who ranks the claim *that we are morally responsible* above the claim *that responsibility is incompatible with determinism* in terms of overall plausibility. In other words, we can distinguish between two different types of libertarian: *resolute responsibility theorists*, and *resolute incompatibility theorists*. The latter are such that if they found out that determinism is in fact true, they would conclude that we are not responsible. The former, however, are such that if they found out that determinism is in fact true, they would conclude that incompatibilism is false – and so we still are responsible.

My point here is simple. To present incompatibilism directly as the view that *if we found out that determinism is true* – or, more simply, that if determinism is true – *then we would have to give up the reactive attitudes* is to neglect, without consideration, the views of the libertarians who are resolute responsibility theorists. But this is to say that it is no part of *what it is* for incompatibilism to be true for the relevant practical thesis to be true.

⁵ Van Inwagen 1983: 221.

⁶ Fischer 2016; Bailey and Seymour 2021, Todd and Rabern forthcoming.

⁷ Another example: dualism. Imagine characterizing dualism – i.e., consciousness and “everything is physical” incompatibilism – as follows: “The dualist contends that if we found out [e.g., from an oracle] that everything is physical, we’d have to conclude that there is no pain.” Or: “According to the dualist, if everything is physical, there is no pain.” Both characterizations are plainly ridiculous. According to the dualist, if everything is physical, she is totally mistaken about the prerequisites of pain.

A final point. Consider another salient property commonly attributed to indicative conditionals: the assertion of an indicative ‘if p , q ’ presupposes that p is a live epistemic possibility. For instance, consider the following:

(5) See, Anders isn’t in his office. If Anders is in his office, the lights are on.

(5) seems infelicitous. Why? Short story far too short: ‘if p , q ’ presupposes that p is epistemically possible.⁸ Now the point. Consider:

(6) If determinism is true, then no one is responsible.

What will the libertarian make of (6)? Prima facie, she shouldn’t *accept* (6), for (6) carries the presupposition that determinism could be true – which the libertarian will reject. Of course, the libertarian could *accommodate* the antecedent in (6); she could say, “Of course, determinism isn’t true. But look, I’m fallible, so if I’m wrong about that, and determinism *is* true, then ...”. But then there is no reason why she couldn’t add, “... then the most natural thing to think is that responsibility is somehow compatible with determinism.” Thus, the libertarian needn’t accept (6). Because she needn’t accept (6), she needn’t accept (1), and thus needn’t accept (3) – i.e., the consequence. But this is to say that the strict modal thesis does not entail the consequence.

Denying (2): Innocent Incompatibilism

Recall the key argument from above:

(1) If we found out determinism is true, we would find out that no one deserves the reactive attitudes.

(2) If we found out that no one deserves the reactive attitudes, we would have to give up the reactive attitudes. So,

(3) If we found out determinism is true, we would have to give up the reactive attitudes.

⁸ Stalnaker 1975. For recent discussion, see Holguín 2020.

We have briefly considered a way in which some incompatibilists will deny (1). But what about (2)? The incompatibilist who rejects (2) may (or may not) admit that to find out that determinism is true would be to find out that no one is responsible. However, she denies that this would be to find out that we thereby have a *requirement*: to stop treating people as responsible – that is, to refrain from engaging in the reactive attitudes.

But first we need to back up and slow down. Obviously, (2) is importantly ambiguous. To whom does “we” refer here, or to whom does “us” refer when Watson writes that “incompatibilists insist that the truth of determinism would require us to take the objective attitude [i.e., give up resentment and indignation] universally”? I thus proceed in two stages. I first consider a reading of this conditional on which *the individual alone* finds out that determinism is true. I then consider an interpretation on which some relevant *community* finds out that determinism is true. First, then, consider the following (from the perspective of an incompatibilist):

(3*) If an oracle told me that determinism is true, then I would have to give up the reactive attitudes.

(Let’s simply grant that the oracle is reliable, and not worry about how we know this, or how the oracle knows what he knows. The point of the oracle is simply to allow us to bypass questions about the mechanism of discovery.⁹) For my own part, *even speaking as an incompatibilist*, I am inclined to reject (3*). Below I consider several different reasons for this conclusion. But first, what is the *prima facie* case for (3*)? The thought appears simple: if an oracle told me that determinism is true, I would thereby come to know that no one is blameworthy (deserves the reactive attitudes), and, necessarily, it is morally wrong to (knowingly) blame those who aren’t blameworthy. In other words, the thought appears to be something like:

Wrong to blame the blameless. Necessarily, it is morally wrong to knowingly blame those who aren’t blameworthy.

