HUNGARIAN PHILOSOPHICAL REVIEW

Vol. 58. (2014/1)

The Journal of the Philosophical Committee
of the Hungarian Academy of Sciences

Are We Free After All?
Reading Huoranszki

Contents

ARE WE FREE AFTER ALL? READING HUORANSZKI

Foreword

Ferenc Huoranszki's book (*Freedom of the Will: A Conditional Analysis*, New York: Routledge, 2011) testifies to the impressive revival of analytical philosophy in Hungary during the last twenty-five years or so. Undoubtedly, Huoranszki has been one of the protagonists of this revival. Readers of this book will quickly understand why. The work bears witness to his philosophical erudition, admirable attention to detail, and tenacious resourcefulness in defending his positions.

Arguably the greatest achievement of the book, however, is to have demonstrated that the possibility and nature of free will is not a sectarian issue of limited interest to the non-specialist. Huoranszki manages to drive the point home that the debate about the compatibility of free will and determinism is intimately tied up with central problems of metaphysics, the philosophy of sciences, and metaethics. The frequent historical references also make a persuasive case that the most influential contributions throughout the history of philosophy have always come from authors who appreciated just how multifaceted the problem of free will was.

This special issue is based on a workshop devoted to Huoranszki's book at the Central European University (CEU) in the autumn of 2011. Each contribution to the workshop was invited to focus on a different chapter of the book. However, the complexity of the book's conceptual apparatus made cross-chapter "transgressions" unavoidable and in fact quite welcome. The editors of the special issue decided to proceed in the same manner, and so some articles cover more than one chapter of *Freedom of the Will*.

We are happy to observe that the workshop and the special issue highlight the renaissance of analytical philosophy in Hungary and the significance of Huoranszki's work in another way as well. Many of the contributors are researchers early in their careers who were inspired and often taught by "Huo", as he is commonly known among friends, colleagues and students. We therefore have every reason to expect that this conversation will continue well into the future.

The editors wish to thank all this issue's authors and anonymous reviewers for their work, as well as David Robert Evans for his painstaking and impeccably

prompt copy-editing of the entire manuscript. We are also grateful to Professor Huoranszki for taking the time to reply to the authors with such care and at such length. We are confident that even hard-boiled connoisseurs of the free will literature will profit from perusing this exchange between Huoranszki and his critics.

Finally, it is our pleasure to acknowledge the financial and organizational support provided by the research project 'What is it to be human?' hosted by CEU's Department of Philosophy and sponsored by the Hungarian Innovation Office (formerly NKTH) and Mag Zrt. In particular, we have greatly appreciated project leader Professor Gábor Betegh's and project member Dr. Anna Réz's dedication, without which neither the workshop nor this special issue would have been possible.

Tibor Bárány, András Szigeti

Ferenc Huoranszki

Freedom of the Will and Responsible Agency

The notion of free will as used by philosophers is a term of art. More often than not, such terms—such as substance, form, intentionality, reasons etc.—are introduced in philosophy in order to single out a problem rather than to solve it. Contrary to the opinion of many modern philosophers, who have been critical of the use of such terms, they can indeed help identify a cluster of problems that are not merely created by the introduction of the technical concept itself. Philosophical discussions often revolve around some such concept. But this does not mean that the problem is only conceptual. Many genuine philosophical issues can be made more easily approachable by sharpening our understanding of such technical concepts.

However, even if the introduction of such technical terms is unavoidable and useful for philosophy, they can lead to confusion unless we carefully circumscribe the task for which we want to use them. It seems to me that sometimes this is exactly the case with the philosophical concept of free will. For there are many different notions of free will which are often used in philosophy, but which set at least partially incompatible tasks for those who want to provide a theory about it. In what follows I shall argue, first, that we have reasons to accept one of the most traditional concepts of free will, one that might be called the metaphysical conception of free will. Then I shall briefly present and defend a particular theory about the freedom of the will so understood.

1. FREE WILL AND AUTONOMY

At least one traditional notion of free will understands it as a specific psychological ability, the ability 'to will' or issue volitions, which, according to this approach, can be done 'freely' or 'unfreely'. This understanding of free will—certainly predominant, for instance, in the Scholastic period—has been the standard target of criticism of many modern philosophers ever since Locke. Locke himself has no objection to the notion of will and willing itself. What he finds 'unintel-

ligible' is the idea that 'freedom' or 'liberty' can qualify the exertion of the will (Locke 1689/1975: 245–246). But more recently, and very influentially, Ryle has argued that talk of volitions as dated token mental events is itself unintelligible (Ryle 1949: 62–69). If they are right, then freedom of the will, understood as the agent's capacity to issue volitions in some specific way, is a chimera.

I am not certain whether a philosophical analysis of intentional human action must rely on the notion of volition, but it is interesting to observe that contemporary psychologists seem to use it often and without any scruple (Lowe 2008: 81–82). This does not, of course, show that the concept is philosophically interesting.[1] But even if I do not find Ryle's arguments against volitions particularly convincing, I share Locke's worry about understanding free will as a capacity of issuing volition freely.

However, there are psychological concepts of free will which seem philosophically more interesting than the idea that human persons can in some sense 'will' freely. The notion of free will is tied to the philosophical problem of human agency. Perhaps God and angels can enjoy freedom of will, but the only way for us to understand it is to look into ourselves. One distinguishing feature of normal human adults as the paradigmatic representatives of agency is that they are capable of autonomous action. When I say this, I have in mind some very broad sense of autonomy. By autonomy I mean the specific way in which agents can control their actions so that they can regard their intentional behavior as their own, and which can warrant their conviction that the result of their actions is their own creation, something that truly originates in them.

It is obvious that not every human action, not even every intentional piece of behavior, is autonomous in this sense. When freedom of the will is understood as a condition of autonomy, we might say that it is a condition of what David Velleman (2000) calls "full-blooded action". It is arguable that "full-blooded actions" play a special role in a philosophical understanding of agency. As Velleman claims "The question is the nature of agency itself, and agency, like any capacity, fully reveals its nature only when fully exercised. We therefore want to know what makes for a paradigm case of action, a full-blooded action, an action par-excellence" (Velleman 2000: 189).[2]

Details apart, full-blooded actions are actions that are performed under some specific actual mental control exercised by the agent performing them. There is much debate in contemporary philosophy of action about how this particular sort of control could best be captured. But this is not my concern here. I merely want to mention that one possible way to characterize actions as "full-blooded

[1] See, for instance, Adams and Mele (1992) who argue against the view—defended most recently, for instance, by Ginet (1990)—that we cannot understand intentional action without postulating the mental acts of volition.

[2] A similar view is developed by Michael E. Bratman (2000).

actions" in Velleman's sense is to say that such actions reveal our agency exactly because they are done of our own free will.[3] Freedom of the will, then, is a condition in which agency is exercised in such full-blooded actions. Hence it is a condition that should be formulated in terms of mental states or processes which can guarantee that agents can exercise full control over what they do.

This understanding of freedom of the will has a natural affinity to its more traditional notion, since it is possible to assume that the relevant kind of control is exercised when agents issue volitions freely.[4] However, it is important to emphasize that the notion of free will as the condition of autonomous or full-blooded action does not require that we introduce volitions as special mental token events by which agents can control their actions. There are a number of different attempts to understand free will as a condition of full-blooded action which do not use the concept of volition in its more traditional sense.[5] The important point is that the notion of free will is here understood as a capacity, the proper use of which can make action full-blooded and thereby autonomous in a broad but important sense.

2. FREE WILL AND RESPONSIBILITY

The concept of free will as a condition of autonomous agency is not the only one in philosophical currency. Normal human adults are not only capable of autonomous actions; they are also the subjects of responsibility ascription. If at least some of such ascriptions are correct, then responsibility for our own actions may also be an essential feature of our agency. In this context, the concept of free will is introduced in order to specify the necessary and/or sufficient conditions of a person's responsibility. Since we are responsible for our full-blooded actions, it might, if anything, be tempting to assume that freedom of the will as a condition of full-blooded actions and freedom of the will as a condition of responsibility are one and the same. Furthermore, it may at first even seem plausible that freedom of the will is a condition of responsibility *exactly because* it aims to capture what makes some of our actions "full-blooded".

I wish, however, to argue that, no matter how plausible this may initially sound, this is nevertheless a mistake. I certainly do not want to argue against philosophers using the notion of free will with the purpose of trying to capture

[3] I am not sure whether Velleman himself would agree with this, but this is not what matters here. He certainly seems to agree that we can be responsible for actions that are not full-blooded in his sense (see Velleman 2000: 127), and that is all that is relevant to the point I want to make here.

[4] See, for instance, Carl Ginet (1990) and Jonathan Lowe (2008).

[5] Such kinds of analysis include, among others, Harry Frankfurt (1971/1988), and also some 'agent causalist' accounts like Timothy O'Connor (2000).

a mental capacity the exertion of which turns a piece of intentional behavior into a full-blooded or autonomous action. But I do want to say that *that* concept of free will is not the same as freedom of the will understood as a condition of moral responsibility, and that conflating the two concepts cannot enhance our understanding of either.[6]

An obvious *prima facie* reason why the conditions of full-blooded actions cannot be the same as the conditions of responsibility is that we are often responsible for things that we omit. And even when we do decide to omit some action, i.e. when our omission is not just the consequence of our negligence, our responsibility can hardly be explained by the way we control the performance of an intentional action. More importantly, however, we are responsible for our negligent behavior, whether or not negligence involves some intentional action. And whatever negligence is, it certainly occurs when some pertinent form of control is absent. Thus, if we are responsible for any such actions and omissions, free will as manifested in our full-blooded actions cannot be a necessary condition of responsibility.

This conclusion is, in fact, granted by many philosophers. They seem to agree that freedom of the will is a sufficient condition of responsibility, but not a necessary one. Consider responsibility for someone's desires, emotions, and beliefs. Few philosophers would deny that we can be responsible for these attitudes even if we cannot have them 'of our own free will'. Or consider the consequences of our actions: we can be responsible for some of them even if we have not intended them, or even if we had only limited, if any, control over them. In fact, one might even argue that the problem of responsibility for omissions and negligent behavior is similar to the problem of our responsibility for the consequences of our actions. Since in such cases we lack direct intentional control, it is at least contentious whether freedom of the will can or cannot be the necessary condition of responsibility for our attitudes and for the consequences of our actions.

However, there is a deeper consideration that supports the view that free will is not a necessary condition of responsibility, but rather expresses the form of control required for full-blooded actions. Responsibility itself, one might say, is a condition: it is a condition under which one can be legitimately *blamed*. Now, if we can blame people for their mental attitudes and for the consequences of their actions, but we do not think that such attitudes and those consequences always depend on an agent having free will, we might insist that free will, whatever it is, cannot be a necessary condition of responsibility. Perhaps in the case of some

[6] In this I agree with Harry Frankfurt (1971/1988), who claims that what *he* calls 'freedom of will' is not an issue about responsibility. However, I obviously disagree with Frankfurt that we should understand freedom of the will on the model of freedom of action, i.e. as having the will one wants to will.

actions, freedom of the will is *sufficient* for an agent's responsibility. And it is sufficient, because such actions best reveal our agency. But since responsibility extends beyond full-blooded agency, free will cannot be a necessary condition of responsibility.

Now I would certainly agree that we are not responsible *only* for our full-blooded actions. However, I want to claim that there is a perfectly natural understanding of free will that regards it as *a necessary, but not sufficient* condition of responsibility for our actions and omissions. Certainly, free will does not seem to be necessary for responsibility for our mental attitudes and for the consequences of our actions. But this does not show that it is not a necessary condition of the direct manifestations of responsible agency, or that the only manifestations of such agency are our full-blooded actions. In general, this does not prove that free will is not an essential aspect of responsible agency. For the conditions of responsible agency and the conditions under which someone can be rightfully blamed do not coincide.

Consider the person who did something wrong under hard circumstances, strong pressure, or serious threat. Perhaps she made a bad choice, and did something, or failed to do something, which she later came to regret. It seems to me that in such cases blaming her would be an inappropriate response. But whatever the appropriate response is, this does not tell us much about whether the person acted, even in those circumstances, as a responsible agent. Or consider the cases of prudential weakness, like someone not following her self-imposed, health-preserving diet. She is responsible, but it seems more appropriate to feel sorry for her or perhaps even trying to help her than it would be to blame her. Blaming itself is not just a mental attitude or an automatic response—like feeling disgust or anger, for instance, might be—but a morally evaluable action for which we might be responsible. Moreover, there are many types of actions—we choose where to live, what job we do, with whom to mate—which are not to be judged morally or otherwise, but which we do perform as responsible agents. Thus it is far from obvious that the conditions of responsible agency should be understood with reference to the conditions in which we can legitimately blame someone.

3. ALTERNATIVE POSSIBILITIES

However, if freedom of the will is not the condition of full-blooded agency, and neither it is necessary for responsibility *simpliciter*, what kind of interesting philosophical work can be done with reference to this concept? And if we do not want to use it in order to identify the special mental capacity the exertion of which is manifested by particular acts of volitions either, is there any traditional meaning of this term that justifies the attempt to understand a theory of free will

as a theory about the conditions of responsible agency? I'd like to think there is. For there is another traditional notion of freedom of the will that aims to capture the metaphysical, rather than psychological, conditions of responsibility. This metaphysical condition is that an agent is responsible for what she did or failed to do only if she could have acted otherwise. And freedom of the will seems to me a necessary condition of responsibility exactly in this sense: that no one can be responsible for an action or omission unless she could have done otherwise.

Saying that freedom of the will can be understood as a metaphysical condition in which we are responsible for what we do or fail to do is not, of course, to say that it can be understood without reference to agency. There is a difference between what a person *can do*, even if *she* doesn't do it, and what *might happen* to her, even if *it* doesn't happen to her. 'Can do otherwise' must express some agent-specific sense of alternatives, and not just the existence of some abstract possibilities. This seems to me an important constraint on any free-will-related understanding of alternative possibilities. If we try to understand the relevant possibilities by 'discovering' which possible worlds are accessible and which are not, we cannot come very close to understanding the conditions of responsibility. I do not object to talking about possibilities in terms of accessibility to possible worlds *per se*. But possible worlds can do their most useful work when we try to model our *modal inferences* in certain contexts. They won't help much in understanding what makes any particular modal statement—like the one that a person could, in given circumstances, have done otherwise—true or false.

In my view, the best way to understand what agents can or cannot do is to rely on counterfactual conditionals. It is with the help of such counterfactuals that we can define the relevant accessibility relations. And which counterfactuals are relevant is entailed by the properties of the objects. Thus, if 'could have done otherwise' is relevant for responsibility for actions and omissions, then it should be understood in terms of counterfactual conditionals related to the agent's properties. It follows that the crucial issue for any approach that interprets the problem of freedom of the will as the problem of agent-specific alternative possibilities is to specify the common characteristics of the relevant properties with the help of the counterfactual conditionals that their ascription entails. There are different ways to interpret the relevant counterfactuals, but here I wish to mention only two that seem to me historically the most influential.

4. FREE WILL AND REASONS

Some versions of the conditional analysis follow the ancient observation that the distinctive feature of the human animal is her rationality. Rationality might mean many things, but one of its important aspects is certainly that the behavior of normal human beings can be responsive to their reasons. And since it is beyond

any doubt that we do not hold people responsible for their actions unless they are able to recognize reasons for or against them, it seems natural to think that the property, the possession of which is relevant for an agent's responsibility, should be understood in terms of responsiveness to reasons. The responsibility-related sense of 'can do otherwise', then, is roughly that an agent is responsible if she is such that, at the time of the action or omission, she would have done otherwise if she had sufficient reason to do so.[7] The fundamental idea behind these proposals is that the capacity to recognize reasons, even if necessary, is not sufficient to hold agents responsible. In order to be responsible, agents also have to be such that they can control their behavior with those reasons.

However, it seems to me that responsiveness to reasons is neither sufficient, nor necessary for responsibility, and that free will as a condition of responsible agency is not a rational capacity at all. *Prima facie*, responsiveness to reasons is not necessary for responsibility, since we can behave irrationally, i.e. can act against our better reasons. But we are responsible for such actions nonetheless. Moreover, people can act and omit actions intentionally, without any reason, and with sufficient reason to do otherwise. But this does not prove that they are not responsible or that they are not acting of their own free will. More importantly, people can have, and act for, their own sufficient reasons *without* being free and responsible. Certainly, the possession of some rational capacities is necessary for most kind of responsibility. But this does not mean that freedom of the will should be understood as a rational capacity.

Speaking somewhat metaphorically, if we try to understand the conditions of responsibility in terms of the capacity to be responsive to reasons, an essential element of our agency disappears. By this condition we can understand only how *agents' reasons* guide their actions, but what we need to understand is how *agents themselves* can control their behavior. And what such control requires is exactly that, by apprehending their relevant reasons, agents may or may not act accordingly. This is exactly why we need the notion of freedom of the will as the ability to act otherwise. It is this kind of freedom which makes it possible for agents to act for their reasons *in such a way* that makes them responsible for what they do or fail to do. Consequently, free will is the necessary metaphysical condition of responsible agency. The capacity to apprehend reasons for actions is only one of its necessary cognitive conditions.

[7] Different versions of this view are developed by Alfred J. Ayer (1982); Donald Davidson (1980); John Bishop (1989); Philip Pettit and Michael Smith (1996). I also argue that John Martin Fischer and Mark Ravizza's 'semi-compatibilist' account of responsibility (Fischer–Ravizza 1998) can also be interpreted as a version of this account of free will. See Huoranszki (2011: 105–107).

5. RESPONSIBILITY, ACTIONS, AND OMISSIONS

There is another tradition in philosophy that aims to understand the agent-relevant sense of 'can do otherwise' with the help of counterfactual conditionals, and which I find worthy of further development. According to that tradition, freedom of the will should be understood as an agent's ability to make choices in order to guide her own behavior. It is tempting to think that, again, such an understanding identifies freedom of the will with a mental capacity, the capacity to make choices the function of which is to issue intentions and thereby guide the agent's action.

However, this is not quite so. On the one hand, an agent's ability to make choices about the particular actions for which they are responsible does indeed play a crucial role in this theory of free will. And the ability to make the relevant choice cannot be understood in terms of reasons for which the agent might act. On the other hand, according to this interpretation of free will as a condition of responsible agency, freedom of the will also involves the possession of certain performance abilities.[8] The fundamental idea is that agents are responsible only if the exertion of their performance abilities can depend in some pertinent way on the possession of their ability to choose.

The first important issue for such a theory of free will is to decide whether or not the relevant performance abilities should be intrinsically or extrinsically identified. It may be tempting to think that the relevant abilities must be intrinsic in the following sense. As I mentioned earlier, most philosophers would agree that the conditions under which agents are responsible for their actions are not exactly the same as the conditions under which they are responsible for the consequences of their actions. Free will can be a condition only of the former because agents can have direct control only over their own actions.

Some philosophers go further, however. They think that since we perform other actions by moving our body, only our own bodily behavior can be under our own direct control.[9] Thus we must be directly responsible for our bodily movement and only derivatively responsible for other intentional actions identified with reference to the movement's results. For instance—to use the well-worn Lockean example—if someone ought to leave a room, she is directly responsible for moving or not moving her body in certain ways, and she is only derivatively responsible for leaving or not leaving the room. Since responsibility for our intentional actions derives from our responsibility for our bodily movements, the

[8] According to Randolph Clarke's useful classification, this is not a 'willist' account of free will. About the difference between the 'willist' views and their rivals, see Clarke (2003: 203). For a criticism about his way of drawing the distinction, see Huoranszki (2011: section 4.4).

