

Compatibilism, Conditionals, and Control

A Response to my Critics

Authors of scholarly papers usually express their gratitude for the comments of their colleagues in a footnote. It is a privilege that I can express mine in the main text, and right at the beginning. In what follows I do my best to respond to many important critical remarks about a work the main purpose of which was to convince readers that the traditional conditional account of free will is not yet defunct and that it can provide the best framework for discussing important issues about agency and responsibility. But whether or not I have managed to convince my readers is not the issue here. I am grateful for the opportunity, in trying to respond to my critics, to restate, and hopefully sharpen, some of my claims on the perennial problem of freedom of the will.

1. COMPATIBILISM AND INCOMPATIBILISM

My main theoretical interest in free will is not the compatibility of freedom and determinism but the nature and limits of the kind of control that is required for free agency, action, and responsibility. Nonetheless, I am a compatibilist regarding free will and physical determinism. I am a compatibilist in that sense, because I fail to see how the truth or falsity of determinism at the level of ‘fundamental physics’ can be relevant to questions about our agency. Here are some reasons why I find incompatibilism problematic, even before any detailed argument is made for or against it.

Suppose Fred starts raising his hand at t in order to open his fridge and thereby to get some fresh milk for breakfast. If incompatibilism of the sort I cannot accept is true, Fred was *strictly unable* to raise his hand a nanosecond earlier. At that time, say at $t-\delta t$, he was exactly in the same physiological and mental conditions as he was at t . He had exactly the same reasons to raise his hand then as he did at t . Nothing has changed in his environment between t and $t-\delta t$ that is relevant for the success of his action. His hands were not tied; no demons conspired to strike down on him if he tried to raise his hand at $t-\delta t$, etc. Incompatibilists say

they understand the sense in which he was nevertheless unable to start raising his hand at $t-\delta t$. I do not quite see why.

Further, incompatibilists say that the reason that Fred was so disabled is cosmological: his inability is explained by the same factors as is Jupiter's inability to have a different position and momentum a hundred years from now. Fred's inability is the consequence of how the universe was a few seconds after the Big Bang, plus at any other moment of cosmic history since then. His inability is cosmologically necessitated by all past instances of the universe, not to mention the future ones which, if determinism is true, also necessitate the present. But even if I could make some sense of the kind of ability that can be lost because cosmic determinism is true, I cannot see how that ability could be relevant to the freedom of our agency.

Finally, if the falsity of physical determinism is a condition of responsibility, then there is a good chance that we shall never be able to find out whether or not we are free agents.¹ For there is a good chance that all the facts that will ever be known for us underdetermine the interpretation of the fundamental physical laws. And it is only those laws as interpreted deterministically which can render poor Fred unable to raise his hand at $t-dt$. The truth of cosmological fatalism always remains a secret for us.

In *Freedom of the Will: A Conditional Analysis* I argue for compatibilism by trying to reject what I take to be the most promising argument for incompatibilism: two versions of van Inwagen's famous consequence argument. Some of my critics challenge my assumption that there are no better versions of the argument for incompatibilism; others claim that my objection to at least one version of the argument is mistaken.

I reject the consequence argument on the grounds that every version of it relies on a principle that we might call the principle of transfer of powerlessness or the principle of inability closure. According to this principle, if S is powerless with respect to P , and P entails Q , then S is also powerless with respect to Q . It still seems to me that it is impossible to construct an argument for the incompatibility of free will and physical determinism without using some premise that presupposes some such closure, and for this reason no such argument can be regarded to be a conclusive refutation of compatibilism.

One central issue with the consequence argument is how to capture the relevant ability that we allegedly lack in the circumstances of cosmological determinism. Gábor Bács, following Bernard Berofsky, suggests that we should capture the relevant ability in terms of unalterability. His claim is that if determinism is true, then no one is able to alter the future. This is supposed to be a

¹ Here I agree with Galen Strawson (1994). I do not, of course, agree with his more general claim about the impossibility of moral responsibility.

consequence of the fact that no one is able to alter the past and the laws, and, if determinism is true, then the past and the laws entail the future (p. 27).²

However, this argument is no stronger than is our understanding of ‘unalterability’. Alteration, to my ears, means change. Of course, we cannot change the past. But neither can we change the future, simply because *the future* is an indexical expression that refers to whatever is actually going to happen. This has nothing to do with either our abilities or determinism. Perhaps there is another sense of the phrase ‘you cannot change it’. In that sense you cannot change an event if there is no way for you to influence its occurrence. But, so understood, we are back to my original worry about the consequence argument: only a fatalist would hold that we cannot influence the future. What follows from this version of the consequence argument is that incompatibilism entails the truth of a sort of fatalism which most incompatibilists would reject as false. It is for this reason that I cannot see how this proposal can rescue the consequence argument, and not because I reject psychological determinism (as does Berofsky, as far as I know).

Bács also suggests that van Inwagen’s own version of the consequence argument is sound, provided we modify van Inwagen’s own understanding of abilities. According to van Inwagen, the relevant ability is an agent’s ability to *exercise* some ability. Bács agrees with me that we cannot characterize in such a way the sort of ability that is relevant for freedom of the will. But Bács disagrees with my point that the ascription of the ability to exercise abilities might lead to a logical contradiction, though he also says that this is only a minor issue (p. 30). I agree that this is a minor issue, but his disagreement provides me with the opportunity to recast my argument, which, as I see it, still stands.

Bács says that the expression ‘ability to exercise an unexercised ability’ involves a scope ambiguity. On a wide scope reading it involves a contradiction, but the narrow scope reading does not involve such a contradiction. I agree with this. However, my point was meant precisely to be that the narrow scope reading, which refers to an ability which is such that it was not actually exercised but could have been exercised, fails to specify the relevant sense of powers or abilities. For the narrow scope reading applies to *any* unexercised ability, never mind whether it is an ability of an agent or an inanimate object, generic or specific, etc. So we need the wide scope reading to specify the allegedly ‘special sense’ of ability, and *that* reading involves a contradiction.

Related to this, Bács also argues that, according to my account, the abilities which are relevant for an agent’s responsibility are to be identified with maximally specific and often extrinsic abilities. But, he says, this account of the relevant abilities obliterates the distinction between abilities and opportunities. The relevant abilities are ‘maximally specific extrinsic determinations of pow-

² All numbers henceforth refer to the pages of this volume, unless otherwise indicated.

ers', like a Ferrari's power 'to go faster than 130km/h with S in its driver seat, S being a cautious driver and the speed limit being 130 km/h, and so on'; these maximally specific powers "can be lost in a deterministic universe according to the first consequence argument" (p. 33).