But matters are not so simple. Again, why should even an incompatibilist doubt (3*)? Several (sometimes overlapping) considerations appear relevant.

⁹ I don’t mean to be dismissive of such questions; indeed, there are highly non-trivial questions about how an “oracle” could know any such thing, or why we would have reason to *believe* this oracle, even if the oracle *did* know such a thing. Cf. Hawthorne’s (2002) appeal to an oracle in the context of debates about dualism.

(a) *Inescapability*. First of all, am I *able* to give up the reactive attitudes? Is it really possible for me, as I actually am, to cease becoming angry and resentful, at least some of the time, while still maintaining my usual day to day relationships – that is, whilst not hermetically isolating myself, or (worse) taking my own life? Perhaps – but this isn't obvious. Needless to say, it is one thing to accept the thesis that no one is blameworthy; it is something entirely else to live moment by moment as if this is true, and to regulate one's emotional life accordingly. However, it is at least very plausible that I would *have to* give up the reactive attitudes only if I am *able* to give up the reactive attitudes. Note: the ability here needn't be anything so strong as the incompatibilist's free-will ability; all that is required is the idea that I *have to* do something only if that thing is appropriately sensitive to my *trying to* do that thing. And it seems questionable that I *could* give up the reactive attitudes, *even if I tried*.

(b) *Difficulty*. However, let me simply grant that I *could* – with sufficient effort – train myself in the relevant way. Even still, it should be uncontroversial that it would at least be *extremely difficult* for me to train myself in such a way that I never blame those around me, to adopt what we earlier called “the objective attitude” at all times. But now the claim that I *must* do this seems to be seriously doubtful. In particular, the claim that I must do this now must be weighed against *other* claims on my time and attention – in other words, it must be weighed in terms of the gains and losses both to my own life and to the lives of those around me. Let's imagine, for purposes of illustration, that I could train myself in the relevant way only by devoting myself to a certain program for three hours a day for three months. (Perhaps this is unrealistic, but it isn't clear what *is* realistic.) But let's also suppose that I am heart surgeon with a busy schedule, and I certainly can't take three hours off per day for three months, at least not without alienating my confused family and friends (who would regard my quest as at best quixotic, if not wholly absurd and misguided), risking my job, jeopardizing the health of my patients, and seriously inconveniencing many others along the way. Well, *must* I do this, for the sake avoiding becoming resentful and indignant with people who I have come to discover do not merit that resentment and indignation? Not obviously. Indeed, this is not obvious even if I am not a heart surgeon, but instead, perhaps, a run-of-the-mill academic with the merely *ordinary* demands on his time and attention.

The point here is a simple one: whatever effort of will I expend on the project of reforming my basic disposition to have the reactive attitudes is effort I cannot spend elsewhere, on *other* things that matter – and perhaps matter much more. This point is naturally related to a

further one. Suppose you had a choice about which of two projects I pursue: (a) spending 30 hours reforming my disposition to blame, so that I less frequently (or perhaps never) blame the blameless, or (b) spending 30 hours at the soup kitchen, or visiting lonely elderly people at a nursing home, or cleaning the polluted river, or ... pick your favorite cause. First, it isn't obvious that the goods here are even commensurable, and second, supposing that they *are* commensurable, it isn't clear that the goods at issue in (a) outweigh the goods at issue in (b). For my own part, for most ordinary people in the kinds of circumstances with which I am most familiar, there is a whole range of activities that I would rather see those people devote their time to, rather than the project of eliminating their dispositions to blame, even on the assumption that those they blame are never in fact blameworthy.