[9] See, for instance, Harry Frankfurt (1982/1988); R. Jay Wallace (1998: 259); and Robert Audi (1993: 233).

performance abilities that are necessary for freedom of the will should be understood as the agent's ability to move her body in a certain manner.

It seems to me, however, that this is an implausible understanding of responsibility-related performance abilities. From the fact that when we do things intentionally we do so *by* moving our bodies, it does not follow that we are directly responsible for our bodily movements, and we are responsible only derivatively for intended results of these movements. In fact, it seems to me that exactly the opposite is true. We are responsible for our bodily movements only derivatively, i.e. to the extent that we do them with the aim to perform an extrinsically identified intentional action. That is, the person in the Lockean example is directly responsible for leaving or not leaving the room, and only derivatively responsible for moving or not moving her body in certain ways.

First of all, the idea that we are directly responsible for our bodily movements because we have direct control over them is a contentious one, to say the least. In most cases we have such control only through and by the representation of the extrinsic result which we aim to attain by moving our body. Our actions are guided directly by the representation of the intended extrinsic results, as in many cases we would not be able to perform the respective bodily movements without having such aims in view. Think of the act of throwing a ball towards the basket, and try to understand or perform a part of that bodily movement without the guiding aim. If you can do that you might be an excellent pantomimic; but you are not necessarily a good basketball player.

Second, the idea that we have more control over our bodily movements than over their consequences is simply an illusion. When the player throws a ball, with the intention of getting it in the basket given the appropriate circumstances (no wind, no obstruction), the action's result is uncertain not because of the external circumstances, but because of the limitations of the player's ability to control her body in the required ways. What exactly the role of our bodily movements is in the performance of our intentional actions is a moot question, but the idea that we have more direct control over them than over the actions that are identified with reference to their aimed results does not seem to me justified at all.

More importantly for the problem of responsibility for actions and omission, however, and independently of the issue of control, it is extrinsically identified actions—i.e. actions identified in terms of some of their results other than the bodily movement—for which agents are directly responsible, for the following reason. Even if freedom of the will as a condition of responsibility is not to be understood in terms of responsiveness to reasons, it must be understood such that it makes it possible for the agent to act upon those reasons. For we judge an agent's behavior partly on grounds of those reasons for which they act. But an agent's reasons for performing an extrinsically identified action and their reasons for performing the bodily movements by which they carry out the former are rather different. A person may have reason to pass a ball to another person

rather than not pass it, and he can have a reason to pass it with his right hand rather than with his left hand or with his leg. But the reasons that ground his choice to pass it are different from the reason for passing it by using his right hand, for instance.

In fact, when it comes to an agent's responsibility, it is only in exceptional cases that we are at all interested in finding out her reasons for the choice of a bodily performance. It is easy to see why. When an assassin aims to kill someone with a gun, we are interested in why she wants to kill. We are not interested in why she has chosen the particular type of gun with which she kills. Analogously, we are not interested in her reasons for pulling the trigger by moving her body in a particular way. Moving our body is a means by which we do other things, but our reasons for choosing the means—if we have any—is not the same as our reasons to aim an action for which the bodily movements are means. Consequently, from the fact that we do other things by moving our bodies, it simply does not follow that we are responsible for the aimed result by virtue of, or as a consequence of, our responsibility for a bodily movement. On the contrary, we are responsible for our bodily movements just because they were meant to be part of some extrinsically identified action.

If the ability to perform actions for which we are directly responsible can be extrinsic, then we can ascribe such abilities only if certain external conditions are also satisfied. Our agency extends beyond our body because the possession of the responsibility relevant performance abilities are sensitive to the circumstances in which we act or fail to act. If someone locks me in a room, I can lose my ability to leave it without undergoing any intrinsic change. And then I am not responsible for not leaving it because I cannot do otherwise, exactly in the sense that I am not able to.

As a consequence, the kind of ability the possession of which is necessary for our responsibility is specific to each situation. Certainly, even in a locked room, I have the general intrinsic ability to move my body in a way that in other circumstances would result in leaving a room. But when we say that it is a necessary condition of an agent's responsibility that she can do otherwise, in the sense that she is able to perform an actually unperformed action, we refer to an ability the possession of which is sensitive to the agent's particular circumstances.

Thus, freedom of the will as a condition of responsibility must be understood as a generic condition that can be applied to all specific circumstances in which we hold agents responsible. For this reason, a theory about the freedom of the will needs to identify the generic metaphysical conditions in which we perform or fail to perform a specific action such that we are responsible for it.

6. THE ABILITY TO DO OTHERWISE

If freedom of the will is the ability to do otherwise—including, as always, the ability to avoid performing some actions that we have actually performed—a theory of free will is a theory about how to understand that ability. Although this is a contentious issue, it seems to me that we can identify abilities only with the help of counterfactual conditionals. It is for this reason that I believe that, despite its recent unpopularity, conditional analysis is the best theory of the freedom of the will we can have.

A conditional analysis need not aim for reduction. Its aim is to specify the relevant abilities, not to show that freedom of the will as the ability to do otherwise is in fact something other than it seems to be. For a long while, conditional analyses of powers meant to show how the use of the so-called 'disposition terms' is compatible with the belief that powers, abilities, or potentialities are not 'real' characteristics of objects. The way in which such reductive analyses usually proceed is that, first, they offer some general account of counterfactual conditionals, and then, relying on that account, they try to analyze powers away in power-free terms. In my view, however, there is no general account of counterfactuals that can be independent of the kind of context in which we want to use them. Thus the question is not how to explain away powers with the help of some such conditionals, but rather how to formulate the conditionals so that they can best express which properties we have in mind when we talk about powers or abilities.[10]

In the most general terms, the conditional analysis of any powers must identify, first, what kind of state or event counts as the power's manifestation; they must then specify those circumstances in which the power would become manifest. Although there are difficult questions about what exactly counts as the manifestation of a power,[11] such questions should not detain us here, since in the context of free will it is obvious that the manifestation of the relevant ability must be the performance of a specific kind of action. A more difficult issue, however, is that of how we can specify the circumstances in which the relevant power would become manifest.

There was a time when philosophers thought it sufficient to specify the relevant circumstances with reference to what they called the 'stimulus' for the manifestation. But it is now generally agreed that it is possible that the stimulus occurs simultaneously with other changes in the object that prevents the occurrence of the manifestation event.[12] There are attempts to amend the analysis, in

[10] I say more on this issue in Huoranszki (2012).

[11] See Jonathan Lowe (2010) and Jennifer McKitrick (2010).

[12] In fact, one of the first such examples was used by Keith Lehrer in his extremely influential objection against the conditional analysis of the ability to do otherwise. See Lehrer (1968/1982).

the reductivist spirit, to identify the relevant circumstances without mentioning the power itself.[13] For reasons that I cannot detail here, I am skeptical about the prospect of any such analysis. But more importantly, if our aim is not reduction but elucidation, we can simply add to the conditions of manifestation that the object does not change its relevant power simultaneously with the occurrence of the stimulus event.[14] For it is just obvious that no object that is losing the power when the stimulus necessary for its manifestation occurs can possibly make that power manifest.

However, there is a more delicate problem about the nature of the manifestation conditions that need to be addressed. This problem arises because powers can be more or less generic, and freedom of the will as a condition of responsibility requires that agents possess some *specific* powers. Consider a person who is sound asleep. It might be true of her that she would have answered an important phone call, if she had chosen to. Nonetheless we would be reluctant to say that she is responsible for not answering the call or that she avoided answering it of her own free will. Although she is certainly able to answer a phone in some generic sense of ability, she was not able to answer it *there* and *then*, and hence she lacked the specific ability or power that is a condition of her responsibility. Moreover, her inability is the result of the fact that she was in a state that made it impossible for her to satisfy the conditions that would result in the alternative form of behavior. Thus, as the possibility of the occurrence of the 'stimulus condition' can depend on some intrinsic ability of the person who is credited with the power, in order to identify the agent's specific power, we need to add to the conditions of manifestation the condition that the person should also have that stimulus-enabling ability.

If we take these observations into account when trying to understand the nature of counterfactual conditionals used with the purpose of specifying an agent's responsibility-related powers, it seems easy to respond to some objections on the basis of which the conditional analysis of the free will as the ability to do otherwise has for a long time been routinely rejected. Freedom of the will is, roughly, the agent's ability to perform some actually unperformed action in the sense that she would do otherwise, if [1] she so chose, and [2] did not change with respect to her ability to perform the action when so choosing, [3] nor with respect to her ability to make a choice about the relevant kind of action.[15]

This formulation is not just an *ad hoc* adjustment made on the traditional simple analysis in order to eschew the standard objections against it. It is a consequence of our earlier considerations about which counterfactuals need to be

[13] The most influential attempt is David Lewis (1997/1999).

[14] See David H. Mellor (2000).

[15] For a more precise formulation of the conditional with replies to some objections to the analysis, see Huoranszki (2011: section 4.2, section 4.3).

used in order to grasp the meaning, and to justify the ascription of, a power. We identify the power with reference to its manifestation. But, certainly, a power cannot be manifested in those circumstances in which it is lost. More interestingly, it would be a mistake to ascribe a power to someone who lacks an ability that is necessary for the occurrence of those circumstances in which the ability can become manifest.

Thus, according to the kind of conditional analysis I propose, free will is a power of the agent that may not be correctly ascribed unless some extrinsic conditions are satisfied, but the exertion of which depends on the occurrence of some internal condition: the agent choosing to perform the actually unperformed action. It is in this sense that freedom of the will explains the nature of control necessary for responsible agency. However, and most importantly, freedom of the will does not require the actual exertion of such control. Freedom of the will is understood as a generic property of agents the possession of which is necessary for their being responsible. Many times, agents are responsible for things they have omitted, but have not even considered doing. But if they are such that they would have done otherwise had they chosen to, and retained their ability to make the relevant choice and to perform the respective kind of action in the given circumstances, they enjoy freedom of the will, and they can correctly be held responsible.

7. FREE WILL AND DETERMINISM

Let us assume that this understanding of free will as the ability to act otherwise is correct, i.e. it captures the sense in which freedom of the will is a condition of responsible agency. Can anyone have the relevant ability in the circumstances of determinism? It seems to me that it is essential to distinguish two senses of determinism here: psychological determinism and physical determinism. Different versions of the conditional analysis of free will are often advanced in order to show that freedom of the will as a condition of responsibility is compatible with both.[16]

However, if responsibility requires that an agent's behavior can depend on their choices, then psychological determinism and freedom of the will do seem to me incompatible. For, as we have seen, the ability to do otherwise, in the free will sense, can correctly be ascribed to someone only in the circumstances in which she

[16] As I see it, this is particularly characteristic of those versions of the analysis that want to understand the agent's responsibility-related capacities in terms of reasons responsiveness, since even psychologically determined behavior can be reasons responsive. However, as I argued earlier in this essay, reasons responsiveness is not sufficient for responsibility, and neither is it necessary. It is only the ability to recognize reason that is indeed necessary for at least moral responsibility, but that ability is complementary, and not the same as freedom of the will.

is also able to make the relevant sort of choice. And the ability to make a choice, whatever else it is, is certainly an ability that essentially involves alternatives.[17]

There are many further questions about how we can correctly understand the relevant alternatives and under what specific circumstances we can assume that the agent has a choice. But if someone has a choice, i.e. she can exercise her ability to choose an act, then she cannot be psychologically determined to choose one option rather than the other. It is certainly true that if I had strong motives for choosing not to perform an action, then I would choose not to do it. But this is a conditional analysis of motives, if anything, and not an analysis of our ability to choose. For the possession of that ability is a condition for acting upon motives in a free and responsible way, and not just a necessary psychological condition of some conscious behavior.

Physical determinism is a different matter. Incompatibilists argue that if physical determinism is true, then we do not have the power to choose and do otherwise. Now I would not mind being an incompatibilist—actually I think I was one once—but, unfortunately, incompatibilism contradicts a view that I am strongly committed to now. It is the view that the analysis of the abilities that are necessary for our agency and responsibility must be independent of our understanding of the nature of physical laws and processes at the fundamental, and hence subpersonal, level. Independence is a kind of dualism, if you like, although what kind of dualism it is remains, of course, a further issue. However, incompatibilists must deny independence, since they believe that if physical determinism turns out to be true—a matter totally independent of considerations about our agency—then that proves without further ado that we are never responsible for what we do.

Intuitions about independence are certainly not sufficient to prove the truth of compatibilism, even if they ground my own conviction that no argument for the incompatibility of physical determinism and the ability to act otherwise can be conclusive. The most important arguments for incompatibility are the different versions of the consequence argument. Such arguments aim to prove that in the circumstances of physical determinism we cannot have the ability to do otherwise, since our actions are the logical consequences of the remote past and the physical laws. Without going into details, my general problem with such arguments is that they are either so strong that they lead to counterintuitive consequences, or they are so weak that they do not prove anything interesting about freedom of the will as the ability to do otherwise.

As to the first horn of the dilemma, if the argument aims to show that in a deterministic world we cannot have the ability or power to do otherwise, it can only

[17] See Jonathan Lowe's illuminating discussion about the nature of the ability to choose in *Personal Agency* (2008). I investigate the nature of choice in more detail in Huoranszki (2011: section 3.4).

prove that at the expense of also proving that *nothing*, no person or physical object, can have an unexercised power at all, no matter how general and how purely physical the relevant power is. For exactly the same kind of argument that is told about an agent's power to act otherwise shall apply in the context of physical powers. The argument for the impossibility of unexercised powers in physically deterministic worlds has nothing whatever to do *specifically* with whether or not the event which did not in fact occur, but which could have occurred, would have been the manifestation of an agent's relevant ability. This, for me, counts as a *reductio ad absurdum* against such kinds of arguments. Perhaps Hume is right that there is no distinction between having a power and exerting it. But not even Hume would say that whether or not there is such a distinction depends on whether or not nomic regularities constitute a logically closed deterministic universe.

Sometimes, however, when it comes to the issue of the compatibility of physical determinism with the ability to do otherwise, the notion of powers is just left behind. The incompatibilist claim is, rather, that if physical determinism is true, then the future is not open. Unlike the former versions of the argument, this version seems to me obviously sound. If the physical universe is deterministic, then there is a sense in which the future is not open, and this is exactly the sense that propositions about the physical state of the universe at one instance, together with the physical laws, entail every proposition about its future physical states just as they entail every proposition about its past physical states. However, it remains an open question whether this sense of 'not being open' is any stronger than the view that the future is not open because a true proposition about the future is, was, and remains true forever. The reason that many believe that the consequence argument is indeed stronger than the argument for fatalism is that they think that the former says something about our powers, while the latter does not. But if it does, we are back to the first horn of the dilemma: if it worked, it would prove that nothing can have an unexercised power unless physical determinism is false. But that is even harder to accept than the view that nothing can be otherwise because there are true propositions about every facts.

As I have said, the power to do otherwise, in the sense related to free will, is an extrinsic ability, and hence whether or not we can ascribe that power to an agent is sensitive to the agent's circumstances. The question of physical determinism is a question about what to include in the relevant circumstances. I do not think that there is, or could be, a general positive response to that question, since the answer depends on the particular kind of action for the performance or omission of which the agent is responsible. But I must confess that it seems to me rather counterintuitive to include the past states of the whole universe, since that would contradict our intuition that there are specific conditions under which a specific ability can be possessed or lost. If the consequence argument were correct, under the circumstances of determinism, *any* past state of the whole universe would be sufficient to deprive us of our ability to perform an actually unperformed action. I

agree that the physical universe is large and powerful, and we are tiny and fragile. But both it and we remain so independently of whether or not physical events unfold according to a deterministic system of laws.

REFERENCES

Adams, Frederick & Mele, Alfred 1992. The Intention/Volition Debate. *Canadian Journal of Philosophy* 22, 323–337.

Audi, Robert 1993. *Action, Intention, and Reason*. Ithaca, NY: Cornell University Press.

Ayer, Alfred J. 1982. Freedom and Necessity. In Gary Watson (ed.), *Free Will*. Oxford: Oxford University Press, 15–23.

Bishop, John 1989. *Natural Agency*. Cambridge UK: Cambridge University Press.

Bratman, Michael E. 2000. Reflection, Planning, and Temporally Extended Agency. *The Philosophical Review* 109, 35–61.

Clarke, Randolph 2003. *Libertarian Accounts of Free Will*. Oxford: Oxford University Press.

Davidson, Donald 1980. Freedom to Act. In his *Essays on Actions and Events*, 63–82. Oxford: Oxford University Press.

Fischer, John Martin & Ravizza, Mark 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge UK: Cambridge University Press.

Frankfurt, Harry 1971/1988. Freedom of the Will and the Concept of a Person. Reprinted in his *The Importance of What we Care About*, 11–25. Cambridge UK: Cambridge University Press.

Frankfurt, Harry 1982/1988. What We Are Morally Responsible For. Reprinted in his *The Importance of What we Care About*, 95–103. Cambridge, UK: Cambridge University Press.

Ginet, Carl 1990. *On Action*. Cambridge: Cambridge University Press.

Huoranszki, Ferenc. 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Huoranszki, Ferenc 2012. Dispositions, Powers, and Counterfactual Conditionals. *Hungarian Philosophical Review* 56, 33–53.

Lehrer, Keith 1968/1982. Cans without Ifs. In Gary Watson (ed.), *Free Will*. Oxford: Oxford University Press, 41–45.

Lewis, David 1997/1999. Finkish Dispositions. Reprinted in his *Papers in Metaphysics and Epistemology*, 133–151. Cambridge UK: Cambridge University Press.

Locke, John 1689/1975. *An Essay Concerning Human Understanding*. Ed. P.H. Nidditch. Oxford: Clarendon Press.

Lowe, Jonathan 2008. *Personal Agency*. Oxford: Oxford University Press.

Lowe, Jonathan 2010. On the Individuation of Powers. In Anna Marmodoro (ed.), *The Metaphysics of Powers. Their Grounding and their Manifestation*. Routledge, 8–26.

McKitrich, Jennifer 2010. Manifestations as Effects. In Anna Marmodoro (ed.), *The Metaphysics of Powers. Their Grounding and their Manifestation*. Routledge, 73–83.

Mellor, David H. 2000. The Semantics and Ontology of Dispositions. *Mind* 109, 551–574.

O'Connor, Timothy 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

Pettit, Philip & Smith, Michael 1996. Freedom in Belief and Desire. *Journal of Philosophy* 93, 429–449.

Ryle, Gilbert 1949. *The Concept of Mind*. London: Hutchison & Co.

Velleman, David 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Wallace, R. Jay 1998. *Responsibility and the Moral Sentiment*. Cambridge MA: Harvard University Press.