I'm not sure whether I have got Bács's point correctly here. I agree that powers, and the corresponding claims about what things or persons can or cannot do, can be more or less specific, and hence generic powers can be retained even in circumstances in which the specific abilities are lost. I'm less certain about the concept of 'extrinsic determination'. If this means that specific powers are *more determinate* than intrinsic ones, then it is certainly true in many cases. But that has nothing to do with the issue of *determinism*. And I fail to see how the consequence argument is connected to the fact that the possession of many abilities is sensitive to some state of the world *at the time when* we ascribe them to the object.

Perhaps the idea is that if it follows from the past and the laws that a power is not exercised at a given time, then objects cannot possess the power itself. But if this is the correct interpretation of Bács's claim, then he merely reformulates the incompatibilist conviction; he does not seem to argue for it. For, contrary to Bács's assumption, if the consequence argument is sound, it applies to *any* unexercised power, no matter how generic or specific it is. Bács seems to agree with me that the argument cannot be sound when it is applied to the question of generic powers. But he does not show how the argument becomes sound when it is applied to maximally specific powers.

Perhaps the idea is that if extrinsic circumstances can be relevant for the possession of a power, then abilities can be sensitive to any state of the cosmos: past, present (future?). Now the past can obviously be relevant for the possession of certain powers to the extent that agents would lack or possess certain powers which they have or fail to have now, if the past had been different. Learned abilities or 'second natures' provide the most obvious examples. However, if the consequence argument is sound, then in a deterministic universe the only ability I can have at this moment is the one which I exercise. Or not even that, for if the possession of my present ability depends on the possession of my ability to influence the cosmic past then *I* simply cannot exercise any ability at all, only the cosmos can.

Bács might respond that the remote past can be relevant for the possession of abilities if we assume that there is an asymmetry between the ability to render a proposition false and the ability to render it true. I argue against closure by observing that in the case of actually exercised abilities we do not require that agents possess the ability to render propositions about the past and laws true. Bács answers that this is irrelevant; for we can have our present and exercised abilities if the relevant propositions are actually true; we need not be able to make them true. However, this alleged asymmetry between the abilities of rendering propositions true and rendering them false seems to be an illusion.

Suppose Fred actually sits at t . Thereby he renders the proposition ‘Fred stands at t ’ false. Thus, obviously, he can render some propositions false. Suppose, further, that there is a set of propositions about the past and the laws, PL^* , which, given determinism, entails that Fred stands at t . By assumption, Fred cannot render PL^* false, but since he actually sits at t , he does, and hence can, render the proposition ‘Fred stands at t ’ false. So closure fails. It is no response to this that Fred *need not* be able to render PL^* false, since it is actually false. For the contentious point is whether, if determinism is true, our *inability* or *powerlessness* with respect to the past and the laws is, logically speaking, compatible with our *ability* to do something that we actually fail to do. As far as logic is concerned, whether PL^* is actually true or false is irrelevant.

Howard Robinson seems to grant that van Inwagen’s arguments fail, but he suggests an alternative, causal argument for incompatibilism. The gist of this argument is that I am not able to do something the opposite of which is strictly causally necessitated. But, given determinism, the past and the laws strictly causally necessitate everything I do. So, if determinism is true, I am not free to do anything other than what I actually do (p. 74). Relatedly, Robinson also argues that there is no good reason to avoid the use of causal language in the argument, as both van Inwagen and I do. These are large issues, and I cannot do more here than scratch the surface of the problem by briefly stating what I think about causation and the causal arguments for incompatibilism.

First of all, unlike van Inwagen, I do not think that causation is a ‘morass’ and that the concept of cause is unrelated to the issue of free agency. What I do think, however, perhaps with a tiny minority, is that causation presupposes the experience of free agency; moreover, it presupposes the experience of free agency precisely in the sense captured by the conditional analysis. Thus we cannot understand the conditions of free agency in causal terms, for the essential condition of the use of causal terms is our experience of free agency.

Second, as Robinson mentions, I deny that causal language has the appropriate modal content for discussing the problem of cosmic determinism, since causal relations are metaphysically contingent and causation can be non-deterministic. Robinson says that the latter worry is irrelevant “because we are discussing determinism” and that “there is no assumption about all causation being deterministic” (p. 75). But my argument was that causation *cannot explain* the relevant sense of necessity because causation can be nondeterministic. One cannot respond to this that deterministic causation explains the modal force of determinism. For that would boil down to the claim that deterministic causation explains determinism. Even if this were true, it would not be very informative. But I believe that this is actually false.

The concept of determinism is independent of the concept of causation, in the sense that determinism can be true without there being any causes. The values of certain parameters of a system can determine the value of some of its

other parameters without causing it. Given the temperature and the pressure of the air exercised on the walls of my room, its volume can be determined. But its volume is not caused by the temperature and the pressure. It was caused by the work of those who built it, if by anything. Given that deterministic physical laws are time symmetric, the present state of a deterministic universe is determined by its future exactly in the same sense as it is determined by its past. But—some possible cases of local backward causation notwithstanding—the future does not cause the past; it certainly does not cause it globally. Provided that we have an acceptable notion of determinism,³ we might characterize some causal processes as deterministic. But causation comes only later, if at all.

And this, I believe, is crucial for the viability of Robinson's version of the consequence argument. Van Inwagen aims to argue from a *logical* truth concerning global determinism to the *metaphysical* impossibility of possessing certain sorts of unexercised abilities. But Robinson wants to argue from causal or nomological necessity to the non-existence of freedom-relevant abilities. However, no traditional compatibilist would grant that if *f*-ing is not nomologically compossible with the past, then *f*-ing is not metaphysically possible. In other words, no traditional compatibilist—including myself—would agree that if *f*-ing is not nomologically compossible with the actual *past* then the agent is deprived of his *present* ability—the ability to do-at-*t* the action in question, as Robinson puts it. The question is precisely whether or not *nomological compossibility with a certain past* can ground the *metaphysical necessity* (i.e. non-contingency) of *every actual action*. According to the incompatibilists, it does. According to the compatibilists, it does not. We cannot just assume that actions are nomologically or causally *necessitated* in the sense that agents lack the ability to do otherwise, since this is precisely the issue at hand.

Finally, I would like to say something about what we might call the *level-based argument* against compatibilism. Such arguments are not versions of the consequence argument, since they do not aim to show that we cannot choose or act otherwise *only* because the universe is deterministic. Rather they claim that if mental events or bodily actions supervene on what happens at the micro-physical level, then micro-physical determinism is incompatible with choice and/or the ability to perform certain actions. Robinson raises the worry that “if you are not a psychological determinist but a physical determinist, where what happens is fixed at a more basic level, then it is not clear that the determining process works *through* choice, rather than rendering it epiphenomenal” (p. 72).