(c) *Harms and benefits.* One salient consideration involved in this judgment is the question of *who is harmed* by my blame, if anyone is harmed at all. If my blaming someone is seriously harmful, and yet she does not *deserve* that harm, then perhaps there is very strong reason against my blaming that person – enough reason to make it a requirement to regulate myself in such a way that I do not do so. First, however, it seems fairly obvious that my hostile blaming attitudes towards Vladimir Putin do not harm, in any straightforward way, Vladimir Putin. Thus, if we are going to find plausible subjects who are non-trivially harmed by my blame, this will have to be subjects with whom I have some kind of personal relationship. Consider, then, a mundane case: my wife sleeps late yet again, hurries out the door, and leaves me with a mess to clean up. In other words: mundane grist for resentment's mill. Suppose I resent her to the appropriate degree and in the appropriate way (on the assumption that she is blameworthy). Is she *harmed* by this resentment, in itself? The question is a difficult one, and one I am not sure how to resolve. But it is at least not obvious that she *is* harmed, or harmed very substantially. Consider another case. My friend Neal promises to pick me up from the airport, and then forgets. I am naturally annoyed and angry. But then suppose I suddenly pass away from a heart attack, standing there waiting for him, in my annoyed and angered condition. Was Neal *harmed* by my harboring these attitudes? Not obviously. Of course, if I survive, Neal might be harmed by a decision of mine no longer to trust him, or no longer to be his friend, on account of his manifest unreliability. But then what is harming him here is not my *blame* – my hostile attitudes in themselves – but instead something else, something that was in any case liable to happen *without* the hostile attitude of resentment. But this is simply a different matter. In other words, in many cases in which it may appear that resentment is harmful, the resentment is merely epiphenomenal with respect to what actually causes the harm.

By way of summing up the above considerations, consider an analogy. Suppose that due to natural facts about me, I am deathly afraid of spiders – not *poisonous* spiders, mind you, but the harmless variety in my vicinity. An oracle then gives me a convincing demonstration that the local spiders are in fact perfectly harmless, and thus reveals to me that my *fearing* those spiders is inappropriate: my *fear* represents them as *fearsome* – worthy of fear – and what I have discovered is that these spiders do *not* merit this attitude: they are not in fact worthy of fear at all. Does this discovery then impose on me a duty to rid myself of this fear? Well, not obviously. First, I may not really be able to overcome the fear anyway, even knowing that it mistakenly represents its object as fearsome; second, even if I could do this, it might be very difficult, and frankly I might have better things to do. Finally, the spiders are not substantially harmed by my fearing them when they don't deserve it. It is thus unclear why the discovery of the fact that the spiders are not fearsome imposes on me an all things considered duty to rid myself of this fear; perhaps I should just recognize this unfortunate fact – that what I fear doesn't merit it – and *carry on*. *Mutatis mutandis*, what we've said about fear likewise goes for blame. To support the key claim from above, we need to be told why the cases are different.¹⁰

The considerations I have been mentioning are all fairly ordinary considerations commonly thought to bear on whether and when I have a duty. Before moving on, however, I want to mention at least one decidedly more speculative source of concern about the key claim, that if I found out that no one deserves the reactive attitudes, I would have to give up the reactive attitudes.

(d) *Demandingness and Authority*.

If incompatibilism is true and yet determinism is true, then “from the start”, as it were, God, the gods, or nature has given us a set of dispositions and practices that presupposed for their fairness something that was false. But what this then suggests is that *fairness* (in this sense) is somehow explanatorily disconnected from what we are like; in other words, there are facts about

¹⁰ Many readers will have plainly noticed that I have adopted these considerations – slightly modified – from P.F. Strawson's “Freedom and Resentment”. Strawson therein suggests that (a) the reactive attitudes are inescapable for us, but that (b) even if we *did* have a choice in this matter, which we don't, this choice could only be made in light of the “gains and losses” to human life. These arguments have universally been taken to be in the service of compatibilism, and commentators have tried – in my opinion in vain – to see in them some way towards a compatibilist result. My argument here is that these points have no tendency whatsoever to support compatibilism; what they show, if they show anything, is that incompatibilism is practically irrelevant. But the thesis that incompatibilism is practically irrelevant must be distinguished from the thesis that it is false.