Gábor Bács

Powers, Possibilities and Ferraris

Historically, compatibilism motivated most conditional analyses of free will. Huoranszki's motivation for the conditional analysis of free will, as advocated in his *Freedom of the Will: A Conditional Analysis*, however, is not compatibilism. He thinks that incompatibilism is false anyway, regardless of the conditional analysis. He argues that the arguments for incompatibilism are not conclusive, and that the intuitions behind incompatibilism can be explained away. This he does in the very first chapter devoted to the problem of free will and determinism (*On Powers and Possibilities*), with the dialectical purpose of fending off, right at the outset, a potential objection to the conditional analysis later to be introduced.

This paper deals with the first chapter of the book. I start with a brief overview of the problem of free will and determinism and also indicate Huoranszki's compatibilist solution to it (section one). Then I turn to the first chapter of the book and present the incompatibilist arguments discussed there and Huoranszki's criticism of them (part two and three). Finally, in the last section I defend one of the incompatibilist arguments against Huoranszki (section four).

One final note of warning before we proceed. As we shall see, Huoranszki's criticisms of the incompatibilist arguments are partly based on his specific views on choice and actional abilities relevant to free will and moral responsibility discussed later in the book. It is simply beyond the scope of this paper to judge the merit of these underlying views, and thereby to judge the overall merit of Huoranszki's criticisms of the incompatibilist arguments.

1. DETERMINISM AND THE CONSEQUENCE ARGUMENT

It is customary to think that we act upon our own free will only if our action is *up to us*. According to a venerable tradition, the 'up-to-usness' of an action means that it is within our *power* to do it or to avoid doing it. So we act upon our own free will only if whatever we do we had the *ability to do otherwise*. The ability to do otherwise implies the *contingency* of actions. Agents have the ability

to act otherwise only if the occurrence of their actual action was not necessary. But the contingency of actions seems to be threatened by *determinism*, the view that the fundamental laws of nature are deterministic. A consequence of deterministic laws is that the occurrence of every event is determined by antecedent conditions and the laws of nature. Since actions are events, if determinism is true, then agents cannot avoid doing what they actually do. Hence, determinism deprives agents of their ability to do otherwise. Free will as the ability to do otherwise is *incompatible* with determinism. That is:

(1) Free will is the ability to do otherwise.
(2) Determinism is not compatible with the ability to do otherwise.
∴ (3) Determinism is not compatible with free will.

Those who accept (3) are the *incompatibilists*; those who reject (3) are the *compatibilists*. Compatibilism can take two forms, depending on whether premise (1) or (2) is rejected. Compatibilists who reject (1) typically hold that what matters for free will and responsibility is only whether agents' actions are sensitive to their *reasons*, and that to this end the truth of determinism is either irrelevant or even necessary for free will (*'only if' compatibilism*). Huoranszki, who defends (1) in the book, claiming that the ability to do otherwise captures best what free will is in the moral responsibility grounding sense, rejects (2) instead. He holds with others that we can be free and morally responsible *even if* determinism is true (*'even if' compatibilism*):

> Even if the actual physical universe is deterministic, human agents can sometimes act freely in the sense that, although they behaved in a certain way, they could have done something else instead. (Huoranszki 2011: 12.)

Huoranszki rejects (2) on the grounds that the most influential incompatibilist argument for it is not conclusive in any of its relevant versions. The argument in question is Peter van Inwagen's *Consequence Argument* who has introduced it in the following informal way:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (Van Inwagen 1983: 56.)

This is just a sketch of an argument whose details can be filled in in a number of different ways. There have been many reconstructions of the consequence argument, but Huoranszki chooses van Inwagen's own reconstructions, because he finds them the most relevant to his account of free will.

Van Inwagen (1983: 55–105) has offered three formal arguments, each differing in terminology and logical structure, of which Huoranszki considers only the first and the third. The first formal argument purports to prove that, if physical determinism is true, agents cannot have the *power to perform* any other action than what they have actually performed; the third formal argument purports to prove that, if physical determinism is true, agents cannot have the *power to choose* any other action than what they have actually chosen to perform. Huoranszki will argue that neither argument is obviously sound, because the first contains a contentious premise and the third relies on a contentious modal principle.

As for the second formal argument, which purports to prove that, if determinism is true, agents have *no access* to any possible world other than the actual world, Huoranszki thinks that it presupposes rather than establishes the truth of incompatibilism, and so does not have much independent force apart from the first and third arguments.

In what follows I shall do what Huoranszki does, and discuss the third argument before the more complex first one.

2. THE THIRD CONSEQUENCE ARGUMENT

Since the consequence argument is concerned with what agents can or cannot do if determinism is true, it must have modal content. The third argument employs modal operators to capture this, which is why van Inwagen calls it 'modal'.

The modal argument contains two modal operators: the logical necessity operator ('□'), and a special operator ('N'), where Np can be read as 'no one has, or ever had, any choice about p'. The argument also uses two inference rules for them:

α. $\Box p \therefore \mathrm{N}p$
β. $\mathrm{N}p, \mathrm{N}(p \supset q) \therefore \mathrm{N}q$

(α) is based on the modal principle that if p is logically necessary then no one has, or ever had, any choice about p. (β) is based on the modal principle that if no one has, or ever had, any choice about p and no one has, or ever had, any choice about *if p then q*, then no one has, or ever had, any choice about q (*No Choice principle*). Technically, it says that the modal operator 'no one has, or ever had, any choice about' is closed under logical implication.

The argument also contains the following abbreviated sentences: P_0 abbreviates a sentence expressing the intrinsic state of the universe in some past moment; L abbreviates a sentence expressing all laws of nature; P abbreviates any sentence about the future state of the universe. Thus equipped, the modal argument runs as follows:

(1) $\Box((P_0 \,\&\, L) \supset P)$	premise, follows from determinism
(2) $\Box(P_0 \supset (L \supset P))$	1, propositional logic
$\therefore$ (3) $N(P_0 \supset (L \supset P))$	2, α
(4) NP_0	premise
$\therefore$ (5) $N(L \supset P)$	3, 4, β
(6) NL	premise
$\therefore$ (7) NP	5, 6, β

According to the modal argument, no one has any choice about the future if determinism is true, because determinism implies that the future logically follows from the past and the laws, and no one has any choice about the past and the laws.

Huoranszki will argue that the No Choice principle (NC) is *not* obviously true, hence the inference rule (β) is not obviously valid, and so the modal argument is not obviously sound. Even van Inwagen has admitted that he had no argument for NC, but he was not concerned about this, because he found NC obvious, or at least more obvious than compatibilism. But Huoranszki thinks that NC is far from being obvious. He argues that the obviousness of NC depends on how we understand the No Choice operator (no one has, or ever had, any choice about), but that NC is not obvious under the most natural interpretations of the No Choice operator.

> The most natural interpretation of the operator when applied to the laws and the past is that we cannot *influence* them by our choices. (Huoranszki 2011: 18, my italics.)

But then it must mean the same in the conclusion: no one can influence the future by his choices if determinism is true. However, argues Huoranszki, this is equivalent to *fatalism*—the view that it is logically impossible to influence the future by our choices. Since not even incompatibilists think that determinism implies fatalism, the conclusion is surely false. Wherein lies the mistake? If we cannot influence the past, but can influence the future, then the transfer of powerlessness from over the past to over the future seems the questionable step. Hence, on this interpretation of the No Choice operator, NC is not obviously true.

Another natural interpretation of the operator in the case of the past and the laws is that no one can *make a choice* about them.

> Certainly, it seems true that there is a sense in which we cannot *make* choices about whether or not propositions expressing the past states of the universe or laws of nature are true since no one can make a choice about what she thinks is impossible for her to affect. (Huoranszki 2011: 18.)

But then the conclusion would be as follows: no one can make a choice about the future if determinism is true. And this Huoranszki finds obviously false, because we can easily find cases in which some past event logically implies some future event, yet an agent could have made a choice about it. His example is the case of a hitman who rejects the contract to kill Bill, who, unbeknownst to the hitman, died of a heart attack the day before. The hitman could not have killed Bill, but nevertheless he did make a choice about it: when he rejected the contract, he did exercise his ability to choose. So even though we cannot make choices about the past and the laws, we can make choices about the future. Thus, on this interpretation of the No Choice operator, NC is likewise not obviously true.

Huoranszki is right about this second interpretation, but, I think, for the wrong reason. If the hitman *knew* that Bill was already dead, he couldn't have made a choice about whether to kill Bill. Nor could he have made a choice if he was a fatalist and *believed* that he could not influence the future by his choices. The exertion of the ability to choose depends on beliefs—true or false. So the ability/inability to choose does not imply that one has or does not have a choice. In my view, the problem with this interpretation is that it renders the No Choice operator *epistemic*, which is inadequate for grounding a metaphysical thesis like incompatibilism.

Huoranszki is also right about the first interpretation. But there are other interpretations. I have already mentioned elsewhere (Bács 2012) Bernard Berofsky's (2002) ingenious proposal that the incompatibilist operator is *unalterability*.

> The Consequence Argument survives this assault because the operator it invokes is unalterability. I am unable to alter the future because I am unable to alter the past plus the laws. (Berofsky 2002: 195.)

It is equally natural, but does not lead to fatalism. Unlike the claim that I am unable to influence (causing to be true) the future by my choices, the claim that I am unable to alter (causing to be false) the future by my choices does not lead to fatalism. The incompatibilist can argue that I can choose to make a proposition true that I cannot choose to make false. If my choice to make a proposition true is psychologically determined by my mental states (belief and desire), then given them I could not have chosen to make it false. Huoranszki can safely

ignore this proposal only because of his views on choice ability. He holds that choices are not psychologically determined by mental states, so making a choice in the same mindset can have different outcomes.

Finally, I should mention a third interpretation of NC that Huoranszki considers briefly.

> According to that interpretation, *S* has a choice about whether or not to A only if it is *open* to *S* at time t both to A and not to A in the future. (Huoranszki 2011: 20, my italics.)

Huoranszki is dissatisfied with this abstract interpretation of NC in terms of openness of future, mainly because he finds it irrelevant. On my part I take this to be the correct interpretation of NC, but will postpone further discussion of openness and opportunity until the last section.

3. THE FIRST CONSEQUENCE ARGUMENT

The first argument does not contain modal operators; as such it is purely extensional. Instead it uses the phrase 'can render a proposition false' to capture the modal content. It refers to the ability to act over the truth-values of propositions.

The argument also contains the following propositions: P_0 expresses the intrinsic state of the universe in the remote past; L expresses the laws of nature; P expresses the intrinsic state of the universe at time *t* when subject *S* refrains from performing some action A. If we assume that no one can render a proposition about the past false, then

(1) If determinism is true, then the conjunction of P_0 and L entails P.
(2) It is not possible that *S* A-ed at *t*, and P be true.
(3) If (2) is true, then if *S* could have A-ed at *t*, *S* could have rendered P false.
(4) If *S* could have rendered P false, and if the conjunction of P_0 and L entails P, then *S* could have rendered the conjunction of P_0 and L false.
(5) If *S* could have rendered the conjunction of P_0 and L false, then *S* could have rendered L false.
(6) *S* could not have rendered L false.
∴ (7) If determinism is true, *S* could not have A-ed at *t*.

According to the argument, given that to be able to act otherwise is to be able to render false a proposition which follows from propositions about the past and the laws which no one is able to render false, it follows that, if determinism

is true, no one is able to act otherwise. In other words, determinism deprives agents of the ability to act otherwise.

Huoranszki rejects the conclusion because it implies that, if determinism is true, nothing can have unexercised powers, and this he finds implausible:

> Now my claim is that if van Inwagen's argument could prove that agents cannot have the power to act otherwise in a deterministic universe, then it would also prove that nothing whatever can have an unexercised power unless physical determinism is false. I think, however, that this is implausible. (Huoranszki 2011: 22.)

Suppose *S* owns a Ferrari, but she is also a cautious driver who respects the legally prescribed speed limit, so *S* never goes faster with her Ferrari than 130km/h. Does her Ferrari nevertheless have the *unexercised power* to go faster than 130km/h? Normally, it should. But, according to the argument, in a determinisitic universe it cannot have. If the Ferrari could have gone faster than 130km/h, it could have rendered P false. But the Ferrari could not have rendered P false, because it could not have rendered the conjunction of P_0 and L false, from which P follows. Not even a Ferrari has the supernatural power to alter the past or violate the laws. So the Ferrari cannot have the unexercised power to go faster than 130km/h in a deterministic universe. But this seems obviously false.

Since the argument is valid, but apparently not sound, one of the premises must be false. Huoranszki thinks that the questionable premise is (4). It says that I can have the ability to act otherwise only if I also have the ability to alter the past or to violate the laws of nature. But, argues Huoranszki, it does not seem to follow that because I am able to perform an actually unperformed action I must also be able to alter the past or to violate the laws. When an agent performs some action she renders a proposition true. But her ability to render a proposition true does not require the ability to render propositions about the past or the laws true. Why should then the ability to render a proposition false require the ability to render propositions about the past or the laws false? Hence (4) does not seem to be true.

The standard response is that the incompatibilist has a different sort of ability in mind when she says that determinism deprives agents of the ability to do otherwise. What she really has in mind is the *ability to exercise an unexercised ability*. Determinism deprives agents of the ability to exercise abilities, not of the abilities themselves. Whence there can be unexercised abilities. The Ferrari retains the unexercised power to go faster than 130km/h even in a deterministic universe. What is denied from the Ferrari in a deterministic universe is the power to exercise it. The Ferrari is *unable* to exercise its power to go faster than 130km/h, because *S* never drives faster than 130km/h.

Huoranszki is quick to dismiss this response on the grounds that the idea of the ability to exercise an unexercised ability not only may lead to an infinite

regress, but seems a downright logical contradiction. However his main problem with this response is that it is based on an erroneous distinction between *general abilities* and the *special ability to exercise them on particular occasions*. One of the central thesis of his book is that this distinction is wrong.

I partly agree with the dismissal. Huoranszki may be right about the regress. If the exertion of an ability requires a further ability, then so does the exertion of the ability to exercise an ability, and so on. This alone could seal the fate of the idea. But I don't think he is right about the logical contradiction. The expression 'the ability to exercise an unexercised ability' seems to involve a scope ambiguity. On a wide scope reading it implies that I could have exercised the unexercised, which is indeed a logically impossible thing to do. But on a narrow scope reading it implies that what was actually unexercised is such that I could have exercised it, which is perfectly legitimate. But this is just a minor issue.

4. IN DEFENCE OF THE FIRST CONSEQUENCE ARGUMENT

The main issue is whether the first consequence argument survives this onslaught. I think it does. In the remainder of this article I will argue on behalf of the incompatibilist and defend the first consequence argument. I will not, however, try to counter these objections. My strategy is simpler than that. I will argue that with certain assumptions the argument can be made effective against Huoranszki's own account of free will as a condition of responsibility.

Let me state first what I think the first consequence argument really purports to prove, by invoking Austin's (1979) famous distinction between the two senses of 'can'. According to Austin there are two different senses in which a person *can* do something, and it is perfectly possible that a person can do something in one sense, but not in the other. One sense involves intrinsic *abilities*; the other sense involves extrinsic *opportunities*. For instance, if you are in a locked room and you have no way to leave it, then you *cannot* leave the room in the opportunity sense of 'can'. You have the ability to leave the room, it's just that you do not have the opportunity to do so. If on the other hand you are in a room with the door open, but sadly you are paralyzed, then you *cannot* leave the room in the ability sense of 'can'. You have the opportunity to leave the room, it's just that you do not have the ability to do so.

If we recast the argument only with 'can', the conclusion would be: *S*'s Ferrari *cannot* go faster than 130km/h in a deterministic universe. But in which sense of 'can'? Obviously not in the ability sense of 'can'. *S*'s Ferrari, say, an *F458 Italia* with an 4.5L V8 engine and direct fuel injection, is internally similar to the *F458 GTE* burning the tarmac well over 130km/h in the Le Mans Series. The power to go faster than 130km/h is an intrinsic power which depends *ceteris paribus* on the internal structure of cars. Determinism does not affect the internal structure

of cars. So determinism cannot deprive *S*'s Ferrari of its *intrinsic power* to go faster than 130km/h. It must, therefore, be in the other sense of 'can' that the Ferrari *cannot* go faster than 130km/h in a deterministic universe. And I think this is exactly what the argument intends to prove. The Ferrari is deprived of the opportunity to go faster than 130km/h; it cannot have the occasion, in a deterministic universe, to show what it is capable of. So determinism deprives *S*'s Ferrari of the *extrinsic opportunity* to exercise its power to go faster than 130km/h. Huoranszki himself says that the consequence arguments aim to show the lack of abstract opportunities or the openness of the future in a deterministic universe.

> Admittedly, if physical determinism is true, there is a sense in which 'the future is not open'. The different versions of the consequence argument all aim to show exactly this. (Huoranszki 2011: 33.)

He thinks, however, that unless the incompatibilist can strike a conceptual link between the abstract opportunity to do otherwise and the agent's ability to do otherwise, the lack of opportunity has no relevance to the issue of free will which concerns our agency. As he says, we can lose and gain abilities without losing or gaining opportunities, and conversely, we can lose or gain opportunities without losing or gaining abilities.

Notice that the controversial notion of the ability to exercise abilities was one such incompatibilist attempt to link opportunity and agency. The ability to do otherwise was identified with the special ability to exercise an unexercised ability on a particular occasion, a supposedly extrinsic property that depended as much on external factors as on agents. It was the *specificity* and *extrinsicness* of the ability to exercise abilities on particular occasions that struck the link between opportunity and agency.

The ability to exercise abilities may lead to nowhere, but it does give an idea. For notice that Huoranszki takes the ability to do otherwise as a *maximally specific extrinsic property*. A central thesis of the book is that free will is a condition of responsibility. The sense in which free will is a condition of responsibility is the ability to do otherwise. We are responsible for the performance or omission of actions in particular circumstances only if we have the ability to do otherwise in those circumstances. But, according to Huoranszki, those actions must be *extrinsically identified*, so the ability to do otherwise as a condition of responsibility must be a maximally specific extrinsic property. Consequently, free will as a condition of responsibility is a maximally specific extrinsic property.

The incompatibilist might, therefore, try to forge a link between extrinsic opportunities, subject to the consequence argument, and Huoranszki's maximally specific extrinsic properties. Here is what she should do. First, she should argue that even though Huoranszki rejects the distinction between general abilities and the special ability to exercise them on particular occasions, he must still

accept a similar distinction between *general abilities* and *their maximally specific determinations*. Huoranszki takes actional abilities, relevant for being able to do otherwise in particular circumstances, to be maximally specific abilities. But maximally specific abilities presuppose more general abilities. The specific ability to speak French presupposes the more general ability to speak. The specific ability to leave a particular room presupposes the more general ability to leave rooms in general, whatever that may come to. And so on. In fact it appears that Huoranszki accepts the distinction.

> Of course, if agents have some specific ability, it does imply that they have the general one just as having the ability to see red implies the ability to see colors. (Huoranszki 2011: 85.)

General abilities will be important to evade the problem of unexercised powers. The incompatibilist could now say that even in a deterministic universe agents can have unexercised powers, in the sense that they can retain their most general powers.