My first point is that I fail to see how the problem about physical closure is related to the issue of freedom and determinism. Most physicalists who accept closure would not, I suppose, hold that physics at the fundamental level is de-

³ Do we? I leave to philosophers of science to decide. For an interesting exposition of the problem see Balázs Gyenis (2013).

terministic. For this reason, Robinson's worry can easily be turned upside down. One may complain that "if you are not a physical determinist but a psychological determinist, where what happens is fixed at a more basic level, then it is not clear that the non-deterministic physical process works through any deterministic psychological process, rather than rendering it epiphenomenal".

I cannot venture a response here to the issues of epiphenomenalism and closure. But I do think that epiphenomenalism about the mental, including epiphenomenalism about choice, is a question distinct from the compatibility of determinism and free will. Like almost everyone else, I assume that epiphenomenalism is false; for if it is not, then the whole issue about agency and compatibility fails to make sense.

However, Robinson's point may not concern the possibility of causally efficacious mental processes, which *qua* mental processes may or may not be deterministic, but rather the action itself that supervenes on the movement of the microphysical particles. Thus, "if the world is closed under physics, then exactly where my body is, what motions it is going through and what noises, if any, it is emitting, will be entirely determined by micro-events that are entirely sub-psychological" (p. 79). László Bernáth claims, in similar spirit, that "if the movement of a set of particles depends on something to some degree, the action connected to this movement depends on the same thing to the same degree as the movement of the set of particles itself" (p. 115).

The first thing to observe here is that whatever we mean in this context by 'determination' is entirely different from *determinism*. The issue now is to which extent *the explanation of the behavior of a composite object* depends on the explanation of the behavior of its parts. Here I can only express my disagreement with Robinson and Bernáth. Cosmic determinism can reveal something only about the physical relation among the whole physical states of the universe; it cannot reveal anything about the nature of local processes, including our intentional actions.

This is not only a matter of 'stances'. Which inspector would be satisfied with the following explanation of an air crash? "Look, I've just read a book that proves that we live in a deterministic universe. This aircraft crashed because its behavior supervenes on the behavior of its constituent particles the position and momentum of which are necessitated by the past states of the cosmos. Given the actual state of the universe, those particles must be now exactly where they are. In fact, although most aircrafts of this type do not crash in similar circumstances, it would have been a miracle, if this one had not." If this is a stance, it is a really silly one. What matters here is the working of the engines, wings, automatic piloting, and so on. No one is interested in the history of the universe, and how it might 'determine' the position and momentum of a set of particles at a certain moment. Moreover, what matters is that the aircraft, by all sensible

human standards, *could have been* put together so that the air crash would not have happened, never mind the causal history of the particles that composed it.

Suppose that the set of particles that compose my body now exists also at times when they do not compose my body. If determinism is true, then their position and momentum is a consequence of the past states of the universe, independently of whether or not they compose my body. But certainly, when they do compose my body, then their temporally local movements must depend on some states of my body. Who would claim that, in the actual local circumstances, the particles that compose *my* fingers would move now (while I'm typing these letters) in the way they do without *my* brain being in the state in which it is?

Well, who indeed? Perhaps the incompatibilists. For according to them, the movement of my fingers depends 'ultimately' on the billions of cosmic states, not on the local states of my brain. But accepting this is not a denial of the ability to do otherwise in a deterministic universe. It amounts to a denial of *my existence*. For my existence as a biological organism, not to mention as a person, is tied to the possession of abilities, like the ability that, normally, *I* can move *my* fingers or *my* whole body, but I cannot move *yours* even if you somehow manage to absorb the set of particles that left my body. If one wants to deny the possibility of free agency on the grounds that we are composed of micro-physical particles of some kind, the movement of which is a consequence of the earlier states of the universe, one ends up inevitably denying the possibility of personal agency itself in a deterministic universe.

2. POWERS AND CONDITIONALS

Freedom of the will as I understand it is the *power* or *ability* to do otherwise, because it is that ability with reference to which we can capture both the alternative possibility condition and the control condition of responsibility. The ability or power to behave otherwise is the sort of alternative possibility that is relevant for agents' responsibility. It is the availability of alternatives in this sense that is necessary for the kind of control which agents must possess in order to be responsible for their actions and omissions. This view is hardly new. In fact, it has a quite respectable history. Its origins can be traced back at least to Augustine's and Boethius' understanding of human freedom.⁴

The reason I want to offer a conditional analysis of free will is that counterfactual conditionals provide the best means of *identifying* the abilities the possession of which is the metaphysical condition of responsibility for actions. The required analysis does not entail reduction, not to mention elimination: who would want to eliminate free will by trying to explain what it is? And why would

⁴ See Tomas Ekenberg (2009).

anyone need to ‘reduce’ it to something else? A good analysis means better understanding: better understanding of something that exists before any effort to analyze it, and of something that can, hopefully, also survive our analysis.

Since free will is an ability, it can also be understood as an unexercised power. Thus, in some respect, it is similar to those properties the ascription of which entails potentialities. The most often discussed properties of that sort are properties expressed by the so-called ‘disposition terms’. At least since John Mackie’s (1973) influential early discussion about dispositions,⁵ ‘dispositions’ and ‘powers’ have most often been used as quasi-synonyms. Dispositions are enlightened philosophers’ powers, so to say. They are not ‘occult qualities’ to the extent that they can refer opaquely to objects’ causally relevant properties. The aim of an analysis of dispositions is to unfold the connection between the ascription of dispositions and the truth of certain causal counterfactuals.

However, I do not think that powers *are* dispositions. Dispositions are not independent of powers, since nothing can have the disposition to *M* unless it also has a power to *M*. But while dispositions imply behavioral tendencies—in a sense, of course, which is compatible with the tendency not being manifested by objects’ actual behavior—powers do not. This difference is crucial for my account of free will, since that account aims to understand free will in terms of powers and not in terms of being disposed to behave in certain ways in certain circumstances.

Free will, as I understand it, is at least one condition of an agent’s responsibility for their actions. Consider our Fred character again, this time visiting his aging mother in hospital. Fred is responsible for what he does because, among other things, he can, in the sense that he is able to, avoid paying the visit. But, being the nice fellow he is, he need not at all be *disposed* to avoid the visit. In general, in order to be responsible for a kind of behavior, one need not be disposed to do otherwise; in fact, normally, people are not so disposed in cases when they are responsible. What they need to have is the power to do otherwise.