fairness, but those facts play no explanatory role in what reality is actually, concretely like. But this then suggests that these facts lack the right kind of relationship to us in order to be fully authoritative for us. *Prima facie*, if someone makes a *demand* of us, the normative force of this demand is a function of at least two things: the *authority* of the one who demands, and the *costs* involved in complying. Roughly: the greater the authority, the more ability to impose more demanding demands; the less demanding the demand, the less authority is needed for the demand to be authoritative. Now, the demand to refrain from the reactive attitudes may be extremely demanding, and impose on me/us serious costs. But when the moral standards that make this demand are themselves shallow in the structure of the universe (cf. Mavrodes 1986) they lack substantial authority, and specifically the authority needed to authoritatively demand that we reform our basic dispositions at substantial cost to ourselves and others. This would not necessarily be to say that incompatibilism is false; rather, it is to say that the moral facts that would make it true would also make a certain kind of moral rationalism false, a moral rationalism on which the moral point of view is always all-things-considered authoritative for us. Perhaps the concern for fairness is a consideration I am not *morally* permitted to ignore, but am nevertheless all-things-considered permitted to ignore. Plainly, the argument here involves an enormous number of complications and further considerations, none of which I can so much as mention. But the considerations here are deep, and are almost always overlooked in discussions concerning the consequences of incompatibilism.

I thus take it as highly non-obvious that, even *qua* resolute incompatibilist, if I find out that determinism is true, then I would have to give up the reactive attitudes – or indeed, do anything differently at all. Of course, I concede that this discovery would occasion certain *attitudinal* changes in me – it would perhaps prompt the thought that the human predicament is in many ways absurd, or unfair, and so on. And it may prompt an ongoing or intermittent reflective ambivalence about both my own dispositions to reactive anger, and that of those around me. But this is not to say that I would have to *do anything about all of this*. Those who say and write otherwise need to make their case.¹¹

¹¹ Pereboom has written:

It is plausible that to a certain degree moral resentment and indignation are beyond our power to affect, and thus even supposing that a hard incompatibilist is thoroughly committed to morality and rationality, and that she is admirably in control of her emotional life, she might nevertheless be unable to eradicate these attitudes. Thus as hard incompatibilists we might expect people to be morally resentful in certain circumstances, and we would judge it to be in an important sense beyond the agent's control when they are. However, we also have the ability to prevent, temper,

But let us now consider the alternative reading of the relevant conditional, where “we” refers to some relevant *community*.

(3) If we found out determinism is true, we would have to give up the reactive attitudes.

The problems in assessing (3) are legion. Who are we to imagine “finding out” that determinism is true – and how? “We” community of philosophers? “We” educated folk who stay on top of the latest scientific discoveries? “We” politicians who determine the trajectory of our nation’s laws and customs? “We” autocrats? Or everyone – literally everyone – all at once? No matter which interpretation we pick, it seems, problems parallel to those considered above will seem to arise: *can* we do this? And supposing we could, would it be all things considered good if we did, or better than other things we could be doing instead – like addressing the problems of poverty, or climate change? (And anyway, why we do *have to* do this for the sake of conforming to the standards of fairness, when those standards play no explanatory role making the universe what it is like?)

Fischer has asked us to imagine a scenario in which a team of physicists makes a series of discoveries, and we subsequently wake up to a *New York Times* headline, “Determinism is true!”¹² Making certain (generous) allowances, this is indeed something we can imagine. But whereas I can imagine a team of physicists “announcing” that determinism is true – and the press being swept along for the ride – I can also imagine laughing at those physicists, in particular, at their startling combination of hubris and naivete. Fischer seems to write as if physicists could “discover” the truth of determinism in much the same way as physicists recently *did* discover the Higgs boson. But these cases are profoundly different; this is certainly not the place to discuss these matters, but note that it is unclear in principle what empirical discoveries *could be* made by physicists that would decisively establish the truth of determinism. (Indeed, we are told by some, the relevant interpretations of the physical theories are *empirically* equivalent.) The questions here are accordingly much more global and theoretical in nature. Further, though we can allow that we should defer to physicists regarding the nature of the *laws of physics*, the question whether all laws of nature reduce to laws of physics (for instance, the question whether there are any *emergent powers*¹³) is itself *not* a question settled by the physicists. Thus, even if physicists converge on the

and sometimes to dispel moral resentment, and, given a belief in hard incompatibilism, we might attempt such measures for the sake of morality and rationality. (2007: 119, Cf. also 2014: 181)

Yes, we might – but would we *have to*?

¹² Fischer 2007, 2012.

¹³ Cf. Wilson 2021: Ch. 8.

thesis that the laws of physics are deterministic, this leaves the question of *determinism* still open: for perhaps there are processes governed by laws that neither reduce to nor supervene upon those laws. Who are the physicists to tell us otherwise?