Second, the incompatibilist should slightly modify the notion of opportunity. The paralyzed person who can leave the room in the opportunity sense lacks the maximally determinate ability to do so because the latter also includes the *intrinsic* ability to walk. The incompatibilist, therefore, should say that his notion of the opportunity to exercise a power must be understood such that an agent has it only if she also has the *power* to be exercised. In this sense the paralyzed person does not have the opportunity to leave the room, even if its door is open, because she lacks the ability to walk. This is not an *ad hoc* move. If someone jumps from an airplane I wouldn't say that she can fly in the opportunity sense. She also needs wings to have the opportunity to fly.

The incompatibilist can now *equate* the notion of opportunity to exercise powers with the notion of the maximally specific extrinsic determinations of those powers. She could say that the sense in which an agent has the opportunity to exercise some general power to act in certain ways in a particular circumstance is that she has the maximally determinate extrinsic ability to act in certain ways *there and then*. For instance, in the situation where I am in a locked room and I have no way to leave it, my lack of opportunity to exercise the general power to leave rooms in that particular circumstance will be equivalent to my lack of the maximally determinate extrinsic ability to leave *that* room. This in effect conforms with what Huoranszki says about the situation.

> Can I leave the room? It seems obvious that I cannot. But it may not seem obvious *why* I cannot. I would say—with Locke—that I cannot because *I'm unable* to: I lack the power or ability to leave it. (…) [If] the door is indeed locked then I do not have the ability to leave the room *there and then*. (Huoranszki 2011: 32.)

Given what Huoranszki says, the incompatibilist has all the more reason to make the identification, because it appears that the set of lack of opportunities *highly overlaps* with the set of lack of maximally determinate abilities.

Once the conceptual link is made between the opportunity to do otherwise and the maximally determinate extrinsic ability to do otherwise, the lack of opportunity will again become relevant for the issue of free will. The incompatibilist can now rerun the first consequence argument this time to argue that determinism deprives *S*'s Ferrari of its *maximally specific extrinsic determination* of the general power to go faster than 130km/h, which is the *power to go faster than 130km/h with S in its driver seat, S being a cautious driver and the speed limit being 130km/h, and so on*. The Ferrari is not deprived of its general intrinsic power to go faster than 130km/h, however. That power remains unexercised. But it is not general powers that matter for free will and responsibility. It is their maximally specific extrinsic determinations, and these can be lost in a deterministic universe according to the first consequence argument.

Finally, the incompatibilist should also say something about the objection to premise (4). Huoranszki says that if the ability to render a proposition *false* implies the further ability to render propositions about the past or the laws *false*, as (4) states, then so must the ability to render a proposition *true* imply the further ability to render propositions about the past or the laws *true*. Since the latter is obviously false, then so is the former. The incompatibilist could reply that there is an *asymmetry* between the ability to render a proposition false and the ability to render a proposition true. The latter does not require the further ability to render propositions about the past or the laws true for the simple reason that they are already *true*!

REFERENCES

Austin, John L. 1979. Ifs and Cans. In *Philosophical Papers*, 205–232. Oxford: Clarendon Press.

Bács, Gábor 2012. Huoranszki Ferenc: Freedom of the Will – A Conditional Analysis. *BUKSZ* 24, 3–4, 307–313. [In Hungarian.]

Berofsky, Bernard 2002. Ifs, Cans, and Free Will: The Issues. In Kane, R. (ed.), *The Oxford Handbook of Free Will*. Oxford: Oxford University Press, 181–201.

Huoranszki, Ferenc. 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

van Inwagen, Peter 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Anna Réz

The Puzzle of Involuntary Omissions

1. INTRODUCTION

It is a general and consensually accepted demand of theories of moral responsibility that they comply with some central and robust intuitions of ours. Common examples go back to Aristotle: we would consider a theory deeply inadequate if it could not explain why we exempt people from responsibility if they are young children or suffer from serious mental illnesses, or if they acted under compulsion or in ignorance. However, once we have accomplished this task, others will arise: most probably our theory will still yield some counterintuitive results once it comes to more complicated cases.

One such group of problematic cases includes involuntary omissions: instances of carelessness, forgetfulness, absent-mindedness, negligence and the like. Examples are numerous: we hold responsible and blameworthy the driver who caused a car accident by not paying sufficient attention; the teenager who forgot to keep her (otherwise sincere) promise to her parents; the babysitter who did not pay heed to one part of a child's dietary restrictions and consequently caused a severe allergic reaction. I assume that the central puzzle about involuntary omissions comes from the acceptance of the following three claims:

(1) We are morally responsible only for those things over which we exercise control.
(2) People do not exercise control when acting carelessly, forgetfully, absent-mindedly, etc.
(3) People are responsible for (at least some of) their involuntary omissions.

It is beyond the scope of this paper to give an analysis of (1) or to argue for its truth—for our present purposes it is enough to accept that due to its intuitive appeal it would be unreasonable to refute it without any further argument.

As we will see, (2) is probably the most often contested claim among the three. Still, it is not hard to see why we tend to think that we do not exercise control over our involuntary omissions. Although there is not any kind of general consensus about the responsibility-relevant notion of control, traditionally the concept of control is tied up with the notions of intentionality, choice and

consciousness. These features of paradigmatically responsible conduct are obviously absent in the case of involuntary omissions: not only do we not *choose* or *intend* to forget, not to notice or not to care about certain things—usually we are not even *aware* of our wrongdoing at the time of its happening. Forgetting, not keeping in mind, not noticing and not paying attention essentially involve the lack of awareness of certain facts, considerations or reasons. Whatever we happen to think about the exact conditions of control-execution, involuntary omissions will most probably fail to fulfill those criteria.

In what follows I will not discuss those accounts which aim to reconcile the tension between the three statements by denying (3) (see e.g. King 2009). Contrary to the central intuitions about responsibility I mentioned earlier, the threat posed to theories of moral responsibility by cases of involuntary omissions can legitimately be rebutted by introducing revisionism—that is, to argue that despite appearances people are not morally responsible for their careless, negligent, forgetful etc. behavior. All the same, here I will restrict my attention to those theories which do intend to account for responsibility for involuntary omissions and examine how successful they are in solving the puzzle. Ferenc Huoranszki writes: "Negligent behavior is a lot more frequent phenomena than intentional wrongdoing." (Huoranszki 2011: 47.) I couldn't agree more; exactly because involuntary wrongdoings make up the vast majority of ordinary moral transgressions, it is of utmost importance to give a feasible account of their place within the scope of responsible agency.

In the following I will first present two promising and popular solutions to the puzzle of involuntary omissions, and discuss their virtues and deficiencies. Then I will turn to Ferenc Huoranszki's treatment of involuntary omissions in *Freedom of the Will: A Conditional Analysis* and point out its advantages over rival accounts. Finally I make an attempt to answer a serious worry concerning the fairness of holding people responsible for things over which they did not exercise actual intentional control.

2. TRACING THEORIES

Indirect or *tracing* theories constitute the most sizeable and popular camp when it comes to explaining responsibility for involuntary omissions. For a clear-cut example of such theories it is worthwhile to examine Holly Smith's account of culpable ignorance (1983). Smith assumes that in all cases of culpable ignorance there is a sequence of actions: a so-called "benighting act", when the agent "fails to improve (or positively impairs) his cognitive position" (Smith 1983: 547), followed by the "unwitting wrongful act". To take her central example: the doctor who, unbeknownst to him, caused a premature infant severe eye damage because he used an unnecessarily high concentration of oxygen, is blameworthy

for blinding the infant, because at a prior time he failed to read the latest issue of the medical journal which published a study describing these effects. According to Smith, the following things must be true in order rightly to blame the agent for committing the unwitting wrongful act (blinding the infant): (1) the benighting act (not reading the journal) must be culpable, i.e., it has to be morally wrong and the agent has to be responsible for committing it, and (2) the unwitting act must fall (known to the agent) "within the risk" of her benighting act, that is, the agent must be aware that with her culpable action or omission she runs the risk of committing the latter unwitting act.

Tracing theories claim that we are indirectly responsible for an involuntary omission if and only if it is a foreseeable result or consequence of an earlier action or omission for which we are directly responsible. The core idea is that the control which we exercise when we perform free and responsible actions is transferred to some of the consequences of the action, and thus the traditional connection between responsibility and control can be reestablished. Tracing theories dissolve the tension with which we started out our discussion by refuting (2): they assume that we do (although indirectly) control our involuntary omissions.

So far, so good. It would be hard to deny that there is something obviously appealing about tracing theories. First, they preserve the connection between control, voluntariness and responsibility. And second, they reflect an important intuition of ours: that exercising control over something means, at the very minimum, that *we can do something about it*. Arguably, we would think differently about our involuntary omissions if we knew for sure that we did not have any means whatsoever to prevent them. Every time we hold someone responsible for such cases we implicitly assume that there was something the agent could have done, even if only in principle, to avoid the wrongdoing.

But tracing theories have constant and notorious problems regarding the scope of responsibility attribution. That is, it seems that indirect theories can explain only a small, if not negligible subset of those cases for which we ordinarily hold people responsible in the absence of voluntary control. To illustrate the typical shortcomings of tracing theories, take George Sher's often cited example, *Hot Dog*:

> Alessandra, a soccer mom, has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family's border collie, Bathsheba, who rides in the back of the van. The pickup has never taken long, so, although it is very hot, Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, she is greeted by a tangled tale of misbehavior, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba unconscious from heat prostration. (Sher 2006: 286–287.)

Most of us would agree that Alessandra is responsible and blameworthy for risking Bathsheba's life, although, in the traditional sense, she did not control her forgetfulness. If we are to explain Alessandra's responsibility by means of an indirect account, we have to trace Alessandra's responsibility back to a prior action or omission over which she had control.

The most fundamental problem, as Sher rightly points out, is that we do not find any suitable candidate for this role. What should have Alessandra done in order to ensure that she wouldn't leave the dog in the car? Since the row at the school was unexpected, Alessandra could see no reason to break a daily routine which had proved to be safe and comfortable for all parties.

Yet, strictly speaking there would have been countless ways to prevent her forgetfulness. For instance, if she hadn't become so deeply irritated by the headmaster's tone, it surely would have come to her mind that Sheba was in the car. However, this obviously won't do, since getting irritated is clearly not something over which we have control. The prior event from which the agent's present responsibility is derived has to be an undisputable case of controlled, responsible agency—otherwise we cannot re-establish the connection between responsibility and control.

Finally, let's say that we find such a prior action or omission. Some would say that at the end of the day Alessandra's mistake was to carry the dog with her instead of leaving her in the air-conditioned apartment. However, as I previously said, the happenings in the school were quite unexpected, so Alessandra could not reasonably foresee that by carrying Sheba she would run the risk of leaving her in the hot car for hours. But how could we hold her responsible for the consequences of her forgetfulness, if she couldn't possibly foresee that her prior voluntary actions and omissions would lead to such a terrible result?

These charges are raised quite frequently against indirect theories (see e.g. Sher 2006 & 2009, Vargas 2005) and are regularly refuted (with varying degrees of success) by the theory's representatives (see e.g. Fischer & Tognazzini's reply to Vargas 2009). Again, according to tracing theories in order to establish the agent's responsibility for her involuntary omission we need to find such a prior action or omission, where the agent's responsibility is undisputed and it was reasonably foreseeable (or, depending on the particular theory, actually foreseen) for her that this prior action or omission might result in the present wrongdoing. When we add up all these requirements, it turns out that cases of indirect responsibility are hard to come by.

3. ATTRIBUTIONISM

Attributionism is a fairly new type of theory of moral responsibility, represented primarily by Thomas Scanlon's concept of responsibility as attributability (1998) and Angela Smith's so-called rational relations view (2005, 2008, 2012).[1] According to Neil Levy's formulation, "on the attributionist account, I am responsible for my attitudes, and my acts and omissions insofar as they express my attitudes, in all cases in which my attributes express my identity as a practical agent. Attitudes are thus expressive of who I am if they belong to the class of *judgment-sensitive attitudes*" (Levy 2005).

The version of attributionism developed by Angela Smith aims to solve the puzzle of involuntary omissions by denying (1): she presents the rational relations view as an alternative to so-called volitional views, "which share a common assumption, namely, that choice, decision, or susceptibility to voluntary control is a necessary condition of responsibility" (Smith 2005: 238). Instead, according to the rational relations view,

> what makes an attitude "ours" in the sense relevant to questions of responsibility and moral assessment is not that we have voluntarily chosen it or that we have voluntary control over it, but that it reflects our own evaluative judgments or appraisals. (Smith 2005: 237.)

Smith's key example is the following: I have forgotten my friend's birthday. Most would say that I am responsible and blameworthy for my forgetfulness. But how can we explain and justify this judgment of responsibility? Smith presents the problem in the following fashion:

> But what, exactly, was the nature of my fault in this case? After all, I did not consciously choose to forget this special day or deliberately decide to ignore it. I did not intend to hurt my friend's feelings or even foresee that my conduct would have this effect. I just forgot. It didn't occur to me. I failed to notice. And yet, despite the apparent involuntariness of this failure, there was no doubt in either of our minds that I was, indeed, responsible for it. (Smith 2005: 236.)

According to the rational relations view, the things which occur to us and those which we completely neglect—our general sensitivity or insensitivity to certain aspects of our environment—can be the proper subject of moral assessment because they express our evaluative judgments about the weight and importance of these things. As Smith summarizes it:

[1] Also, Gary Watson (1996) and Pamela Hieronymi (2006, 2008) are often considered as attributionists.

> if one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one's tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person. If this is so, then the fact that a person fails to take note of such factors in certain circumstances is at least some indication that she does not accept this evaluative judgment. (Smith 2005: 244.)

Smith's argument goes as follows: if one holds the evaluative judgment that *x* is important, then one will be disposed to notice factors relevant to *x*'s welfare. Thus, by using contraposition, we can conclude that, if someone is not disposed to notice relevant factors to *x*'s welfare, she does not hold the evaluative judgment that *x* is important. This lack of judgment is why we hold her responsible.

The attributionist solution to the problem of involuntary omissions has some major advantages. First, as we have seen in the discussion of tracing theories, authors have constant difficulties with establishing the connection between responsibility and voluntary control in such cases. By simply denying the relevance of any such connection, attributionist accounts relieve the discussion of this burden and offer a relatively easy solution to the puzzle. Second, attributionist accounts do an excellent job in identifying the *content* of moral criticism. Indeed, it seems to be the case that I am blameworthy because I do not care enough about my friend, not because of any conscious or voluntary action or omission of mine. It is the lack of concern which triggers moral criticism.

However, it was exactly the idea of rational inferences (the very heart of the rational relations view) from the agent's attitudes or conduct to her evaluative judgments that was recently to come under fire. Matt King (2009) argues that such inferences are usually based on repeated evidences, while blameworthiness for behaving negligently does not presuppose any such regularities:

> The power of the evidential relation surely rests on the reliability of the inference from conduct to ill qualities of will. The reliability of such an inference requires, it seems, some regularity in its connections. (…) Of course, any conduct can count as some evidence for the underlying quality of will, but we generally require more before we are justified in actually drawing the inference. (…) But ascriptions of responsibility in cases of negligence need not rest on regularities. (…) [O]ne transgression is sufficient for negligence, and if negligence itself is to be sufficient for responsibility, then it seems that quality of will views (on the evidential reading) fare no better in explaining it, for the transgression itself won't be sufficient evidence for an ill quality of will. (King 2009: 584.)

Holly Smith echoes King when she writes:

> In such cases, no *stable* faulty attitude could be attributed to the agent in light of his or her one-time failure to take notice. Indeed it may not even be plausible to ascribe to the agent a *momentary* faulty attitude of the kind shown by an exhausted soldier who shoots at a movement in the house he is searching, too tired to care about the risk that he is shooting an innocent civilian rather than an enemy combatant.
>
> In cases such as these, in which we can't reasonably impute a faulty evaluative attitude to the negligent agent, the attributionist strategy for imputing blame to the agent for her culpably ignorant act seems to fail. (Smith 2011: 120.)

The problem which King and Smith highlight is especially apparent in our previous example of Alessandra and Sheba. There is no hint in the story which would suggest that Alessandra's failure to notice that Sheba is in danger would be due to any kind of disregard or general carelessness toward the dog. But we can equally imagine an alternative scenario in which Alessandra's failure to take notice of Sheba is in perfect harmony with her views about the relative unimportance of animal life or her low opinion of Sheba in particular. Either way, Alessandra's behavior remains an instance of both responsible and blameworthy conduct. Involuntary omissions such as forgetfulness and negligence, *pace* Smith, can come about with or without evaluative commitments being manifested in them—thus we cannot discriminate between responsible and non-responsible agency on this basis.

4. CONDITIONAL ANALYSIS

For those who have read the introductory article of this volume it is not necessary to give a detailed overview of Ferenc Huoranszki's account of free will, understood as a (the crucial?) condition of moral responsibility for our actions and omissions (as opposed to responsibility for consequences and mental states, of which Huoranszki's theory does not aim to give an account). Thus I will restrict my attention to the central claims of *Freedom of the Will: a Conditional Analysis*, without going into the details of the metaphysical and action theoretical discussions which underlie them.

Following classical compatibilist authors (primarily John Locke and G. E. Moore), Huoranszki develops a conditional account of free will, according to which:

> *S*'s will is free in the sense of having the ability to perform an actually unperformed action A at *t* iff *S* would have done A, if (1) *S* had chosen so and (2) had not changed with respect to her ability to perform A at *t* and (3) had not changed with respect to her ability to make a choice about whether or not to perform A at *t*. (Huoranszki 2011: 66.)

In Chapter 3, Huoranszki discusses at length the case of involuntary omissions and their connection to intentional control and the ability to choose and act otherwise. In his argument, he takes it for granted that people are often morally responsible for their involuntary omissions: his examples involve cases of carelessness, forgetfulness and negligence. Thus he accepts (3), which leaves him with two options: to deny either (1) or (2). Huoranszki opts for the first strategy: while he heavily criticizes tracing theories (on slightly different grounds than I do), he maintains that agents can possess freedom of the will without exercising actual intentional control. Also, although he admits that it is a legitimate move to choose intentional control over freedom of the will as the condition which grounds moral responsibility, he argues that it is the latter which we should consider as the relevant criterion—that is, contrary to (1), control is *not* a necessary condition of responsibility.

It would be a mistake, however, to conclude that the conditional analysis offered by Huoranszki completely dismisses the significance of control in establishing responsibility. The core idea of the conditional analysis is that freedom of the will is an ability, which the agent did or did not possess at the time of her action or omission: the ability to choose and act otherwise. That is, although in most cases of involuntary omissions the agent did not exercise (neither direct nor indirect) intentional control over her omission, she still retained her ability to avoid her omission, had she chosen to do so (and she had the ability to so choose). She did not exercise control—but she could have exercised it.

I find this analysis compelling and illuminating. The intuitive reason why we can attribute responsibility to people for their involuntary omissions, despite the unconscious and unintended nature of their conduct, is that those abilities which would allow them to prevent the omission seem unimpaired. Neither some general incompetence of the agent, nor momentary epistemic or physical obstacles can explain the omission, which suggests—in accordance with the conditional analysis—that the omission comes about because the agent *fails to exercise* her (otherwise intact) abilities.

Huoranszki's account of involuntary omissions has major advantages over tracing and attributionist theories. On the one hand, while preserving the link between responsibility and control execution, contrary to tracing theories it does not restrict drastically the scope of responsibility. We often retain our ability to choose and act otherwise, despite the fact that we do not actually exercise it, neither prior to nor at the time of our omission; whereas tracing theories would,

contrary to our ongoing moral practices, exclude these cases from responsible agency, according to the conditional analysis in these cases we do possess freedom of the will and consequently are potentially appropriate targets of responsibility attribution.