Thus, the distinction between the conditions in which one can have the power to do something and in which one is disposed to do it plays a significant role in my conditional account of the ability to do otherwise, even if the idea of that account originates in the work of a philosopher who does not distinguish them. As far as the relation between powers and counterfactual conditionals is concerned, my views derive from Hugh Mellor’s work on dispositions. Mellor claims that the ascription of dispositions entails conditionals, but no conditional itself entails a disposition. This is so because it is the possession of some dispositional property that grounds the truth of the relevant conditionals. It follows that a conditional analysis of disposition *terms* need not be ‘reductive’. Its purpose

⁵ In a later paper Mackie explicitly claims that he sees no reason to distinguish powers from dispositions. See Mackie (1977: 362).

is not to show that such terms do not refer to genuine properties, or that they do so only to the extent that they express the conditions of some kind of causal interaction between events.

Of course, any property's instantiation might require the presence of something ontologically more fundamental. But admitting this does not require a reduction of the dispositional to the non-dispositional. In fact, I must admit that I doubt that any *property* has ever been reduced to something else in this way. The possession of one dispositional property might be explained by the possession of some other one. But this only means that the presence of a disposition or a cluster of dispositions (like fragility or temperature) in an object is explained by the presence of some other disposition (like some kind of molecular bonding or mean kinetic energy).⁶ This is so even if the explanation is a form of entailment; after all, it is just very hard to see how something which is not dispositional can entail anything that is.

The 'one-way entailment' from dispositions to conditionals does not mean that conditional analysis is a nonstarter. It does however mean that we need to assume that the property to be analyzed can actually be possessed by the object to which we ascribe it. If the power or disposition to be analyzed are properties of objects in the same sense in which their 'qualities' are properties, then we can introduce a further condition into our analysis which requires that the object does not change with respect to the possession of the relevant power during the period of its would-be manifestation.⁷

Adding this condition can fence off the conditional analysis of free will from some traditional counterexamples. These counterexamples are based on the possibility that agents can lose or acquire an ability to perform an action as a result of choosing to perform that action.⁸ It is this type of counterexample that has been historically most influential against the conditional analysis of free will as the ability to act otherwise. Independently of the debates about free will, similar counterexamples have been raised against the traditional conditional analysis of dispositions as well. Such examples are called now 'finkish dispositions', following an important exchange between Charles Martin (1994) and David Lewis (1997) on the conditional analysis of dispositions.

In her insightful reconstruction of the most recent debates about conditional analyses of dispositions, Zsófia Zvolenszky complains that my rejection of the simple conditional analysis is not radical enough. I say that "how objects would behave in specific circumstances is not sufficient to grant them a power or ability". But what I *should* say is that "how objects would behave in specific circumstances is *neither necessary nor sufficient* to grant them a power or ability" (p. 59).

⁶ See David H. Mellor (1974).

⁷ See David H. Mellor (2000).

⁸ See Keith Lehrer (1968/1982).

Certainly, Zvolenszky is right in that *the truth of the conditional* as formulated in the simple conditional analysis is neither necessary nor sufficient for the ascription of a disposition. Cases when an object *loses* a disposition when the circumstances of manifestation occur show that the truth of the conditional is not *necessary* for the ascription; cases in which the object *acquires* a disposition when the circumstances of manifestation occur show that the truth of the simple conditional is not *sufficient* for the ascription.

Nonetheless, I still cannot see the possibility of identifying a power without any reference to how objects possessing it would behave in certain circumstances. It is one thing to say that the *truth* of the conditional as formulated in simple conditional analysis is not sufficient for ascribing a power. It seems another thing to claim that we can identify a power or an ability *without any* reference to how the objects possessing them would behave in certain circumstances.

If I understand Zvolenszky correctly, she argues that the possibility of ‘masked’ and ‘mimicked’ dispositions renders impossible to analyze powers with reference to how objects would behave in certain circumstances. In my work I discuss only such cases in which objects might *change* some of their powers in the circumstances in which they are about to become manifest, i.e. I discuss only the possibility of ‘finkish’ powers. However, it has been argued that the conditional analysis of dispositions can fail even in cases when objects do not change their dispositions; their dispositions are only ‘masked’ or ‘mimicked’. Good wrapping might save a glass from splintering when it is dropped, but the glass remains fragile nonetheless. The glass’s fragility is masked. And even non-fragile things, like a landmine, can splinter when they are dropped. The landmine’s behavior mimics fragility.

Such examples are often cited against the most influential attempt to modify, in a reductivist spirit, the conditional analysis: David Lewis’s (1997) revised conditional analysis. And since some new versions of the conditional analysis of free will rely exactly on that revised account, the examples also seem relevant for the analysis of agents’ abilities to act otherwise (Vihvelin 2004). In my view, however, the conditional analysis as revised by Lewis cannot be deployed to rescue the conditional analysis of free will. And, more importantly, I do not think that the possibility of masks and mimics is a problem for the conditional analysis of powers, even if they might be a problem for the conditional analysis of dispositions.

According to David Lewis’s view, objects can possess dispositions only by virtue of having some intrinsic properties which are supposed to serve as the disposition’s ‘intrinsic causal basis’. The retention of a basis is a necessary condition for the manifestation of the disposition. The reference to the ‘intrinsic basis’, instead of the disposition itself, has the promise to render the analysis ‘non-circular’. My first remark about this is that although the reference to the assumedly non-dispositional base does reflect an ontological commitment—‘no free standing dispositions’: ‘deep down’ everything must be ‘intrinsic and quali-

tative or categorical’—it can hardly make the analysis less circular unless we can somehow identify the basis without referring to the disposition to be analyzed.

However, as Lewis himself recognizes, we cannot. For the only way for us to single out the relevant intrinsic quality is to say that the property, *whatever it is*, that fulfills the role specified by the conditional must be the basis of the analyzed disposition.⁹ The relevant intrinsic properties might be different in each object that has the disposition, or they might be replaced by another in the same objects at every moment. For this reason, the retention of the relevant intrinsic base cannot be a necessary condition of the possession of a disposition. But even if it were, Lewis’s analysis would be as ‘circular’ as the simpler account that I prefer. Instead of adding to the conditions that the disposition itself is retained, it adds to them that whatever is the basis of the disposition is retained; but that basis can only be identified *qua* basis of the disposition that is thus also retained. Thus the reason for enriching our ontology with this extra property cannot be that its introduction renders *the analysis* of dispositions ‘non-circular’.