The point here is that if we are to imagine suddenly gaining decisive evidence for the truth of determinism, then we are entering into the realms of science fiction.¹⁴ I have no objection to entering such realms, but we should tread cautiously when assessing *what we must do* in certain science fiction scenarios; the contours of these scenarios are often unclear. For instance: perhaps aliens (or AI overlords?) emerge, and perhaps, through some convincing demonstration, they show us a series of parallel universes that are physical duplicates of our universe up to the moment of their intervention – there is my counterpart, doing *exactly* what I did, down to the finest details! – and perhaps this forces on us the realization that any universe with our starting conditions and laws is a duplicate of our universe, which is to say: everything we have done has been an inevitable consequence of the state of the universe “at the beginning”. Well, what must we do, after this revelation? Frankly, I struggle to know how to begin to answer this question. But is it outrageous to suggest that we might collectively decide that though this new revelation did overturn our impression of ourselves as responsible for what we do and become, well, our priorities remain in place, life must go on – and so *carry on as you were*? I don’t know, and this is partly because it is unclear what scenario I’m even imagining, and plainly lack the space to fill in more of the details one way or another. But then my point: philosophers who write that, on incompatibilism, if *we* came to discover that determinism is true, then we would have to give up the reactive attitudes, have substantially more work to do to show us that this is true. *How* are we making this discovery, *who* is making it, and in what circumstances? Of course, on incompatibilism, there may be *some* way of answering these questions such that, given those answers, the people in this scenario now (all things considered) must relinquish what reactive attitudes they may have. But it could be that the imagined people in the imagined circumstances bear very little similarity to us as we currently are, and so what these people must do given this discovery tells us nothing about what *we* must do given this discovery.

Let us sum up. It is one thing to say that there is no world in which anyone both deserves blame and is determined, and quite another to say that all agents at determined worlds have an all-things-considered moral requirement to abandon the practices of blame. It could be true that no agent at any determined world deserves blame; it could also be that some agents at some determined worlds aren’t required to abandon their practices of blame. (And it could be

¹⁴ Cf. the science-fiction deterministic thought experiment described in Todd 2019.

that one of those deterministic worlds is our own.) At the very least, this question is an open question, and one that can't be settled by analytic facts alone.

Conclusion

My overarching point in this essay is that we need to discuss “incompatibilism” more carefully. First, we need to be very careful before writing, “According to the incompatibilist, if determinism is true, no one is responsible.” This goes for many incompatibilists, but not all – not those libertarians we called resolute responsibility theorists (or “flip-floppers”). Second, we need to be careful not to directly present incompatibilism in terms of an alleged conditional *requirement* – the requirement to give up some relevant attitude or practice, conditional on determinism. On this front, let us revisit Watson’s claim from above – that incompatibilists insist that the truth of determinism would require us to take the objective attitude universally. As I have argued, this characterization won’t do, at least not without further argument. At most, we can say the following: according to the incompatibilist who is *resolute*, the truth of determinism would give us *pro tanto* reason to take the objective attitude universally. That blame is or would be undeserved or unfair is certainly a *pro tanto* reason against blame. But whether and in what circumstances this *pro tanto* reason amounts to a *decisive* reason is a difficult further question – and one that deserves much more attention than it has thus far been given.

Let me finish this essay with one remark about how these final points bear on the longstanding, torturous dispute between the compatibilists and the incompatibilists about responsibility and determinism. My sense – which I certainly cannot here justify – is that many compatibilists are motivated by the key thought that *nothing could show that we have to give all of this up*. That is, nothing could show that we have to give up what certainly seem to be central features of our “form of life” – our general readiness to respond to others with reactive attitudes. But if I am right, this is a thought with which the incompatibilist can in principle agree. If what is most plausible about compatibilism is that *nothing could require us to give the attitudes up*, and if what is most plausible about incompatibilism is that *no one who is determined is also fairly blamed* – that is, if incompatibilism simply and directly seems true¹⁵ – then perhaps these are two thoughts that can ultimately be reconciled. But this is a discussion we must leave for another occasion.¹⁶

¹⁵ For instance, on the basis of the manipulation arguments; cf. Todd 2013 and 2017, and Pereboom 2014: Ch 4.

¹⁶ For helpful comments on previous drafts of this essay, I wish to thank Michael McKenna, Mark Balaguer, Daniel Telech, Simon-Pierre Chevarie-Cossette, Brian Rabern, Andrew Bailey, Max Kiener, and Justin Capes.