On the other hand, although having free will means having the ability to choose and act otherwise, this ability is highly specific and can be identified only by referring to such external features which obtain at the time of the action. Thus, contrary to attributionism, according to conditional analysis the agent's responsibility does not stand or fall on such fairly long-standing mental states (evaluative judgments), which might well be absent without thereby undermining either responsibility or blameworthiness.

5. CONTROL AND FAIRNESS

Despite the apparent advantages of conditional analysis, proponents of tracing theories might still insist that it is unfair to hold someone responsible for her involuntary omission, given that it wasn't a foreseeable consequence of her intentionally controlled conduct. They may ask: what is the moral significance of an unexercised ability of choice? Cases of carelessness, forgetfulness or negligence arise exactly because certain reasons, facts and considerations do not even cross the agent's mind. But how could we fairly hold anyone responsible for something over which they did not have any kind of conscious control?

There is something definitely compelling in this line of thought, and it is doubtful whether it could be silenced for good by any argument. Here I only offer another consideration which can help to strengthen the position of Huoranszki's account.

When supporters of tracing theories hold to the condition of actual direct or indirect intentional control, they rely on an attractive moral principle, which we often implicitly assume while talking about the conditions of responsibility—i.e., the principle that it would be unfair to hold someone responsible for things beyond their control. Since, as George Sher astutely points out, this principle "is more often baldly asserted than carefully defended" (Sher 2005: 180), its exact content and the concepts involved are rarely discussed. Probably the most pressing issue is what we mean by something being "beyond our control". Tracing theories interpret the principle as stating that it is unfair to hold someone responsible for things over which she did not exercise actual (direct or indirect) control.

However, our ordinary way of talking about control suggests that this is not the most natural interpretation of the principle. This becomes especially apparent when—as we often do—we talk about losing or not having control over something. For the sake of simplicity I will here illustrate my point with an ex-

ample in which we talk about controlling an external object instead of our own behavior.

Let's suppose that a perfectly competent driver causes a car accident which could easily have been avoided. When the police inquire as to what happened, the driver claims that she lost control over the car. Naturally, the police ask for further details: was the brake broken or the steering wheel disabled? Did she have a seizure, making her unable to handle the situation? Her reply is this: "No, everything functioned perfectly. It's just that I did not press the brake. It did not cross my mind." Not only is this a ridiculous excuse, it is also a deeply puzzling one: we did not get any adequate explanation of why the accident occurred. But whatever explains the accident, it won't be the lack of control: although we might say that the driver did not control her car (although this claim might also sound somewhat odd), it is indisputable that she *could have controlled* it—and this is the fundamental question which we aim to settle when talking about someone retaining or losing control over something.

Obviously, this case is not analogous to our previous example *Hot Dog*. Alessandra's failure to rescue Sheba is not mysterious at all: her forgetfulness is adequately explained by the school row which distracted her attention and thus broke the usual course of events. Also, it is an intriguing issue how serious a distraction should be to exempt the agent from responsibility. If upon arrival at the school Alessandra had been informed that her son was in the intensive care unit of the local hospital, we would excuse her forgetfulness. In a similar vein, if the driver's failure to push the brake were explained by his catching sight of some brutal and violent crime taking place on the street, we would consider letting her off the hook. However, the existence of these apparently exempting conditions further strengthens Huoranszki's position, since arguably what we consider in both cases is whether the distraction was large enough to deprive the agent of her ability to choose (and consequently act) otherwise.[2]

The car driver example suggests that—contrary to what tracing theories assume—the principle that we cannot fairly be held responsible for things beyond our control should be interpreted, faithfully to the ordinary usage of the terms involved, as a claim concerning our *ability* to exercise control instead of the presence or absence of *actual* intentional control. Once we have replied to the tracing theorist's worry about unfairness by showing it to be unwarranted, we removed a major obstacle which rendered it more difficult to endorse the conditional analysis of involuntary omissions.

[2] I'm grateful to my anonymous referee for pushing me to clarify these points.

6. CONCLUSION

Cases of involuntary omissions such as absentmindedness or negligence pose a major challenge to any theory of moral responsibility which aims to explain how an agent can be responsible for an omission despite its taking place without their conscious or intentional activity. In this paper, I first presented two popular solutions to this puzzle and presented their most serious drawbacks. Then I argued that the conditional analysis of free will offered by Ferenc Huoranszki proves to be more successful and intuitively more illuminating than its rivals. Although the exact details of the theory are in many respects crucial, I suspect that any theory of responsibility which shares Huoranszki's emphasis on the *ability* to exercise intentional control, and therefore which analyzes involuntary omissions in terms of the agent's *failing* to exercise this ability, can do an equally good job in explaining and justifying responsibility attribution for these common instances of human agency. Without this conceptual framework, however, the puzzle of involuntary omissions might well remain unresolved.

REFERENCES

Fischer, John M. & Tognazzini, Neil 2009. The Truth about Tracing. *Noûs* 43, 531–556.
Hieronymi, Pamela 2006. Controlling Attitudes. *Pacific Philosophical Quarterly* 87, 45–74.
Hieronymi, Pamela 2008. Responsibility for Believing. *Synthese* 161, 357–373.
Huoranszki, Ferenc 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.
King, Matt 2009. The Problem with Negligence. *Social Theory and Practice* 35, 577–595.
Levy, Neil 2005. 2005. The Good, the Bad and the Blameworthy. *Journal of Ethics & Social Philosophy* 1, no. 2.
Scanlon, Thomas 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
Sher, George 2005. Kantian Fairness. *Philosophical Issues* 15, 179–192.
Sher, George 2006. Out of Control. *Ethics* 116, 285–301.
Sher, George 2009. *Who Knew?* New York: Oxford University Press.
Smith, Angela M. 2005. Responsibility for Attitudes. *Ethics* 115, 236–271.
Smith, Angela M. 2008. Control, responsibility, and moral assessment. *Philosophical Studies* 138, 367–392.
Smith, Angela M. 2012. Attributability, Answerability, and Accountability: In Defense of a Unified Account. *Ethics* 122, 575–589.
Smith, Holly M. 1983. Culpable Ignorance. *Philosophical Review* 92, 543–571.
Smith, Holly M. 2011. Non-Tracing Cases of Culpable Ignorance. *Criminal Law and Philosophy* 5, 115–146.
Vargas, Manuel 2005. The Trouble with Tracing. *Midwest Studies in Philosophy* 29, 269–291.
Watson, Gary 1996. Two Faces of Responsibility. *Philosophical Topics* 24, 227–248. Reprinted in *Agency and Answerability*, 260–288. Oxford: Clarendon Press, 2004.

Zsófia Zvolenszky

Conditionals, Dispositions, and Free Will

1. INTRODUCTION

In Chapter 4 of his *Freedom of the Will: A Conditional Analysis*, Ferenc Huoranszki offers the following account advertised in the title of the book:

> Huoranszki's conditional analysis of free will
> *S*'s will is free wrt an unperformed action *A* iff
> S would have done *A*, if
> (i) *S* had chosen to perform *A*, and
> (ii) there is no change in *S*'s ability to perform *A*, and
> (iii) there is no change in *S*'s ability to make a choice about whether to perform *A*.[1]

Conditional accounts of free will—like G. E. Moore's, Thomas Hobbes' and David Hume's—echo a variant of clause (i) of this analysis. Why include the other two clauses—in particular, clause (ii)? It is here that a debate in metaphysics and the philosophy of language, about dispositions, will prove relevant and instructive, providing interesting parallels, conceptual clarifications, and also additional indirect support for Huoranszki's conditional account of free will.

About dispositions in a nutshell. Picture a slice of perishable chocolate cake: it has the disposition to spoil if left outside the refrigerator for a prolonged period. But as things stand, the slice never spoils, for I gobble it up as soon as it's purchased; still, while in existence, the slice retained this ever unactualized disposition of perishability. A highly influential account—defended, for example, by Gilbert Ryle, Nelson Goodman and W. V. Quine—has it that ascriptions of dispositions like perishability, water-solubility, fragility, and so on should be given a conditional analysis along the following lines:

[1] I use the standard abbreviations: 'wrt' stands for 'with respect to'; 'iff' for the biconditional connective 'if and only if'. For simplicity, the temporal qualification '*S*'s will is free at time *t*' is left implicit throughout in this as well as other definitions.

SIMPLE conditional account of disposition ascriptions
An object/person/substance *N* is disposed to *M* under *C* iff
N would *M* if
(I) it were the case that *C*.

That is, the slice of cake, while in existence, had the disposition to spoil under the condition of being left outside the fridge for a prolonged period just in case it would have spoiled had the condition obtained.

Notice that clause (i) of HUORANSZKI and clause (I) of SIMPLE are extremely similar: both provide an analysis in terms of a subjunctive conditional (past- or present-tense, respectively) of the form '*P* would have been the case, if it were the case that *C*' and '*P* would be the case if it were the case that *C*'.[2] In Section 4, we'll see that the similarities run deeper than that: the difficulties emerging in the context of giving a conditional account of dispositions point the way toward further reasons to include a clause parallel to HUORANSZKI's (ii) in the conditional analysis of disposition ascriptions. By way of stage setting, in Section 2, I will trace some of the reasons why Huoranszki departs from Moore's classic version of the conditional analysis of free will. In Section 3, I will give some preliminaries on conditional analyses of disposition ascriptions. Concluding remarks will follow in Section 5. Along the way, my aim is also to reconsider the role and interrelations of the various counterexamples to the sufficiency and necessity of the conditional analysis of disposition ascriptions; these constitute crucial clarifications not only for the dispositions debate but also for Huoranszki's arguments for his version of the conditional analysis of free will.

2. CONDITIONALS AND FREE WILL

In developing his own conditional analysis of free will, Huoranszki (2011: section 4.1) takes as his starting point Moore's classic proposal (1912: 220–221), reconstructing it along the following lines:

MOORE's conditional analysis of free will
S's will is free wrt an unperformed action *A* iff
(i′) *S* would have done *A*, if *S* had chosen to perform *A*,
(ii′) *S* could have chosen to make a choice about performing *A*,
(iii′) no-one can predict whether or not *S* chooses to perform *A*.

[2] Throughout, I am assuming in the background Lewis's (1973) possible worlds semantics for subjunctive conditionals, according to which, roughly, a conditional of the form *P* would be/would have been the case, if it were/had been the case that *C*' is true iff all *C*-worlds (world in which *C* is true) most similar to the actual world are *P*-worlds (Lewis 1973).

According to MOORE, (i')–(iii') are individually necessary and jointly sufficient for *S* to have free will with respect to performing an action. Of these, Huoranszki keeps (i'), finding fault with (ii') and (iii'). Clause (ii'), he argues, leads to an infinite regress and also fails to be a necessary condition for having free will (a point argued for in Chapter 3 of his book). Clause (iii'), he argues, is neither necessary nor sufficient for having free will to perform an action: Huoranszki is free to refuse a bowl of zucchini soup even though those who know him well can easily predict his refusal; meanwhile, the unpredictability of events (like various weather phenomena) gives no guarantee whatsoever that they are free to occur.

Huoranszki (2011: section 4.2) offers his own clauses (ii) and (iii), above, to provide what he considers an adequate conditional analysis of free will.[3] His clause (i) is the same as (i') in MOORE—we'll henceforth refer to it simply as clause (i).[4] But he thinks (i) by itself would be insufficient to define free will because of an objection of Keith Lehrer's (1968/1982): it is logically possible that the very act of choosing or not choosing *A* affects one's ability to perform the action in question. HUORANSZKI's clause (ii) is intended to block such a possibility. (HUORANSZKI's clause (iii) is intended to block yet another counterexample of Lehrer's which we won't discuss here.)

Let's spell out Lehrer's objection to clause (i) in four steps:

First step. It is logically possible that (1), (2) and (3) hold for some action *a*.

(1) *s* can perform *a* (in the sense of having free will wrt to *a*) only if *s* chooses to perform *a*.
(2) *s* doesn't choose to perform *a*.
(3) *s* would have done *a* if *s* had chosen to perform *a*.

Second step. From (1) and (2) the *modus tollens* inference schema, below, yields (4):

Modus tollens:
if *P* then *Q* (which is equivalent to *P* only if *Q*)
not *Q*
Therefore, not *P*

[3] Huoranszki's broader aim is to use the conditional analysis to expose a problem with Peter van Inwagen's Consequence Argument (discussed in Huoranszki's Chapter 2), according to which from determinism it follows that our actions are not up to us.

[4] For ease of exposition, whenever it's harmless in the context of the paper, I'll be deliberately "sloppy" in glossing over discrepancies like the following: (i') includes the conditional consequent while (i) doesn't.

(4) *s* cannot perform *a* (in the sense of having free will wrt to *a*).

Third step. (3) states that the action *a* satisfies clause (i).

Fourth step. An action like *a* shows clause (i) to be insufficient to define free will, for *a* satisfies (i) (according to (3) above), yet the subject in question does not have free will wrt to *a* (according to (4) above).

HUORANSZKI's (ii) serves to exclude actions like *a*, for which (1) holds. We have so far left as abstract what an action fitting *a*'s parameters would be. Lehrer offers a far-fetched example:

> Suppose that, unknown to myself, a small object has been implanted in my brain, and that when a button is pushed by a demonic being who implanted this object, I became temporarily paralyzed and unable to act. My not choosing to perform an act might cause the button to be pushed and thereby render me unable to act. (Lehrer 1968/1982: 44.)

In the context of dispositions (in Section 4), we'll encounter some more realistic examples that are analogous to this one, along with close parallels between the conditional analyses of free will and of disposition ascriptions.

We should note already that in his conditional analysis, Huoranszki proposes to construe freedom of the will as a special ability/unactualized power: the ability/power to act otherwise. Moreover, this very ability/power is featured in clause (ii), making the analysis nonreductive (a point that we will revisit in the last section).

How do dispositions come into this picture? There are various ways we might understand disposition ascriptions like the following: "Huoranszki is disposed to ride a bike"; "This piece of cake is perishable". We might take the first to mean, on the one hand, that Huoranszki is inclined/prone/has a tendency to ride a bike, or, on the other hand, that he has a power/ability to do so. Huoranszki is interested in this latter sense of 'is disposed to…' and other dispositional predicates ('is perishable/fragile/edible/lethal' etc.). Notice that this is a natural move, given that many claims about dispositions don't involve habit or recurrence: a cake's edibility does not mean it can be eaten more than once, a cup's fragility does not mean it can break more than once, and a poison capsule's being lethal doesn't mean it can kill more than once. We are thus looking at analyzing dispositional predicates in the sense of "a substance's, an object's, or a person's *power to behave in certain ways* in certain kind of circumstances, even if they *never* behave that way" (Huoranszki 2011: 60; emphases in the original). This is the sense of 'is disposed to' that we aim to capture via a conditional analysis.

If we construe dispositional predicates as referring to abilities/powers, and freedom of the will as the ability/power to act otherwise, then—unless there are reasons warranting special treatment for the latter—MOORE is a specific application of a general, conditional-based account of disposition ascriptions. In this case, an analog of Lehrer's counterexample to clause (i) should also arise for a conditional account of disposition ascriptions. According to Huoranszki (2011: 61–63), the conditional account of dispositions is indeed subject to a Lehrer-analog counterexample, which serves to point us in the right direction about how clause (i) should be supplemented: by including clause (ii) of HUORANSZKI.

Let's see how various counterexamples, the Lehrer-analog one included, have shaped the discussion about conditional analyses of disposition ascriptions. Setting right the roles and interconnections of the various counterexamples does, I think, shed light on the dispositions debate and carries ramifications for Huoranszki's line of argument for his own account of free will.

3. CONDITIONALS AND DISPOSITIONS: MASKS AND MIMICKS

It is widely assumed that all dispositional predicates can be understood as involving a condition of manifestation: for example, 'is perishable' is about being disposed to spoil under a certain condition, say, being left out of the fridge for an extended period. This way, all disposition ascriptions fall under the SIMPLE conditional analysis of dispositions, repeated here:

> SIMPLE conditional account of disposition ascriptions
> An object/person/substance *N* is disposed to *M* under *C* iff
> *N* would *M* if
> (I) it were the case that *C*.

A battery of counterexamples challenge this analysis from two directions.

Some counterexamples call into question whether the analysis provides *necessary* conditions: the first half of the biconditional in SIMPLE might be true while the second is false, showing that the truth of the second half is not necessary for the truth of the first half; call these anti-necessity T–F counterexamples. For a porcelain cup to be fragile—to be disposed to break when dropped, say—it is not necessary that the cup would break if dropped. This is shown by *masking cases*: if the fragile cup has suitable protective packaging, it remains fragile, yet it would not break if it were dropped and all other circumstances remained maximally similar to actuality (including the packaging). The packaging is an extrinsic feature that *masks* the cup's disposition to break (Johnston 1992: 233; see also Bird 1998).

Some counterexamples indicate that SIMPLE fails to provide *sufficient* conditions: the first half of the biconditional in SIMPLE might be false while the second half is true, showing that the truth of the second half is not sufficient for the truth of the first half; call these anti-sufficiency F–T counterexamples. A non-fragile object might be such that it would break when dropped. This is shown by *mimicking cases* (discussed also by A. D. Smith 1977: 441, 444):

> A gold chalice is not fragile but an angel has taken a dislike to it because its garishness borders on sacrilege and so has decided to shatter it when it is dropped. Even though the gold chalice would shatter when dropped, this does not make it fragile because something extrinsic to the chalice is the cause of the breaking. (Johnston 1992: 232.)

> When a styrofoam dish is struck, it makes a distinctive sound. When the Hater of Styrofoam hears this sound, he comes and tears the dish apart by brute force. So, when the Hater is within earshot, styrofoam dishes are disposed to end up broken if struck. (Lewis 1997: 153.)

Because of the angel, an extrinsic factor, the chalice *mimicks* the behavior of a fragile object without being fragile. Because of the Styrofoam Hater, an extrinsic factor, the styrofoam dish mimicks the behavior of an object disposed to break when struck, without having that disposition.

David Lewis (1997) appealed to intrinsic properties in his influential reformed analysis, which is thought to handle some of the counterexamples against SIMPLE. His justification was that plausibly, "dispositions are an intrinsic matter" (Lewis 1997: 147)—their causal bases are properties that are intrinsic to the object (Lewis 1997: 155); for example, a porcelain cup's fragility is due to its intrinsic properties, as is a gold chalice's non-fragility. (Lewis argues that without the intrinsicness restriction on properties, we would run into the problem that the chalice itself has the disposition to break.)[5] Below I have simplified Lewis's proposal along the lines of Sungho Choi – Michael Fara (2012: section 1.4):

[5] Lewis (1997: 155) offers an ordinary example of his own (an analogous line can be made about the porcelain cup with protective packaging):

> … to placate those who will not be convinced by fantastic examples, I offer the case of Willie. Willie is a dangerous man to mess with. Why so? Willie is a weakling and a pacifist. But Willie has a big brother—a very big brother—who is neither a weakling nor a pacifist. Willie has the extrinsic property of being protected by such a brother; and it is Willie's having the extrinsic property that would cause anyone who messed about with Willie to come to grief. If we allowed extrinsic properties to serve as causal bases of dispositions, we would have to say that Willie's *own* disposition makes him a dangerous man to mess about with. But we very much do not want to say that. We want to say instead that the disposition that protects Willie is a disposition of Willie's brother. And the reason why is that the disposition is an intrinsic property of Willie's brother. (Emphasis in the original.)