But the alleged circularity of these analyses is a pseudo-problem. Both attempts to revise the simple analysis are informative and non-empty, and thus ‘circularity objections’ to them fail to have good grounds. The reason why Lewis’s analysis of dispositions cannot help in the analysis of the abilities that are relevant for free will is that such abilities are, more often than not, *extrinsic*. To use a painfully boring but nonetheless helpful example from Locke, I can change with respect to my room-leaving ability simply by being locked into a room. Never mind whether that property is ‘irreducibly dispositional’ or ‘qualitative’: intuitively, when I am locked in, nothing has intrinsically changed *in* me, but I have lost a power and hence I am not responsible for not leaving the room.

It is here that the problem of masks becomes important. Many philosophers would say that in such circumstances I do have the ‘intrinsic disposition to leave rooms’, but this disposition is ‘masked’: in the circumstances I cannot manifest it. I must admit I have certain difficulties with understanding how the power to leave rooms can be an intrinsic property of anyone at all.¹⁰ My power to stand up and walk is indeed intrinsic, but those powers of mine are neither lost nor masked when I’m locked in a room. My room-leaving ability is extrinsic in the first place. No one should perform any kind of surgery on me in order to deprive me of it; it is enough to change my environment in certain ways. But this means that my power is not only ‘masked’ in such circumstances: I simply fail to have it.

⁹ See David Lewis’s important posthumous paper ‘Ramseyan Humility’ (2009).

¹⁰ Of course, I know how to leave the room, and that know-how is intrinsic to me. But abilities are distinct from know-how and I can lose them even when I know how to perform an action. See Kieran Setiya (2000).

Consider the standard example for a masked disposition, the fragile glass in a safe package. Is the glass disposed to break when it is dropped? Observe the ambiguity in this question: is that glass *qua* glass still disposed to break? Of course it is; that's why it is so carefully wrapped up after all. Is *that particular glass* disposed to break when it is dropped? Of course it is *not*; that's why it is so carefully wrapped up, after all! According to the first reading, the disposition is there, but it cannot be manifested *even* in the standard circumstances of its manifestation. According to the second reading, the object simply lacks the disposition, *because* it would not manifest it even in the standard conditions of its manifestations. Since the object can have both properties, the ascription of those properties cannot contradict each other. And they do not so contradict each other, because the object can have a *generic intrinsic disposition* that is 'masked' in the circumstances, and at the same time it can lack a *more specific extrinsic disposition* the possession of which requires the 'collaboration' of circumstances.

It is here that the distinction between dispositions and powers becomes crucial. For it might sound strange indeed to apply dispositional terms to properties whose instantiation is *very* sensitive to the changes of external circumstances. Dispositional terms express behavioral tendencies of objects, or often of *kinds* of objects—like the fragility of things made of non-hardened glass—and such a role just seems to vanish if we make them too specific. On the one hand, of course, fragile objects remain fragile when wrapped up in a safe package. On the other hand, *they cannot break* in those conditions even when dropped. Since they cannot break, they lack the specific power to break.

But when it comes to the problem of human freedom as a condition of responsibility, we are not interested in how someone is generally disposed to behave; at least, our interest in this is only secondary. Rather, we are interested in the question of whether *a particular person in the particular circumstances* had the power or ability to do otherwise, i.e. we are interested in specific powers the possession of which often depends on extrinsic circumstances.

Consider again the person locked in the room which she cannot escape. Suppose you agree that we want to identify the agent's abilities in this situation in order to decide whether or not she was responsible for not leaving the room. Then, if you say that her ability is only masked, i.e. she *does* have the ability to leave the room, then you must also hold her responsible for not leaving it. Alternatively, you must deny that the alternative possibilities that are relevant for agents' responsibility should be identified in terms of their abilities. I'd rather say that the agent lacks the specific ability to leave the room, and this is why she is not responsible for staying in. But then abilities relevant for responsibility cannot be masked; they can only be lost, even if they are lost because of some change extrinsic to the agent.

The point I am driving at is that the powers that are relevant for our responsibility, unlike dispositions, *cannot be masked*. Some unfavorable circumstances can

simply deprive agents of their responsibility-relevant powers. Which circumstances count as ‘unfavorable’ is a moot question, of course. Is a serious threat enough? Is the presence of a Frankfurt-style counterfactual intervener in the background enough? I’m not sure about the first, but I deny the second. A non-realized potentiality of manipulation cannot deprive an agent of the ability to do otherwise. For it is precisely the difference between active interference and the inactive presence of the manipulator which explains our intuition that the non-manipulated agent is responsible. This means that in Frankfurt-style cases there is a change with respect to agents’ abilities in the actual and in the counterfactual situation. It is exactly that change which explains why we have different intuitions about agents’ responsibility in the two situations. Thus such cases seem to me more akin to the cases of finks than they are to the cases of masks.

3. REASONS AND MOTIVES

László Bernáth notes that, according to my account, akratic actions need not be explained with reference to an agent’s reasons. An action can be psychologically explained by the agent’s character and motive without assuming that the agent must have had *some* reason on which she acted. But Bernáth claims that even if it is indeed so, this is irrelevant as far as the plausibility of restrictivist libertarianism and the idea of ‘self-forming actions’ are concerned (pp. 107–108). But it seems relevant to me.

If reasons are normative in the sense that agents’ reasons subjectively justify their actions, and agents’ own mental states justify what they do only in exceptional cases, if at all, then agents’ reasons for actions cannot be their mental states. However, Kane *must* assume that agents’ reasons are their mental states, since that assumption plays a crucial role in his attempt to explain how nondeterministic brain processes can be relevant for libertarian free choice.¹¹ One of Kane’s main concerns is to explain how neural indeterminacy can ground libertarian control without rendering choices random, arbitrary or irrational.¹² His answer to that question is, roughly, that brain states are also competing reason states. Thus, whichever state is realized by a decision, the agent must have a reason for her action.¹³ Bernáth might be right that one can construct a libertarian account in some respect similar to Kane’s without assuming that reasons are psychological states, but that won’t be Kane’s own account.

¹¹ Actually, following Davidson and Audi, Kane explicitly endorses the mental state model of reasons and reasons explanation. See Robert Kane (1996).

¹² See particularly Robert Kane (1999).

¹³ The original idea is due to Robert Nozick (1980: 295).

Since Kane's model aims to explain the relevant connection between physically nondeterministic processes and rational actions, he must assume not only that reasons as agents' psychological states can compete for causing one of the incompatible actions, but also that there is an objective single case probability (a kind of propensity) for each set of the competing reasons to cause the agent's decision and action. I have tried to defend this conception of choice against some objections, but I certainly would not commit myself to this theory. As Dániel Corsano notes, the idea of ascribing propensities to reasons to cause actions is problematic in many ways, and I fully agree with that.