References

- Bailey, Andrew M. & Seymour, Amy. 2021. "In defense of flip-flopping," *Synthese* 199: 13907-13924.
- Fischer, John Martin and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fischer, John Martin. 2007. "Compatibilism," In *Four Views on Free Will*, ed. M. Vargas. Malden, MA: Blackwell Press. Pp. 44-84.
- Fischer, John Martin. 2012. "Semicompatibilism and its Rivals," *Journal of Ethics* 16: 117-43.
- Fischer, John Martin. 2016. "Libertarianism and the Problem of Flip-flopping," in Timpe and Speak, eds., *Free Will and Theism*. Oxford: Oxford University Press.
- Hawthorne, John. 2002. "Advice for physicalists," *Philosophical Studies* 109: 17-52.
- Holguín, Ben. 2020. "Knowledge in the Face of Conspiracy Conditionals," *Linguistics and Philosophy* 44: 737 – 771.
- Khoo, Justin. 2022. *The Meaning of "If"*. Oxford: Oxford University Press.
- Mavrodes, George. 1986. "Religion and the Queerness of Morality," In William Wainwright & Robert Audi (eds.), *Rationality, Religious Belief, and Moral Commitment: New Essays in the Philosophy of Religion*. Cornell University Press. pp. 213--26.
- Mellor, D. H. 1993. "How to believe a conditional," *The Journal of Philosophy* 90: 233–248.
- Pereboom, Derk. 2007. "Hard Incompatibilism". In *Four Views on Free Will*, ed. M. Vargas. Malden, MA: Blackwell Press.
- Pereboom, Derk. 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Ramsey, Frank. 1931. *The Foundations of Mathematics and other Logical Essays*. London: Kegan Paul, Trench, Trubner & Co.
- Stalnaker, Robert. 1975. "Indicative Conditionals," *Philosophia* 5: 269–286.
- Strawson, P.F. 1962. "Freedom and Resentment," *Proceedings of the British Academy* 48: 1–25.
- Steward, Helen. 2021. "What Is Determinism? Why We Should Ditch the Entailment Definition," In Marco Hausmann & Jörg Noller (eds.), *Free Will: Historical and Analytic Perspectives*, Springer Verlag. pp. 17-43. 2021.
- Todd, Patrick. 2013. "Defending (a modified version of) the Zygote Argument," *Philosophical Studies* 164: 189-203.
- Todd, Patrick. 2017. "Manipulation Arguments and the Freedom to do Otherwise," *Philosophy and Phenomenological Research* 95 (2): 395-407

- Todd, Patrick. 2019. “The Replication Argument for Incompatibilism,” *Erkenntnis*. 84 (6): 1341-1359.
- Todd, Patrick and Brian Rabern. Forthcoming. “Resisting the Epistemic Argument for Compatibilism,” *Philosophical Studies*.
- Van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- Watson, Gary. (1987). “Responsibility and the limits of evil: Variations on a Strawsonian theme,” In F. Schoeman (Ed.), *Responsibility, character and the emotions: New essays in moral psychology* (pp. 256–286). Cambridge University Press. Reprinted in Watson, *Agency and Answerability: Selected Essays*. Oxford: Oxford University Press. 2004.
- Wilson, Jessica. 2021. *Metaphysical Emergence*. Oxford: Oxford University Press.

Further Reading

When we ask whether “responsibility” is or is not consistent with determinism, there are different notions of responsibility we may have in mind; the focus on “reactive attitudes” is most intimately associated with Strawson 1962. Various philosophers have written or suggested that on incompatibilism, if determinism is true, we must abandon the reactive attitudes; see Watson 1987 and Fischer 2016. Bailey and Seymour 2021 is a recent reply to Fischer’s criticisms of “flip-flopping”; see also Todd and Rabern forthcoming.

Biographical Note

Patrick Todd is a Chancellor’s Fellow/Lecturer at the University of Edinburgh, where he has been based since 2013. For the 2022/23 and 2023/24 academic years, he is based in Lund, Sweden, as a Senior Research Fellow at the Lund/Gothenburg Responsibility Project. He writes on various aspects of the problems of free will and moral responsibility, including issues regarding incompatibilism and compatibilism, and the standing to blame. He has recently published a book in metaphysics with Oxford University Press, *The Open Future: Why Future Contingents are All False*.