INTRINSIC-property-based conditional account of disposition ascriptions
An object/person/substance *N* is disposed to *M* under *C* iff
there is an intrinsic property *B* that *N* has such that *C* and *B* would jointly cause *N* to *M*, if
(I) it were the case that *C*, and
(II) *N* were to retain *B* for a sufficient time.

According to Michael Fara (2009), INTRINSIC …

> … avoids the problem of "mimicking"… It is true that the gold chalice, watched over by the destructive angel, would shatter if it were dropped. But the chalice has no intrinsic property which would contribute to causing the shattering—the angel alone would cause the chalice to shatter. (Fara 2009: section 2.3.)

Fara is suggesting that the chalice no longer presents an F–T counterexample given the amendments in INTRINSIC, because the second half of the biconditional in INTRINSIC comes out false (as does the first half): no intrinsic property of the chalice contributes to causing the shattering.[6]

We might, however, think it is unclear that INTRINSIC avoids the chalice counterexample. For we might reason: the chalice does have an intrinsic property *p* giving rise to the chalice's extreme garishness, and *p* causes the angel's wrath, in turn causing the shattering of the chalice; and with *p*, we can make the second half of the biconditional true, bringing back the original F–T problem the chalice had presented for SIMPLE. I find this objection to Fara compelling; it draws support from two considerations.

First, just one paragraph later, Fara points out that the T–F masking counterexample with the packaged porcelain cup remains unresolved by INTRINSIC, for the biconditional's first half remains true (the cup is still fragile/disposed to break when dropped) while the second half is still false:

> the cup does have an intrinsic property which would join with the dropping in causing it to shatter *if the packing were absent*. But since the packing *isn't* absent (and wouldn't be absent if the cup were dropped), the [second half of the biconditional], in this instance, is false. (Fara 2009: section 2.3; emphases in the original.)

[6] Fara's is a far more plausible take on Lewis's example of Willie protected by the brother who is dangerous to mess with (in the previous footnote): the second side of INTRINSIC comes out false the same way as the first if we substitute "Willie is disposed to be dangerous under the condition of being messed with"; for there is arguably no *intrinsic* property of Willie's to fit the analysis. Willie's case, formerly a counterexample to SIMPLE, is no longer such. (But even if we were to accept this, the chalice case would still remain a counterexample to INTRINSIC.)

But if an extrinsic factor like the packaging material remains in place when evaluating the subjunctive conditional (because the conditional's antecedent instructs us to consider worlds in which there is no departure from the actual world except for the cup being dropped, so all worlds under consideration retain the packaging), making the biconditional's second half false, then in the chalice example, we likewise have no grounds for excluding the vengeful angel's presence from the worlds under consideration (angelless worlds would constitute a gratuitous departure from actuality), and, in all such worlds, the chalice does break (thanks to the angel), so the second half of INTRINSIC comes out true, contrary to Fara's point that it is clearly false. The chalice example does, after all, remain an F–T anti-sufficiency counterexample.

Second, the following quote from a substantially revised version of Fara (2009) discusses the second mimicking case about the styrofoam dish, reversing Fara's previous verdict: the new version claims that the styrofoam example (along with other mimicking cases) remains unresolved by INTRINSIC:

> [INTRINSIC] doesn't avoid the problem of mimickers. The styrofoam dish, if struck, would break by the mimicking operation of the Hater of Styrofoam. Note that, if struck, the dish would retain for a sufficient time an intrinsic property, say, the microstructure responsible for its distinctive sound and, further, this intrinsic property would be a cause of the breaking. The prediction by [INTRINSIC] is therefore that the styrofoam dish is disposed to break when struck, which might be claimed to be counterintuitive. (Choi–Fara 2012: section 1.4.)

The reason why Choi and Fara think the styrofoam example remains a problem for INTRINSIC parallels exactly the objection I had formulated about the chalice: given the presence of the Hater of Styrofoam, there is some intrinsic property of the styrofoam dish giving rise to the sound that the Hater loathes and causing the dish to break (at the hands of the Hater), say Choi and Fara; given the presence of the vengeful angel, there is some intrinsic property of the chalice that inspires the angel's wrath, causing the chalice to break (given the angel's intervention), say I.

So far, compared to SIMPLE, INTRINSIC hasn't made any headway on the counterexamples (coming up in Section 4); before moving on to Lehrer-analog counterexamples with which INTRINSIC does help, let us take a closer look at the counterexamples now on the table.

Just how common are masking and mimicking cases? Certainly, the mimicking examples about angelic wrath over a chalice and about the Styrofoam Hater both seem rather exotic. Fara (2005: 76, 81) mentions mimicking cases in passing only, but stresses that masking cases need not be extraordinary:

> It is worth noticing that masking is a commonplace phenomenon: dispositions of objects are being masked all the time. I'm disposed to go to sleep when I'm tired; but this disposition is sometimes masked by too much street noise. Cylinders of rubber are disposed to roll when placed on an inclined plane; but this disposition can be masked by applying a car's break. A piece of wood in a vacuum chamber is no less disposed to burn when heated than is its aerated counterpart (if dispositions are intrinsic properties, then this has to be granted); but wood won't burn when heated in a vacuum. The masking of dispositions is such a humdrum occurrence that any adequate account of disposition ascriptions must accommodate it. (Fara 2005: 50.)

Notice a key conceptual connection between the mimicking and masking cases discussed above: they are two sides of the same coin in that the fragile cup with protective packaging has its fragility masked, meanwhile, because of the packaging, the cup mimicks non-fragility. And because of the angel's wrath, the gold chalice mimicks fragility, meanwhile, its non-fragile nature is masked by the angel's wrath. Mimicking cases, then, are no more exotic than masking cases. For numerous scenarios in which a disposition is masked, a corresponding opposite disposition is being mimicked and *vice versa*: toxicity/malleability masked amounts to non-toxicity/non-malleability mimicked and *vice versa*; toxicity/malleability mimicked amounts to non-toxicity/non-malleability masked and *vice versa*. So if masking cases are commonplace, as Fara suggests, then so too are mimicking cases.

Not all possible examples work this way. Consider a cup made of melamine resin: a kind of hard plastic that is considerably less breakable than glass or porcelain, but which is still somewhat breakable. Suppose that a melamine cup is such that it would survive some (regular, household-style) dropping events but not all of them. Then the melamine cup is neither fragile nor non-fragile: it is both false that if it were dropped it would break and that if it were dropped it wouldn't break. The cup is *in-between with respect to fragility*. Now imagine a scenario in which our vengeful angel specializes in the melamine cup, making sure it breaks when dropped. This is a case of fragility mimicked without non-fragility being masked. Imagine another scenario in which partial protective packaging is put on a porcelain cup, so it doesn't become non-fragile but about as sturdy as a melamine cup. Then the porcelain cup's fragility has been masked without non-fragility being mimicked.

(Another example of in-between status: suppose Huoranszki owns a trekking bike, which he rides around the city, and a road bike for countryside outings. It is not the case that Huoranszki is disposed to take his trekking bike when riding a bicycle; nor is it the case that he is disposed to not take his trekking bike when riding a bicycle. For this disposition–opposite disposition pair, he has neither. Huoranszki is *in-between with respect to taking his trekking bike when he goes for a bicycle ride*.)

These sorts of examples do not undermine the point I have been making, for two reasons. On the one hand, the examples discussed in the dispositions debates aren't like the in-between cases. Indeed, to get a really compelling masking case, it helps to go all the way: mask fragility in such a way that the object *never* breaks. And the same goes for mimicking cases: a truly striking example has an extremely sturdy, clearly non-fragile object (like the gold chalice) that mimics fragility due to an extrinsic factor like the vengeful angel. On the other hand, there are still plenty of pairs of dispositions and their opposites for which the conceptual connection I have been describing is in place and that suffices for my point that masking and mimicking cases do, to a large extent, overlap. To underscore this point about how common it is that masking cases are at once mimicking cases of the opposite disposition and *vice versa*, consider that more specific conditions of manifestation alter in-between status: our melamine cup is disposed to break when dropped from a sufficient height, say 20 meters, and it is not disposed to stay intact when dropped from 20 meters. And Huoranszki is disposed to take his trekking bike when riding a bicycle in town, and he is disposed to leave his trekking bike at home when riding a bicycle out of town. In what follows, I will focus on examples of disposition ascriptions in which the subject has in-between status with respect to neither the disposition in question nor its opposite.

That mimicking and masking cases go hand in hand indicates a further connection between the counterexamples: in the dispositions literature, *the anti-necessity T–F counterexamples about a disposition (fragility, toxicity, malleability) masked are at once anti-sufficiency F–T counterexamples about an opposite disposition (non-fragility, toxicity, non-malleability) mimicked, and the other way round.*

In the light of this, it is curious that Fara (2009: section 2.3) sees an asymmetry between the two types of cases with respect to INTRINSIC, which, according to him, "avoids the problem of 'mimicking'" but "does not help with the problem of masking". After all, if INTRINSIC stumbles on the packaging material masking fragility (a T–F counterexample), then it also stumbles on the packaging material mimicking non-fragility (an F–T counterexample). And if INTRINSIC successfully handles the chalice case as one in which fragility is mimicked (an F–T counterexample), it would at once handle a case of non-fragility being masked (a T–F counterexample). After all (reverting, for simplicity, to SIMPLE for a moment), being disposed to *b* if *d* is analyzed as *b*-ing if *d* were the case, while the opposite disposition—being disposed to not-*b* if *d*—is analyzed as not-*b*-ing if *d* were the case (substitute, say, 'break' for *b* and something like 'being dropped' for *d*); now, for INTRINSIC (and also for SIMPLE), we can produce substitution pairs that differ only in that one has *b* throughout in place of *M*, while the other has not-*b*, while the two substitution instances are alike in truth value and describe

the exact same scenario (for example, fragility masked/non-fragility mimicked or fragility mimicked/non-fragility masked).[7]

Someone might object that the symmetry I'm drawing between masking and mimicking cases relies on there being, for each dispositional predicate of the form 'disposed to *M* under *C*' a dispositional predicate of the form 'disposed to not-*M* under *C*', an unwarranted assumption. To this objection I have four responses. First, even if fragility/non-fragility were an isolated pair of dispositions for which we could reformulate a mimicking case as a masking case, one could no longer claim that INTRINSIC successfully handles mimicks but not masks, as Fara (2009) does; after all, *some* scenarios would be at once masking cases of one disposition and mimicking cases of the opposite disposition. Second, whatever metaphysical reasons someone might have for denying that for every disposition there is an opposite that is also a disposition, the burden of proof is on her side to show this; and even if she managed to show this, it is unclear if such a metaphysical claim about dispositions would cast doubt on the *ascription of* a negative disposition. Third, the metaphysical project just mentioned is complicated by the fact that formulating the positive/negative disposition distinction is nontrivial given that not breaking when dropped seems equivalent to *remaining intact* when dropped. Then it is not at all obvious which of this pair would be the positive and which the negative disposition: the disposition to break (that is, the disposition not to remain intact) or the disposition to remain intact (that is, the disposition not to break). One option that suggests itself is that positive dispositions, unlike negative ones, are about undergoing some sort of drastic change (like dissolving, shattering) when the condition of manifestation obtains. (Defining what counts as drastic change and what doesn't is already nontrivial.) But it is not obvious that this will work either, given examples like being heat resistant or sound proof, which seem like clear candidates for dispositions, yet objects that have them remain (largely) unchanged when the manifestation conditions obtain. Fourth, it does seem perfectly clear that for a broad range of dispositional predicates, their opposites *are* dispositional predicates also (fragility, malleability, water-solubility, to name but a few); if so, then if masking cases aren't exotic then mimicking cases aren't either; at the very least, many mimicking cases aren't as outlandish as the angel and styrofoam examples might lead us to believe.

If we go all the way and claim that for all disposition ascriptions there is a corresponding ascription of the opposite disposition, then we may still think that

[7] There are two further options for substitution pairs (for the examples that aren't in-between dispositions): the case of an ordinary porcelain cup gives rise to two true biconditionals, one with both halves true, T–T (… disposed to break when dropped…), the other with both halves false F–F (…disposed not to break when dropped…). The same holds for the case of an ordinary gold chalice, except the "…disposed to break…" biconditional yields F–F, while the "…disposed to not break…" biconditional yields T–T.

there are two separate types of problems that the chalice and the packaged cup present. But crucially, the philosophically interesting dimension of difference between them is not that

- the chalice presents a mimicking case while the packaged cup presents a masking one; or
- the chalice presents an anti-sufficiency counterexample while the packaged cup presents an anti-necessity one; in other words,
- the chalice presents an F–T counterexample while the packaged cup presents a T–F one.

The reader might wonder if such basic points—about logical and conceptual connections between the masking and mimicking cases—are worth making. But when considering the dispositions debate (as we have done), it becomes clear that these points have fallen into disregard and iterating them clarifies matters. We will shortly recognize their relevance in the context of Huoranszki's appeal to the analogy between the analysis of free will and that of disposition ascriptions.

We have so far considered two conditional analyses of disposition ascriptions, SIMPLE and INTRINSIC, and found that neither can handle the cases with the chalice, the styrofoam cup or the packaged cup. In the next section we will consider yet another pair of counterexamples to SIMPLE, and uncover parallels between the free will and the dispositions debates.

4. CONDITIONALS AND FINKISH DISPOSITIONS: FREE WILL REVISITED

In the previous section, we came to the conclusion that what are called masking and mimicking cases are quite commonplace. For over half a century, another pair of unusual-seeming counterexamples has received extensive attention in the dispositions literature: examples of so-called finkish dispositions, due to C. B. Martin (his examples appeared in print much later, in 1994). A finkish disposition "would straight away vanish if put to the test… A finkishly fragile thing is fragile, sure enough, so long as it is not struck. But if it were struck, it would straight away cease to be fragile, and it would not break" (Lewis 1997: 144).

Let's say that a wire is live when it is disposed to conduct electricity, and dead when it is not so disposed. Now consider a *dead wire that is finkishly dead*: its deadness "cops out" or finks out precisely when its condition of manifestation—being touched by a conductor—occurs:

> The wire … is connected to a machine, an *electro-fink*, which can provide itself with reliable information as to exactly when a wire connected to it is touched by a conductor. When such contact occurs the electro-fink reacts … by making the wire live for the duration of the contact. In the absence of contact the wire is dead. … In sum, the electro-fink ensures that the wire is live when and only when the conductor touches it. (Martin 1994: 2–3; emphasis in the original.)

Call this the *electro-fink case*.

Consider also a *live wire that is finkishly live*: with the electro-fink operating in reverse cycle, "the wire is dead when and only when a conductor touches it" (Martin 1994: 3). Call this the *reverse-cycle electro-fink case*.

The electro-fink and the reverse electro-fink examples have it in common that they involve wires with dispositions that disappear just as their manifestation conditions obtain. These examples aren't as exotic as they might first seem: George Molnar (1999) draws attention to the fact that reverse electro-finks are quite common, in the form of fuses—metal bits that melt in order to interrupt excessive electric current.

And the electro-fink and similar examples bring us right back to an exact parallel to Lehrer's counterexample to clause (i) of the conditional analysis of free will. Huoranszki (2011: 61, 63) discusses Goldman's (1970: 199–200) example of a finkishly non-soluble sugar cube whose non-solubility "finks out", making it water-soluble all of a sudden when the cube is immersed in water. Then the following is true of the cube, *c*: "If *c* is not immersed in water, it is not soluble" or, equivalently, "*c* is soluble only if it is immersed in water", or, equivalently, "If *c* is soluble, then it is immersed in water". As far as appearances go, the cube behaves just like a soluble sugar cube, yet it is not soluble. Let us see how the Lehrer-analog objection to the SIMPLE analysis of disposition ascriptions goes, formulated in terms of the electro-fink case:

First step. (1′), (2′) and (3′) hold for the electro-fink case's finkishly dead wire *i*.

(1′) *i* is live iff it is touched by a conductor; this has two parts:
  (1′a) *i* is live only if it is touched by a conductor.
  (1′b) *i* is live if it is touched by a conductor.
(2′) *i* is not touched by a conductor.
(3′) If *i* were touched by a conductor, it would conduct electricity.

Second step. From (1′a) and (2′), the *modus tollens* inference schema yields (4′):

(4′) *i* is not live.

Third step. (3′) states that *i* satisfies the left hand side of SIMPLE.

Fourth step. An object like *i* shows SIMPLE's clause (I) to be insufficient to define disposition ascriptions, for *i* satisfies (I) (according to (3′) above), yet *i* is not live (according to (4′) above). Hence *with the electro-fink case, for the dispositional predicate 'is live', we have a Lehrer-analog anti-sufficiency F–T counterexample to the SIMPLE conditional analysis of disposition ascriptions*.

Let's say that quasi-mimicking/-masking a property involves mimicking/masking a property when the object's realization conditions don't obtain without mimicking/masking it when the realization conditions do obtain.

The wire *i* quasi-mimicks being live when it is actually dead. Just as the finkishly non-soluble sugar cube described above quasi-mimicks being soluble. To put it differently, *i*'s being a dead wire is quasi-masked when it is touched by a conductor. And the sugar cube's non-solubility is quasi-masked when it is immersed in water. This sounds very much like the chalice and the styrofoam examples we considered above as mimicking cases. Indeed, the electro-fink and the mimicking cases are similar apart from one key respect: the gold chalice and the styrofoam cup are fragile all along, before and after the manifestation condition for fragility—being dropped or struck—transpires; by contrast, the wire *i* is dead before being subjected to the electro-fink and live after. Hence, *with the electro-fink case, for the dispositional predicate 'is dead', we have a Lehrer-analog anti-necessity T–F counterexample to the SIMPLE conditional analysis of disposition ascriptions*.

The reverse electro-fink case involves a live wire (a fuse!) whose disposition is removed—quasi-masked—by the reverse electro-fink. This bears similarity to masking-type cases like the packaged porcelain cup, the key difference being that the cup remains fragile when surviving the drop in one piece, while the reverse-electro-finked wire stops being live as soon as a conductor touches it. Notice, though, that (just as before) the reverse-electro-finked wire can also be construed as an object that quasi-mimicks being dead.

We can now see that David Lewis's proposed INTRINSIC (replacing SIMPLE) is, by adding clause (II), *custom tailored to handle the electro-fink and reverse electro-fink cases*. Here is why. Clause (II) specifies an additional subjunctive antecedent: for an object/substance/person *N* and for an intrinsic property of *N B*, "if *N* were to retain *B* for a sufficient time". This clause takes care of the electro-fink case precisely because in actuality, the wire in question *doesn't* retain its disposition of being dead when touched by a conductor; and if one thinks (along with Lewis) that dispositions are an intrinsic matter, then she expects that the wire has *some* intrinsic property responsible for its deadness, a disposition that the wire would lose when it is touched by a conductor. This way, a scenario most similar to actuality in which (I) holds is one in which (II) does not obtain. By considering

scenarios most similar to actuality in which both clauses obtain, we are looking at a different set of possible worlds than before, and in none of those worlds is the wire electro-finked. The reverse electro-fink case is handled in much the same way for the very same reason: the situations most similar to actuality in which the reverse-electro-finked live wire is touched by a conductor, are ones in which (II) does not obtain; the wire does not remain live; worlds most similar to the actual one in which both (I) and (II) obtain form a different set, none of whose members feature the reverse electro-fink on the wire.