Nonetheless, I still wish to say something in defense of Kane's view. Corsano claims—correctly, as far as I can see—that if the strengths of reasons are measured in terms of the probability of bringing about the corresponding actions, then there should be some possible situations in which the agent fails to act upon the reason that actually explained his choice. But this seems to be implausible. This is an accurate observation. However, on my reading of Kane's theory, reasons need to have different propensities to bring about actions *only* in situations of 'torn decisions', i.e. in situations where agents have nearly equally strong reasons for performing incompatible actions. And it is less implausible that agents would choose otherwise in some such situations of that type than it is in those cases in which they have a straightforward preference for one type of action over another, no matter how strong that preference is. Of course, this answer invites other difficulties. How can we distinguish such situations from the rest? And why should we attribute a special significance to such 'torn decisions' in the formation of the self and responsible agency? I do not see any plausible answer to these questions, but that is a different matter.

Corsano also criticizes my understanding of pathological aversion claiming that 'pathological' is a medical term, the meaning of which is shifting so that pathological aversion might be compatible with the intentional performance of the averted action in certain situations. I agree. Perhaps using the term 'pathological' was a mistake. I wanted to make it clear only that not *every* form of aversion exempts agents from responsibility. My aversion to seeing blood can exempt me from responsibility for not helping someone who is injured only if it somehow makes it psychologically impossible for me to help. It is very hard to understand the nature of such aversion, and using a medical term might well have been unfortunate (even if in practice we often attempt to identify that type of aversion on medical grounds).

Similarly, Judit Szalai in her informative and challenging discussion about the nature of obsessive-compulsive disorder says that "obsessive-compulsive disorder is not as interesting for the philosophy of action and free will as it is generally thought to be" (p. 89). First, I want to note that if this is so then it is a welcome consequence for the account of free will that I suggest. Psychologi-

cal compulsion is often mentioned as a counterexample to conditional analysis on the grounds that compulsive agents would have done otherwise, if they had chosen so, but we do not hold them responsible. But if such agents do consider themselves responsible, and they are also able to choose and refrain from their compulsive behavior, then the possibility of compulsive behavior is not an objection to the analysis.

Szalai's knowledge of the philosophical and psychological literature on OCD far exceeds mine, and I see no reason to take issue with anything she says about the nature of the phenomenon. Nonetheless, I am somewhat hesitant to accept her conclusion that what we—somewhat misleadingly, I concur—call 'psychological compulsion' does not really undermine responsibility. Certainly, compulsive agents' behavior is *voluntary*, at least in the sense that it is purposeful and intentional. As Szalai says, I do not think that psychological compulsion can be understood as a peculiar form of irectic or motivational state. Compulsive agents are not literally forced to do what they do by some 'irresistible motive'. In general, motive explanations have a logic that seems to me incompatible with the existence of such 'mental forces'. For this reason I think that any psychological state that exempts agents from responsibility must be a cognitive deficiency.

Nonetheless, the issue of responsibility seems to me a bit more complicated. I do not want to—and as stated above I certainly need not to—insist that OCD patients are not responsible. However, consider the following situation: someone forgets to take her son home from his school because she concentrates so much on cleaning the house again and again in order to avoid the risk of some supposedly dangerous infection. *Prima facie*, we would certainly hold this person responsible for her omission. But suppose further that we learn that she suffers from a serious obsessive-compulsive disorder in respect of that type of action. Would this information not change *our* intuition about her moral responsibility? And if the answer is yes, what explains our intuition?

Szalai notes that OCD patients often suffer from obsessive thoughts and images. But this, in itself, is not responsibility undermining. Many 'normal' persons suffer from some recurrent or almost obsessive thought, but this need not affect their behavior in a way which we would regard as incompatible with their responsibility. Szalai also claims that OCD patients often say that they have reason to do what they do. But the question seems to me that of whether or not they can also believe that, at least in certain circumstances, they have *stronger* reasons for stopping doing what they are doing. How can we explain that at least some of them look for medical help? It seems that they do so because they believe that they have better reason to avoid doing what they do than continuing to do it.

I do indeed argue that in many cases irrational behavior need not be explained by any of the agents' reasons. But I do not deny that in many other cases we can

explain an agent's irrational behavior by some of their reasons, provided they also have much better reasons to avoid doing what they do. What the irrational agent cannot say is that her reasons for performing the irrational action were stronger than her reasons for avoiding it. Perhaps that is the case whenever we think that an OCD patient is not responsible for omissions that are explained by her pathological state.

As far as I can see, we can have only two possible grounds to exempt persons from responsibility in those cases when they are able to perform an action that they actually fail to perform. One is that their rational capacities are impaired: they lack the capacity to recognize the reasons for their actions in the relevant situation appropriately. Alternatively, we must say that, although they are able to recognize these reasons, they are not able to adjust their behavior accordingly because they lack the relevant ability to make a choice about what to do. Both deficiencies are of a cognitive nature, but sometimes it is not easy to understand them from the 'normal' perspective.

It is clear that in certain cases agents *do not perform* an action which they ought to perform not because they were unable to perform it, but because, in the circumstances, they could not make a choice about whether or not to perform it. A more difficult question is how it is possible to be *unable to refrain from performing* an intentional—and in that sense voluntary—action. If such actions are possible, then the only explanation of why agents cannot avoid doing what they do seems to be that they cannot make a choice about it, and in this sense their behavior is 'compelled'. Whether there are actual cases of such compulsion, and, if there are, whether OCD patients provide the best example for it, remains an open question.

4. CHARACTER AND CONTROL

Robinson calls me a libertarian compatibilist, and this is indeed a good label for my view. While I believe that the truth or falsity of determinism at the micro-physical, sub-personal level is irrelevant for the philosophical account of free agency, I am a libertarian in the classical 18th century sense of libertarianism. Before Laplace, not many philosophers were interested in the compatibility of free will and *physical* determinism, simply because they did not have the relevant concept of physical determinism.¹⁴ But many philosophers were interested in the issue of how motives explain actions. The so-called 'necessitarians' held that agents' motives necessitate their actions exactly like physical causes were sup-

¹⁴ Although the current issue of physical determinism and free will was to a certain extent anticipated by theological discussions about the compatibility of divine omniscience and omnipotence with free choice.

posed to necessitate their effects. Their rivals claimed that motives can explain actions without necessitating them.¹⁵ I follow the second group of philosophers since in my view the way conscious motives explain intentional actions is fundamentally different from the way physical causes explain physical effects, or the way being afraid explains trembling hands and sweating, or the way unconscious desires explain anxiety and slips of tongue.

It is interesting to note, however, that some of the most influential libertarians, like Peter van Inwagen and Robert Kane, *are* necessitarians in the early modern sense. They believe that motives explain actions by necessitating them, except in a few special situations when conflicting motives are nearly equally strong, so that none of them can necessitate the performance of an action in the presence of the other. On that account, motives work like forces, except that they do not superimpose on each other like physical forces do, and hence only one of the opposing forces can become efficient in the production of action.