At this point, we have come back to the connection between free will and dispositions: recall that Huoranszki (2011: 61) views accounts of dispositions as analyzing powers/abilities, and being free to act otherwise as a specific application of that analysis to a special power/ability. Further, Huoranszki, below, suggests that problems with SIMPLE indicate why MOORE's clause (i) should be supplemented with clause (ii):

> [Lehrer's] objection is correct exactly because the simple conditional analysis [of free will] is mistaken. The analogy between the Lehrer-Goldman debate about the analysis of our ability to perform an actually unperformed action on the one hand and the problem of finkish dispositions on the other is the key to understand how the original Moorean analysis can be revised and made correct. As [examples about finkish dispositions show], the [SIMPLE] conditional analysis fails *because how objects (or agents) would behave in specific circumstances is not sufficient to grant them a power or ability*. Our definition must also include a clause saying that they do not lose or acquire that ability when the circumstances in which the disposition is about to become manifest obtain. (Huoranszki 2011: 63; my emphasis. Zs. Z.)

I am sympathetic to Huoranszki's final conclusion. Moreover, I think it carries interesting implications for the dispositions debate for reasons that are considerably more decisive than what is given in this passage: I will spell them out at the end of this section.

I disagree with Huoranszki on an intermediate step, italicized above. It should read: "how objects would behave in specific circumstances is *neither necessary nor sufficient* to grant them a power or ability". After all, the electro-fink case (as well as its reverse counterpart) can be construed simultaneously as anti-sufficiency as well as anti-necessity counterexamples. And even if we don't call a wire's being dead a disposition (a questionable assumption to begin with, as I argued in the previous section), one of the finkish dispositions—the reverse electro-fink case with the live wire turning dead when touched by a conductor—is still an anti-*necessity* counterexample to SIMPLE.

In addition, calling for a supplementary condition to MOORE's (i) in the way that Huoranszki does in this passage is not an ordinary case of a definition being too broad, suffering from insufficiency, in need of being narrowed by one

or more additional conditions—along the lines of first "defining" 'zucchini' too broadly as a type of summer squash but (since the condition is satisfied by various squashes that aren't zucchini, such as the acorn squash) subsequently narrowing it by adding conditions like 'cylindrical in shape' or 'usually green'. Using this model for what clause (ii) is doing ignores the fact that the second half of the biconditional in HUORANSZKI does not merely list narrowing conditions: instead, it consists of a *past subjunctive conditional whose antecedent is being supplemented with additional conditions* (similarly to the move from SIMPLE to INTRINSIC). But that does something other than narrowing the second half of the biconditional: it narrows it in some respects and broadens it in others (similarly to when we added clause (II) to SIMPLE). This is also shown by the fact that in the case of finkish dispositions, adding an analog of Huoranszki's clause (ii) to SIMPLE would *simultaneously* take care of both the anti-sufficiency and the anti-necessity counterexamples, both the electro-fink and the reverse electro-fink case:

NONREDUCTIVE conditional account of disposition ascriptions
An object/person/substance *N* is disposed to *M* under *C* iff
N would *M* if
(I) it were the case that *C*, and
(II′) *N* were to retain its disposition to *M* when *C* for a sufficient time.

Here is why NONREDUCTIVE is not simply further narrowing the right hand side of SIMPLE. With respect to the disposition of being live, the electro-finked dead wire *i* no longer presents an anti-sufficiency F–T counterexample because we are considering only those (counterfactual) scenarios in which *i* is touched by a conductor (by I), and the *electro-fink* no longer interferes with *i*'s *originally dead* disposition (by II′); the biconditional's second half claims that, among these situations, the scenarios most similar to the actual world all have *i* conducting electricity. Crucially, the set of most similar worlds is not a *narrowing down*, a *subset* of the set selected by (I) alone (as in SIMPLE). And, with respect to the disposition of being live, the reverse-electro-finked live wire *i′* no longer presents an anti-necessity T–F counterexample, because we are considering only those (counterfactual) scenarios in which *i′* is touched by a conductor (by I), and the *reverse electro-fink* no longer interferes with *i′*'s *originally live* disposition (by II′); and among those, the scenarios most similar to the actual world all have *i′* conducting electricity. As before, crucially, the set of most similar worlds is not a *narrowing down* or a *subset* of the set selected by (I) alone (as in SIMPLE).

NONREDUCTIVE is special in that the analysis in the second half itself includes the very disposition to be analyzed.[8] HUORANSZKI is another instance of a nonreductive analysis in which the very ability to perform an actually unperformed action, *the ability/power to act otherwise*, is featured in the analysis. Huoranszki (2011: 68–71) considers this nonreductive feature of his proposal; we'll do the same in the upcoming section. Beforehand, let's consider how NONREDUCTIVE (inspired by HUORANSZKI's clause (ii)) clarifies some of the issues surrounding the analysis of disposition ascriptions.

Notice that NONREDUCTIVE does not resolve the scenarios with the packaged porcelain cup, the styrofoam cup or the golden chalice: as things stand, all three remain counterexamples.[9] In this respect, just like INTRINSIC, NONREDUCTIVE seems custom tailored to handle the finkish cases.[10] Clearly, those who want to defend this account have to say something further about these outstanding problem cases. Why think NONREDUCTIVE is a step in the right direction then? Problems arising in connection with a recent proposal that received considerable attention—Michael Fara's (2005) so-called habitual analysis, which aimed to counter the weak points of INTRINSIC—confronts problems that make the inclusion of something like (II′) seem highly attractive.

Habitual claims—like "Mary smokes when she gets home from work" or "Huoranszki rides a bike to work"—"have something to do with what is *normally*, *or typically*, *or generally* the case" (Fara 2005: 64). Nonetheless, Fara insists that we shouldn't analyze them in terms of conditionals with an adverbial prefix like "Normally, if Huoranszki goes to work, he rides a bike".

> Fara (2005) proposes that instead of using conditionals to give an analysis of dispositions we should use what he calls *habituals*. Habituals are a commonplace device for characterizing how an object habitually behaves. [Huoranszki rides a bike

[8] Bird (1998) and Molnar (1999) are prominent proponents of nonreductive analyses of disposition ascriptions.

[9] After all, the chalice and the styrofoam cup actually retain their nonfragility (and the packaged cup its fragility), and are *actually* dropped, yet in the actual world, they break (don't break). So the counterfactual's antecedent conditions hold for the actual world while the consequent doesn't. This is not surprising given that clause (II′) of NONREDUCTIVE (in parallel with INTRINSIC) is custom-tailored to handle cases in which a disposition is lost/gained precisely when the realization condition is about to obtain. This is a feature that the finkish examples have but the cup examples lack.

[10] Of course, these sorts of problems don't arise in the context of the free will debate, in which mimick-type cases don't seem to be possible, and—depending on what we think about free will being compatible with some forms of the ability to act otherwise being masked—masking-type cases might not arise either. Fara (2008)—unlike Huoranszki (if I understand him correctly)—does think the ability to act otherwise can be masked without removing the ability.

> to work], for example, is true even if, occasionally [Huoranszki rides the tram to work]. In this sense, we can think of habituals as expressing universal generalizations that tolerate exceptions. (Choi–Fara 2012: section 1.4.)

Fara's HABITUAL analysis of disposition ascriptions
An object/person *N* is disposed to *M* when *C* iff
(I′) there is an intrinsic property that *N* has in virtue of which it *M*s when *C*.

HABITUAL is custom-tailored to handle the masking cases: the porcelain cup is fragile even when on occasion it is packaged. It has an intrinsic feature in virtue of which it usually breaks when dropped. Fara takes this to be the analysis' selling point: "The best reason to prefer the Habitual Account is that it can solve the problem of masking" (Fara 2005: 71). HABITUAL adopts from Lewis (1997) the inclusion of an intrinsic property. The account is successful at handling all counterexamples to SIMPLE that we have discussed: the angel's presence around the gold chalice is exceptional; in the majority of the relevant scenarios (actual and counterfactual) in which the chalice is dropped, it doesn't shatter, making the habitual "The chalice shatters when dropped" false despite the exceptions. Moreover, the chalice's not breaking is a matter of its intrinsic features, so the disposition ascription "The chalice is fragile/disposed to shatter when dropped" is false according to HABITUAL, just as we wanted. Parallel explanations make "The styrofoam cup within earshot of the Styrofoam Hater is fragile", "The electro-finked wire is live" likewise false, and "The reverse-electro-finked wire is live" true. All promising results so far. But HABITUAL faces profound problems.

Juhani Yli-Vakkuri (2010) forcefully argues that despite Fara's insistence that his is a non-conditional analysis, HABITUAL is logically equivalent to a *ceteris paribus* conditional analysis like the following:

Ceteris Paribus (CP) conditional analysis of disposition ascriptions
An object/person/substance *N* is disposed to *M* under *C* iff
ceteris paribus N would *M* if
(I) it were the case that *C*.

The idea is that an object is disposed to break when dropped just in case, other things being equal, the object would break if dropped. Such a modification of SIMPLE is supposed to exclude exactly those cases that are counterexamples to SIMPLE: masking, mimicking, electro-fink and reverse electro-fink cases. But the move to CP renders the analysis vacuous, claim C. B. Martin (1994: 5–6) and Yli-Vakkuri (2010: 664–665). For the way to understand CP is that it is supposed to exclude all cases that are counterexamples to SIMPLE. But what such cases have in common is no more than that they are counterexamples to SIMPLE. So if Yli-

Vakkuri is right that HABITUAL reduces to CP, then the former is just as vacuous as the latter.

There are two further considerations of my own that also point in the direction that HABITUAL is ultimately vacuous: Fara's remarks about the context sensitivity of HABITUAL on the one hand, and about so-called entrenched finks on the other. We'll consider these in turn.

First, Fara (2005: 74–76) imagines that the porcelain cup is an extremely valuable museum piece that specialists have carefully packaged; suppose the packaging is invisible. A vandal is determined to break the cup, does not know about the packaging, and deploys all manners of destruction but does not succeed. Exasperated, he says

(5) The damn cup just doesn't break when struck!

The museum specialists respond:

(6) Oh, but it *does* break when struck (that's why we protected it in the first place).

How can both the vandal's and the specialists' utterance be true? Fara's response: "the cup does have an intrinsic property—say its weak molecular bonding—in virtue of which, if the *museum specialists* were to utter (6), *they* would speak truly" (Fara 2005: 76; emphases in the original). Meanwhile, Fara is suggesting that in the context of the vandal's utterance, this particular property will not make the second half of HABITUAL true, nor will any other intrinsic property of the cup's. But why exactly is that the case? Fara gives no explanation apart from saying that (5) in the vandal's mouth seems true and that context dependence can allow for (5) and (6) being simultaneously true, in the same way that two people can simultaneously make true utterances by one saying "I am happy" and the other "I'm not happy". But that does not *explain* how the vandal's utterance comes out true rather than false according to HABITUAL.

Second, Fara (2005: 76–78) makes a distinction between transient versus entrenched finks. A transiently finkish wire is attached to the electro-fink only "temporarily, or rarely, or sporadically". Fara gives two examples of entrenched finkishness:

> Imagine that the copper wire, when placed in the circuit, is immediately attached to the [reverse electro-] fink, perhaps as a safety precaution… We might say that being attached to the fink is a "way of life" for the wire. This would be a case of *entrenched finkishness*. … If being attached to such a device is a way of life for the wire, then it seems absurd to say that the wire is disposed to conduct electricity when touched by a conductor…

> [s]uppose a sturdy wooden barrel, which we might ordinarily describe as being disposed to roll when pushed, is nailed to the floor of a seafood restaurant, to add to the restaurant's nautical décor. We can imagine that the barrel has been on display in the restaurant for several years, and that it regularly resists attempts of drunken patrons to roll it by pushing. Again, it seems clear that this barrel, unlike most others, is *not* disposed to roll when pushed; instead it is disposed to stay perfectly still when pushed, as the unsuccessful attempts of the drunken patrons attest. (Fara 2005: 77–78; italics in the original, my underlining. Zs. Z.)

The underlined parts highlight Fara's appeal to what he considers decisive linguistic intuitions to the effect that the wire with the entrenched reverse-electro-fink is not disposed to conduct electricity, and that the nailed-down barrel is not disposed to roll when pushed, so (7) is false.

(7) The barrel is disposed to roll when pushed.

Fara claims that HABITUAL yields just this verdict: despite intrinsic features of the wire and the barrel, the right hand side of the biconditional comes out false due to the enduring extrinsic finks—because the wire's *not* conducting electricity is not an exception but the rule for that particular wire (given the entrenched fink); and because the barrel's *not* rolling is likewise not an exception but the rule for that particular barrel.

I see a problem with Fara's take on the cases of entrenched finkishness. It is unclear that HABITUAL makes the barrel's immobility unexceptional; that depends on what the range of relevant scenarios is among which immobility is unexceptional. But as we saw in the previous paragraph, utterances of disposition ascriptions are context-sensitive, so we can imagine, besides the false utterance Fara is highlighting, a true utterance of (7) in the case of entrenched finkishness (the same way we had imagined the vandal and the museum specialists). So somehow the range of relevant scenarios in the two contexts has to be such that immobility is exceptional in the case of the true utterance but unexceptional in the case of the false one. There is, however, no guidance whatsoever from HABITUAL as to how we might achieve this other than by fiat: for each utterance, the range, whatever it is, has to be such that the truth value comes out right. But that is just as vacuous as saying that an utterance is true when it's true and false when it's false.

In sum, there are several reasons to think HABITUAL is vacuous. The account can accommodate any intuition (putative or real) about utterances involving disposition ascriptions; but such extreme flexibility breeds lack of explanatory power.

The possibility of entrenched finkishness has implications that go beyond HABITUAL. Recall that it had seemed as though INTRINSIC was making headway

with the finkish cases even if not the others (the chalice and the packaged cup). But if we grant the intuitions Fara is appealing to in the entrenched finkishness cases—admitting that (7) is false when said about a barrel whose "way of life" is being nailed down in a restaurant—then we have undermined INTRINSIC. For the very idea behind INTRINSIC had been that disposition ascriptions are an intrinsic matter and extrinsic features, whether they be transient or entrenched, make no difference to what dispositions the object in question does or doesn't have. This way, what hope of progress INTRINSIC brought relative to SIMPLE is tarnished.

To be sure, NONREDUCTIVE is still susceptible to counterexamples like the chalice and the packaged porcelain cup. Yet in the light of the shortcomings of various reductive analyses of disposition ascriptions (SIMPLE, INTRINSIC, HABITUAL), NONREDUCTIVE is showing renewed promise: at the very least, it can handle finkish cases across the board, be they transient or entrenched.

5. WHY NONREDUCTIVE DOESN'T MEAN NONSTARTER

Is a nonreductive analysis of free will such as HUORANSZKI worth having? Huoranszki (2011: section 1.4) gives an affirmative answer, claiming that the reference to ability retention in clause (ii) …

> …is innocent and does not make our analysis viciously circular. The second occurrence of the ability in the analysans stipulates only that whatever the analysandum refers to will not be altered by the circumstances that the analysans specifies. And this can be included in the specification of the circumstances among which the ability would become manifest without making the analysis uninformative. (Huoranszki 2011: 68.)

Shortly after, Huoranszki claims that a nonreductive analysis is the only one worth having, and suggests that the reasons for this are illuminated by the problems that reductive analyses of disposition ascriptions encounter:

> My version of the conditional analysis does not aim to provide a reductive analysis of the abilities relevant for freedom of the will. In fact, I do not think that any such analysis is worth seeking or can be given. To see why, let me compare again my analysis of free will to the analysis of dispositions. A reductive analysis of dispositions would require that we analyze the meaning of every power which can be (truly) ascribed to an object in terms of occurrent and/or categorical properties. However, I doubt that any such analysis is possible at least for two reasons. (Huoranszki 2011: 69.)

Huoranszki's second reason is that in spelling out the appropriate manifestation conditions for a dispositional notion like fragility (a kind of power), we inevitably resort to including powers in our analysis. For example, it won't do to say that a fragile object is one that would break if it were struck; we need to say something like "if it were struck *by a hard object*". Such an analysis of fragility is illuminating even though it involves hardness (another power). We can appreciate this point even more if we consider in-between objects like the neither fragile nor non-fragile melamine cup which does have more specific dispositions, like being disposed to break when dropped with a certain minimum speed/from a certain minimum height, etc.

Part of Huoranszki's first reason is that accounts in terms of intrinsic properties (like INTRINSIC and HABITUAL) are problematic because it is doubtful that every disposition is intrinsic.[11] The intuitions surrounding entrenched finkishness cases—(7) being false—provide further support for this claim.

I concur with Huoranszki's conclusion: it is unlikely that disposition ascriptions are susceptible to a reductive analysis. But the reasons I spelled out in this paper are distinct from Huoranszki's. When it comes to disposition ascriptions, we can hope to handle finkish cases along the lines of INTRINSIC. Masking and mimicking cases—related in a crucial way that is not usually recognized in the literature—might move us towards HABITUAL. But then we are faced with cases of entrenched finkishness undermining not only HABITUAL but also casting doubt on what INTRINSIC had achieved. The way out seems to take us in the same nonreductive direction that Huoranszki's clause (ii) does.[12]

REFERENCES

Bird, Alexander 1998. Dispositions and Antidotes. *The Philosophical Quarterly* 48. 227–234.

Choi, Sungho – Michael Fara 2012. Dispositions. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition). URL = <http://plato.stanford.edu/ archives/ spr2012/ entries/dispositions/>.

Fara, Michael 2005. Dispositions as Habituals. *Noûs* 39. 43–82.

Fara, Michael 2008. Masked Abilities and Compatibilism. *Mind* 11. 843–865.

[11] The other part of Huoranszki's first reason is this: an analysis in terms of the internal structure of an object can hardly avoid mentioning dispositions: "it seems that an internal structure can 'realize' a disposition only if the *structure itself* has certain dispositional features" (Huoranszki 2011: 69; emphasis in the original); an analysis like INTRINSIC does not, in the end, present a reductive, disposition-free alternative then.

[12] This paper has benefitted from comments by two anonymous referees and the audience at the CEU conference on Huoranszki's book, especially Ferenc Huoranszki; I thank them all. The present research was supported by the Bolyai János Research Fellowship of the Hungarian Academy of Sciences (MTA), and Grant No. K-19648, entitled Integrative Argumentation Studies, of the Hungarian Scientific Research Fund (OTKA).

Fara, Michael 2009. Dispositions. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2009 Edition). URL = <http://plato.stanford.edu/ archives/sum2009/ entries/dispositions/>.

Goldman, Alvin I. 1970. *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.

Huoranszki, Ferenc 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Johnston, Mark 1992. How to Speak of the Colors. *Philosophical Studies* 68, 221–263.