I do not find this model of motive explanation plausible at all, but I do agree with such libertarians, called restrictivists, on one point. Van Inwagen (1989) argues that a libertarian must admit restrictivism, because the argument for it relies on the same kind of premises as the alleged proof of incompatibilism does. This is an important insight, but an insight that can help the compatibilist. Since what this shows is that if the argument for restrictivism is not sound, then we have yet another reason to doubt the soundness of arguments for incompatibilism.

I argue that restrictivist libertarians are committed to two implausible theses. First, they believe that agents' will cannot be free unless they have motives that are nearly equally strong but which would cause incompatible actions. This means that most of our actions, no matter how morally or prudentially important they are, do not reveal our free agency. Furthermore, most of the time we behave as compulsive agents do according to the standard theory: at the time of action our motivational states make it inevitable for us to act otherwise. If there is a difference at all between the behavior of a free and a compulsive agent, it must lie somewhere in their past.

Hence the second implausible thesis. Restrictivists postulate that there is a special class of actions which have a distinct role in the formation of an agent's character and motives. The actions belonging to that class do not enjoy special status because they are chosen in some particularly demanding moral situation. They may or may not be. They have special status simply because they are performed in a situation in which the agent has motives with nearly equal force for more than one incompatible action.

¹⁵ For an excellent summary of these early debates see James A. Harris (2005).

Let me grant for the sake of argument that our responsibility could indeed be grounded in such choices. It is still not clear how indeterminism at the sub-personal level could possibly help understand the nature and character-forming effects of these choices. Suppose I once made some such decisions, and having made them in the past grounds my responsibility in the present. Suppose, further, that physicists have found that the best interpretation of fundamental physical laws is deterministic; this means, roughly, that the differential equations with which physicists model such processes have a unique solution. Would this discovery make my earlier ‘torn decisions’ lighter? Did I then have no conflicting motives? Did I not agonize about what to do? Did my earlier decisions fail to influence my present character? Which aspect of my personal history would be changed by that discovery? Libertarian restrictivists claim that they can clearly see the answer. I for one cannot.

But I find the restrictivist story implausible already at the personal level. Although I do not have any view of how our character develops, I do have a view of what kind of actions can reveal our character: only those that are not psychologically necessitated by agents’ motives. So my aim in the book was not to deny that psychological states can necessitate or determine actions. My aim was to argue that whenever an action is so determined it cannot reveal an agent’s character, and hence the idea of restrictivism is inconsistent.

Suppose—perhaps contrary to the facts—that addiction is a form of compulsion. If we believe that addicts can be responsible, then we think this because their present compulsive behavior is the result of their previous choices and actions. But we certainly do not believe that what the addicts do *now* ‘reveal their real character’. We might believe that their present behavior reveals the character they had at the time when they made choices that eventually resulted—intentionally or unintentionally—in addiction. But this means that they must already have had a character when they made those choices, and they were obviously able to make choices then. If we could not choose and act otherwise when we have strong motives, motivated actions could not reveal our character at all.

Robinson argues that an action can be free only if it is not predictable (p. 70), whereas Bernáth argues against my claim that, in normal cases, motives do not determine actions. As he says, “not only psychological experts but also ordinary people who are good judges of character can predict decisions” (p. 113). Interestingly, again, in the early modern period, it was necessitarians, and *not* their libertarian rivals, who held that if an action is not psychologically determined, then agents’ behavior must be unpredictable. Strangely enough, in our time it is libertarians who claim that unpredictability is a condition of ‘libertarian choice’ (I assume that both Robinson and Bernáth argue for a libertarian position). In

this debate, I am still with the early modern anti-necessitarians.¹⁶ I am convinced that predictability of behavior has nothing to do with free will, choice, or the ability to do otherwise.

It is obvious that unpredictability is sufficient neither for freedom nor for its absence. Unpredictable weather is neither free nor unfree. And the most unpredictable forms of human agency, like schizophrenic or paranoid behavior, are the least free. (Some post-modern thinkers might disagree, but then they cannot have in mind freedom as a condition of moral responsibility.) The idea that unpredictability is at least a necessary condition of human freedom derives from the prejudice that successful prediction *must* be based on some sort of causal necessitation.¹⁷ This seems to me false even in the case of inanimate objects, but it is certainly wrong about intentional human behavior. We normally predict intentional human behavior by ascribing certain aims to agents. We also assume, quite plausibly, that they act in order to achieve those aims. But it is totally idle to add to this that having those aims must causally necessitate their choices. Given that we know virtually nothing of the nature of the alleged causal connection between motives and choices, the postulation of such a connection can hardly make a person's behavior *more predictable*.

Relatedly, Robinson claims that “the conditional analysis is really saying that different choices would have come about *under different circumstances*, where *difference of choice alone* does not constitute ‘different *circumstances*’, in the sense intended,” and hence that my theory “is not really a conditional theory at all” (p. 72). Perhaps not. I certainly wouldn't object to not calling my theory conditional. I myself also say at one point that calling my preferred account ‘conditional’ can give rise to misleading associations. I do argue, however, that what we are interested in when we ascribe responsibility to agents is whether or not they would have done otherwise, if they had chosen to, assuming that in those circumstances they were able to make a choice.

The ability to choose is a condition of responsibility, since it renders it possible for us to control whether or not, or exactly when, we act upon our motives and reasons. Indeed, I cannot give a conditional analysis of choice, but that has nothing to do with the possibility of choosing otherwise. We can give a perfectly

¹⁶ I cannot help invoking here Jacob Bryant's succinct response to Joseph Priestly, one of the most influential early modern necessitarians: “granting that people in the same circumstances would always act uniformly in the same manner: yet in respect to the mind and freedom of choice, I do not see how they are at all affected. If I had full liberty to choose in one instance, I should have the same in another; and even if I were to repeat it an hundred times. You insist, that the repetition of the same act must be the effect of necessity. But if that, which I do, be the result of forecast and reason, it will at all times be an instance of my freedom in respect to election”. (Cited by Harris 2005: 170.) For a similar view in contemporary discussions see Jonathan Lowe (2008).

¹⁷ As John McDowell (1981: 154) says, this is “the quasi-hydraulic conception of how reason explanation account for actions”.

good conditional analysis of a coin's power to fall heads or tails when thrown, as well as many other nondeterministic powers. My reason to reject a conditional analysis of choice is that there are cases when we are responsible, not for *making* a certain choice, but rather for not making any, although we could, like in the cases of negligence, carelessness or forgetfulness. In such cases, agents possess the ability to choose and have reason to exercise it, but they do not. This shows that even if agents' responsibility requires that they would have done otherwise had they chosen to and retained the ability to make the relevant choice, responsibility does not require agents' actual control over the exertion of those abilities. It is for this reason that the relevant ability of choice does not seem to be susceptible to conditional analysis.¹⁸

Anna Réz notes that even if my account of free will does not do worse—or perhaps does better—than others in explaining the conditions in which negligent agents are responsible, such cases remain puzzling, since they seem to defy the otherwise very intuitive control condition of responsibility. As she says, “Cases of carelessness, forgetfulness or negligence arise exactly because certain reasons, facts, considerations do not even cross the agent's mind. But how could we fairly hold anyone responsible for something over which she did not have any kind of conscious control?” (p. 42).

One way to solve the problem is to try to introduce diachronic conditions of responsibility, and then say that responsibility for such behavior—usually, but not always, some omission—‘traces back’ to some earlier acts over which the agent did have control. I have the same problem with such proposals as I do with restrictivism as discussed above. I agree that there are cases in which the ‘tracing strategy’ can work, but it seems that it works only in a very special and restricted class of cases. Few of us have ever ‘trained themselves consciously and intentionally’ to be absent-minded, negligent, or careless. And it is very hard to understand what it would mean to make ourselves into conscientious and scrupulous agents rather than absent-minded and careless—particularly when some specific, unexpected situation occurs. Nonetheless, sometimes we can be responsible for our behavior even if we seem to have neither direct nor indirect control over it.

I do not pretend to be able to solve this puzzle, but I do hold a certain view about responsibility and control which is, in my estimation, not very widely shared among philosophers who are interested in the problem of responsibility. It seems to me that it is *a feature of the human condition* that we do not have perfect control over many aspects of our life for which we are nonetheless responsible. This means that when we want to understand the nature of control which is

¹⁸ Although I agree with Robinson that Moore's proposed condition—that we would have made a choice had we chosen so—might be more promising than I thought it was when I wrote the book.

required for responsibility we should not be guided by the thought that *responsibility* and *luck* are not compatible.

Most of us would think that the stoics' search for a condition in which we can totally neutralize the effect of luck on our happiness is in vain. Why should we have a different opinion about theories that aim to neutralize the effect of luck on our responsibility? Just as our happiness depends partly on us and partly on some personally uncontrollable circumstances, so does our responsibility. It might be difficult to accept this in an age dominated by 'positive thinking', but one need not be a 'depressive realist' in order to see the irresistible influence of fate on our lives, both in matters of happiness and in matters of responsibility.

This view of mine about the human condition does, however, have consequences as to how we should understand the connection between someone's being responsible, her taking responsibility, and others holding her responsible. Certainly if one is responsible then, in some sense, one ought to be held responsible; at least one ought not to be held non-responsible. However 'holding responsible' is an ambiguous expression. In one sense the claim is that if someone is responsible then we ought to *judge* that she is responsible, simply because judging otherwise would be a mistake. In another sense, 'holding responsible' means possessing a moral attitude toward someone else; that is why we can qualify it, as Anna Réz does in the passage quoted above, as fair or unfair.

If I am right about luck and responsibility, then from the fact that in certain circumstances it is correct to judge that someone is responsible, say, for her negligent behavior, it does not follow at all that it would be fair to cultivate a moral attitude toward her. In many cases of prudential weakness, it is more adequate to feel compassion toward, and offer help to, the weak than it is to 'hold her responsible' or blame her. Exactly in the same way, in some cases of morally relevant omissions, even if we judge correctly that the agent is responsible, because in fact she is, we should think twice before we 'hold the agent responsible' in the sense which is compatible with blame.

Being responsible is one thing; it is the possession of certain powers that enable us to control our behavior in certain ways. Holding someone else responsible, in the sense of feeling it justified to blame, is another. In many cases when agents fail to exercise active control over what they do, we would still think ill of them, if they denied their own responsibility. But this is compatible with thinking equally ill of those who would blame them for their bad moral luck. We cannot control every aspect of our own fate. But we ought to control our own stance towards the fate of others.

REFERENCES

- Ekenberg, Tomas 2009. Power and Activity in early Medieval Philosophy. In Juhani Pietarinen & Valtteri Viljanen (eds.), *The World as Active Power*, Leiden and Boston: Brill, 89–111.
- Gyenis, Balázs 2013. Determinizmus és interpretáció. *Magyar Filozófiai Szemle* 57, 85–100. [In Hungarian.]
- Harris, James A. 2005. *Of Liberty and Necessity. The Free Will Debate in Eighteenth-Century British Philosophy*. Oxford: Oxford University Press.
- Kane, Robert 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kane, Robert 1999. Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism. *The Journal of Philosophy* 96, 217–240.
- Lehrer, Keith 1968/1982. Cans without Ifs. In Gary Watson (ed.), *Free Will*. Oxford: Oxford University Press, 41–45.
- Lewis, David 1997. Finkish Dispositions. *The Philosophical Quarterly* 47, 143–158.
- Lewis, David 2009. Ramseyan Humility. In David Braddon-Mitchell & Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*. Cambridge MA: MIT Press, 203–222.
- Lowe, Jonathan 2008. *Personal Agency*. Oxford: Oxford University Press.
- Mackie, John L. 1973. Dispositions and powers. In his *Truth, Probability, Paradox*, 120–153. Oxford: Clarendon Press.
- Mackie, John L. 1977. Dispositions, Grounds, and Causes. *Synthese* 34, 361–370.
- Martin, Charles Burton 1994. Dispositions and Conditionals. *The Philosophical Quarterly* 44, 1–8.
- McDowell, John 1981. Non-cognitivism and rule-following. In S. Holtzman & Christopher M. Leich (eds.), *Wittgenstein: To Follow a Rule*. London: Routledge and Kegan Paul, 141–162.
- Mellor, David H. 1974. In Defence of Dispositions. *The Philosophical Review* 83, 157–181.
- Mellor, David H. 2000. The Semantics and Ontology of Dispositions. *Mind* 109, 757–780.
- Nozick, Robert 1980. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Setiya, Kieran 2000. Practical Knowledge. *Ethics* 118, 388–409.
- Strawson, Galen 1994. The Impossibility of Moral Responsibility. *Philosophical Studies* 75, 5–24.
- van Inwagen, Peter 1989. When is the Will Free? *Philosophical Perspectives* 3, 399–422.
- Vihvelin, Kadri 2004. Free Will Demystified: A Dispositionalist Account. *Philosophical Topics* 32, 427–450.