Lehrer, Keith 1968/1982. Cans without Ifs. In Gary Watson (ed.), *Free Will*. Oxford: Oxford University Press, 41–45.

Lewis, David 1973. *Counterfactuals*. Oxford: Blackwell.

Lewis, David 1997. Finkish Dispositions. *The Philosophical Quarterly* 47, 143–158.

Martin, Charles Burton 1994. Dispositions and Conditionals. *The Philosophical Quarterly* 44, 1–8.

Molnar, George 1999. Are Dispositions Reducible? *The Philosophical Quarterly* 49, 1–17.

Moore, George Edward 1912. *Ethics*. New York: Henry Holt.

Smith, David A. 1977. Dispositional Properties. *Mind* 86, 439–445.

Yli-Vakkuri, Juhani 2010. Conditional and Habitual Analyses of Disposition Ascriptions. *The Philosophical Quarterly* 60, 624–630.

Howard Robinson

Ferenc Huoranszki's Libertarian Compatibilism

1. INTRODUCTION: HOW I SEE THE OVERALL STRUCTURE OF HUORANSZKI'S ARGUMENT

(a) Huoranszki's 'dualism'

Huoranszki (2011) is a dualist in the following sense:

> (a) he is a libertarian within the psychological;
> (b) but thinks that Free Will is neutral on the question of physical determinism.

So he thinks he can combine a psychic libertarianism—giving free will everything we ever really wanted—with a deterministic physics.

The reason for (a) is that choice is a *sui generis* phenomenon that cannot be analysed in terms of the quasi-mechanistic interplay of belief-desire phenomena that psychological determinism requires.

> [The belief-desire] picture, it seems to me, misses something fundamentally important about our human agency. Putting beliefs and goal-determining 'desires' with appropriate contents together (no matter how complex those contents and attitudes are) might be sufficient to understand how a goal-directed optimalizing machine works, but no responsible human agency will ever emerge from this picture. What's missing is agents' ability to control their actions by their choices. (Huoranszki 2011: 116.)

The reason for (b) is two-fold. (i) The Consequence Argument that purports to demonstrate the incompatibility of free will and determinism does not work (Huoranszki 2011: 12–27). (ii) The conditional analysis, which does work, is compatible with physical determinism (Huoranszki 2011: 54–74). It also accommodates choice in a way that makes the conditionality of freedom different from that of other dispositions, e.g. fragility (Huoranszki 2011: 75–95).

(b) Huoranszki's compatibilism

The overall rationale of the book could be put in the following form

1. The dynamic of psychological discourse is indeterminist but not chancy or random.
2. The dynamic of psychological discourse is autonomous—that is, not tied to the dynamics of physical discourse—even if the physical is deterministic.
3. Psychological discourse can be taken in a fully realist manner.

Therefore

4. There is real free will whether or not the physical realm is deterministic.

Huoranszki argues for 1 in the way mentioned above and in his discussion of the role of reason. And for 2 by his rejection of the Consequence Argument (henceforth 'CA'). 3 is, I think, assumed. I shall try to cast doubt on whether 3 is possible in a world 'closed under physics' later.

My main objective here is to discuss chapter 4 "The conditional analysis of free will", but I cannot do that without saying something about the special nature of choice and how, in my view, the CA should be expressed.

I also want to say something about whether Huoranszki can get away with his dualism.

2. HUORANSZKI'S INTERPRETATION OF MOORE'S CONDITIONAL THEORY

Moore's statement of the conditional theory [quoted-cum-paraphrased] is as follows.

> It is, therefore, quite certain (1) that we often *would* have *acted* differently if we had chosen to; (2) that similarly we often would have *chosen* differently, *if* we had so to choose; and (3) that it was almost always *possible* that we should have chosen differently, in the sense that no man could know for certain that we should *not* have so chosen. All these three things are facts, and all of them are quite consistent with the principle of causality. Can anyone undertake to say for certain that none of these three facts and *no* combination of them will justify us in saying we have Free Will? (Moore 1912: 220–221; Huoranszki 2011: 55.)

Huoranszki takes exception to (2) and (3), but I think he misunderstands them both. In relation to (2), he quotes the following passage from Moore.

> [T]here is no doubt it is often true that we should have chosen to do a particular thing if we had chosen to make the choice; and that this is a very important sense in which it is often in our power to make a choice. (Moore 1912: 219; Huoranszki 2011: 56.)

In the light of the first part of this quotation, he takes (2) as implying that when we choose, we have chosen to choose (and we might have chosen to choose differently), and he thinks that this leads to a regress. No doubt it does, but this is not the most plausible interpretation of what Moore says. The phrase 'if we had chosen to make the choice' is not most plausibly or charitably interpreted as implying a regress where we choose our choices, but as meaning simply 'if we had chosen to make the *other* choice—that is, if we had chosen differently'. This, I think, is all Moore needs to make the point he wants to make.

On (3), Huoranszki claims that the ability of others to predict our choices is irrelevant to whether they are free—we can often predict what our friends will do. But I think that Moore is making a point common to the compatibilist case, namely that our lack of omniscience is important to our possession of a sense of freedom, both in ourselves and others. If we really could detect all the causes, it would be hard to see actions as free.

Huoranszki challenges Moore's version of the conditional theory in order to clear the way for his own version of it. If what I say above is correct, Moore's statement of the theory may be closer to what the conditionalist must say.

3. THE ABILITY TO ACT OTHERWISE: LEHRER'S OBJECTION

Lehrer (1968/1982) argues that there is a case under which it is false that

(i) *S* could have done *A*

Although it is true that

(ii) *S* would have done *A* if he had chosen to do *A*.

This is because (ii) is consistent with

(iii) *S* could have done *A* only if he had chosen to do *A*

And

(iv) It is not the case that *S* chose to do *A*

But (iii) and (iv) entail

(v) It is not the case that *S* could have done *A*

There is, however, an ambiguity between

(a) It could have been the case that *S* did *A* only if he chose to do *A*.

And

(b) *S* would have had the ability to do *A* only if he had chosen to do *A*.

(b) is distinctly strange, because choosing to do something is not usually—and perhaps can hardly be—what endows one with the ability to do it: rather you can effectively choose it only if you have the ability. What we seem to have here is a case of a finkish disposition. Finkish dispositions (the term was invented by C. B. Martin 1994: 2–3) are defined by Huoranszki as follows.

> Finkish dispositions have a special feature: either objects lose them in circumstances in which they are about to be actualized, or, inversely, objects acquire them only in the circumstances in which they are about to be actualized, and they disappear otherwise. (Huoranszki 2011: 63.)

An ability that you acquire only when you choose to actualize it is clearly finkish. These are clearly not dispositions of the normal or natural kind. Huoranszki is, therefore, only restoring the natural sense when he provides the following account.

> *S*'s will is free in the sense of having the ability to perform an actually unperformed action *A* at *t* iff S would have done *A*, if (1) *S* had chosen to and (2) had not changed with respect to her ability to perform *A* at *t* and (3) had not changed with respect to her ability to make a choice about whether or not to perform *A* at *t*. (Huoranszki 2011: 66.)

It is worth noting that Frankfurt's (1969) classic objection to the 'could have done otherwise' condition on free will is no better than—is more or less a version of—the invocation of finkish dispositions, and just as implausible. The fact that an evil scientist can intervene and prevent you from making a choice you are about to make does not alter the facts about your *natural* capacities, *ceteris paribus*, which is all that disposition and capacity talk, whether of humans or any other kind of object, can consist in.

Compare this to Moore's three conditions. Huoranszki's are just clarifications of the conditionality involved in (1). But being conditional just on choice is, indeed, libertarianism—you would have done otherwise simply if you had chosen to. The compatibilist normally wants to say something about conditions for making a different choice—that is what Moore's (2) does.

There is an important issue in the interpretation of the conditional analysis here. In a sense, conditionality on choice alone is not what the conditional analysis is about. 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. I'm tempted to say that Huoranszki's theory is not really a conditional theory at all. The rationale (according to me) of the book, given in 1–4 above, does not mention conditionals. This does not mean that abilities are not essential to Huoranszki's account, but these abilities are not Mooreanly conditional, for there are not explained in terms of different circumstances producing different choices.

4. THE CONSEQUENCE ARGUMENT

(a) The two versions

Huoranszki discusses and claims to refute two versions of CA. The first centres on the supposed transitivity of the operator 'has no choice over'. Its central claim is that if *S has no choice over* the truth of *P*, which expresses the total state of the universe at some time in the past, well before *S* existed, and if *S has no choice over* the truth *L*, which expresses all the laws of nature, then, given determinism, *S* also *has no choice over* the truth of *A* which states what he is now doing, or will be doing in five minutes.

The problem with this argument is as follows. The compatibilist says that the causal mechanism operates through—among other things—the choices we make, though these themselves are determined. So it will not do to say that *S* has no choice over *A*, because *S*'s choice is one of the (caused and causal) factors that bring about *A*. 'Has no choice over', is not, therefore a transitive relation, as the argument requires. Huoranszki is therefore correct in his dismissal of this argument.

There is, however, the important following reservation, which will become relevant later. If you are a psychological determinist, then the transitivity of the relation does not hold, because choices are amongst the things that determine outcomes. But 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. We will return to this later.

The second argument is as follows. '*Po*' represents the complete state of the universe at some time before *S*'s birth; '*L*' represents all the laws of nature, and '*Q*' represents some event coincident with a time at which we believe *S* might act.

[1] If determinism is true, then the conjunction of *Po* and *L* entails *Q*.
[2] It is not possible that *S* could have *A*-ed at *t*, and *Q* be true.
[3] If [2] is true, then if *S* could have *A*-ed at *t*, *S* could have rendered *Q* false.
[4] If S could have rendered *Q* false, and if the conjunction of *Po* and *L* entails *Q*, then *S* could have rendered the conjunction of *Po* and *L* false.
[5] If *S* could have rendered the conjunction of *Po* and *L* false, then *S* could have rendered *L* false.
[6] S could not have rendered *L* false.

Therefore

[7] If determinism is true, *S* could not have *A*-ed at *t*.

Huoranszki deals with this argument in a way very similar to his treatment of the previous one. He argues that the fact that one does not exercise an ability on a given occasion does not show that one has lost it. This holds true even when there is a deterministic explanation of why one did not exercise it at a given time. (After all, the fact that a fragile glass did not break when it was *not* dropped does not mean it lost its fragility: or, if you want a positive disposition, the fact that the gunpowder did not explode when no-one lit the fuse does not mean it did not possess its explosive capacity throughout.) Huoranszki brings this home by substituting for '*A*' in the argument 'spoke in the last five minutes'. If one did not so speak, it does not mean one lost one's ability to speak in that time, only that one did not exercise it.

This response to the argument as it stands is plausible, but I think that this only shows that the argument should be formulated in a slightly different way.

I think it should go as follows.

(1) I am not free to do something that I do not have the causal power to do.
(2) I do not have the causal power to do something the opposite of which is strictly causally necessitated by factors beyond my control.

Therefore

(3) I am not free to do something the opposite of which is strictly causally necessitated by factors beyond my control. {1,2 Hypothetical syllogism}
(4) Initial state [*Po*] and laws of nature [*L*] are factors beyond my control.

Therefore

(5) Anything strictly causally necessitated by *Po* and *L* is something the opposite of which I am not free to do. {3,4 MP}
(6) If determinism is true, *Po* and *L* strictly causally necessitate all my actions.
(7) Determinism is true.

Therefore

(8) *Po* and *L* strictly causally necessitate all my actions.

Therefore

(9) I am not free to do the opposite of anything that I actually do. {5,8 MP}

The assumptions are 1, 2, 4, 6, 7. No-one disputes 4, 6 follows from definition of determinism and 7 is *ex hypothesi*. So the compatibilist must dispute 1 or 2, and I do not see how.

(b) Van Inwagen and avoiding the appeal to 'cause'

Van Inwagen's version of CA is stated in terms of the entailment of what will happen by the antecedent conditions and the laws; there is no explicit mention of causation, whereas I appeal to causal powers. This is no accident.

Van Inwagen says

> The reader will note that the horrible little word 'cause' does not appear in this definition [of determinism]. Causation is a morass in which I for one refuse to set foot. (Van Inwagen 1983: 65.)

This applies not just to this definition, but to the whole statement of the argument.

Huoranszki, more explanatorily, says

> For a long while, this question was formulated in terms of causes, more precisely, as the problem of how 'universal causation' is compatible with human freedom. …
>
> This way of formulating the problem has, however, lost popularity in the last couple of decades because the argument from universal causation relies on two assumptions that many philosophers would reject. First, the argument assumes that the occurrence of the cause must metaphysically necessitate its effect. It is in this sense that universal causation renders events non-contingent. Hume has famous-

> ly claimed, however, that we 'can always *conceive* any effect to follow from any cause…' and that 'whatever we conceive is possible, at least in the metaphysical sense' … Second, the argument presupposes that causation must be deterministic … (Huoranszki 2011: 12.)

Huoranszki's second worry is irrelevant because we are discussing determinism, and that so the assumption of determinism would beg no questions: but, in fact, there is no assumption about all causation being deterministic in my argument. Nor, as far as I can see, is there anything in it that a Humean about causation (that is, a regularity-cum-counterfactual theorist, like, for example, David Lewis) could object to. Someone might argue that, as the 'powers' conception of causation is anti-Humean, the expression 'causal powers' rules out a Humean interpretation of cause. This is not so. The expression 'possessing a causal power' in context merely means the possibility of exercising (as opposed to merely possessing) at a given time an ability one possesses, however that is interpreted metaphysically.

I do not see that this argument is dependent on any particular (and, hence, controversial) understanding of causation. So I do not see the force of van Inwagen's objection to employing that term.

Van Inwagen tries to avoid objections like Huoranszki's by arguing that we are not disputing abilities in general but the power to do particular things on particular occasions. So it is not the existence of some general dispositional or ability-state that is at stake, but rather the possibility of someone's acting in a certain way at a particular time. This surely captures more accurately than any general ascription of a capacity what we mean when we assert that someone was free to do something at a particular time. This could be built into the argument as follows.

(1') I am not free to do something at *t* if I do not have the causal power at *t* to do that thing at *t*.

So it is not merely a case of having-at-*t* the relevant capacity, but having the relevant capacity to do-at-*t* the action in question.

(2') I do not have the causal power at *t* to do at *t* something the opposite of which is nomologically necessitated by factors beyond my control.

Therefore

(3') I am not free to do at *t* something the opposite of which is nomologically necessitated by factors beyond my control.

(4) [As before.]

(5') Anything nomologically necessitated by Po and L is something the opposite of which I am not free to do. {3,4 MP}
(6') If determinism is true, *Po* and *L* nomologically necessitate all my actions.
(7) [As above.]

Therefore

(8') *Po* and *L* nomologically necessitate all my actions.

Therefore

(9) [As above.]

The only way of resisting this argument is to insist that to be 'free to *A* at *t*' is no more than having, at *t*, the general capacity for *A*-ing, in the same way as a glass is fragile at *t* because if someone had dropped it, it would have broken. I think it is clear that this does not capture the idea of free *choice*, as I shall argue in the next section.

(c) Competition between causes, part (i)

The question of to what extent causal lines, or types of causation, might conflict, and how, has been a perennial issue.

Moore's account of the issue (quoted by Lehrer 1966: 189) is as follows.

> All that is certain about the matter is: (1) that, if we have Free Will, it must be true, in *some* sense, that we sometimes *could* have done what we did not do; (2) that, if everything is caused, it must be true in *some* sense, that we *never could* have done, what we did not do. What is very *un*certain, and what certainly needs to be investigated, is whether these two meanings of the world 'could' are the same. (Moore 1912: 131.)

Lehrer adds

> The really crucial question to be answered here is the following: Is it logically consistent to say both that a person could have done otherwise, in the sense of "could" related to free will, *and* that he could not have done otherwise, in the sense of "could" related to causation? The reason why this question is crucial is that if the answer to the question is negative, then free will and determinism are logically inconsistent, even if the two senses of "could" mentioned above are quite different. (Lehrer 1966: 190.)

The purpose of the conditional analysis is to give the 'non-causal', 'freedom-related' sense of 'could'. This might be thought of as replacing (1) in my argument with

> (1'') I *am* free to do something that I do not have the causal power to do, provided that I could have done it in the conditional sense, i. e., I would have done it, if I had chosen to.

This suggests that Moore and the conditional analysis accept my version of the argument, meaning by that it accepts that *if* one employs the causal-related sense of 'could' (rather than the 'would have done otherwise in certain different circumstances' sense) in one's definition of freedom, then freedom and determinism are inconsistent with each other. My version of the argument might seem, therefore, to be simply question-begging, because it employs the wrong sense of 'could'—the causal, not the conditional one. But what is the 'freedom' sense of 'could'? Simply to invoke the conditional 'would have if…' sense is just too weak. The window would have behaved differently—it would not have shattered—if it had not been hit by the stone, but that does not make it a free agent, so conditionality alone is not enough. Of course, the difference in the case of the window does not run through a causal line involving a choice, but, if determinism is true, it is not the choice that makes the difference—it merely executes the mandate of nature which is antecedently determined—a nature which could never in fact have been different. (If you want to allow for quantum indeterminacy you could add '—except by random indeterminacy, never by deliberate choice'.) It seems to me that Huoranszki wants conditionality *upon choice alone*, but that presupposes that choice itself is not determined, for if it is determined then the action is not conditional on the choice alone but is equally conditional on the factors that determine the choice. Maybe Moore, too, in claiming that there are two different senses of 'could', is also trying to privilege choice as a determining factor in a way that is actually inconsistent with determinism. You could, of course, argue that there is a 'freedom' sense of 'could', just in the sense that some, and only some, determined processes and counterfactuals run through choices; but then you could equally well claim that there is a 'weather' sense of 'could' on the grounds that some processes and counterfactuals run through weather events ('if the rain had not stopped, the match would have been cancelled'). The point is that these distinctions are not between *causal* and *some other non-causal* senses of 'could', but only between the different elements that might be contained in straightforwardly causal processes.

So what the Conditional Argument shows is that the only sense of freedom open to the compatibilist in this area is 'if the world had been different in some of its initial conditions (including mid-process random events) or laws so as to affect my choices, then I would have done something different'. But this seems

a very attenuated sense of freedom, because 'if things had been originally different, different things would consequently have occurred' applies equally to events that have nothing to do with choice or freedom, as it does to choices.

5. AN ARGUMENT FOR LIBERTARIANISM (MEETING HUORANSZKI HALF WAY!)

(a) The argument

Fred gets up on Saturday morning. He decides to go shopping. On the one hand, this decision is not a random or chance event because

(i) he usually goes shopping on Saturday morning;
(ii) he realizes he needs some milk, is running out of coffee and fancies some fresh fish for dinner.

But it is not determined because

(iii) though he usually goes shopping on Saturday, this is a habit, not an addiction—if he decides he really needs nothing, he stays in and reads a book, with some relief.
(iv) There is no acceptable way of turning this behaviour into a deterministic scenario, for the following reasons

(a) There are no deterministic laws relating usually shopping, needing milk and coffee, wanting fish and going shopping. There are only loose generalizations.
(b) The situation cannot be turned into a deterministic one by adding further psychological-cum-conditioning circumstances. E.g. if your mother held you *this* way…and you usually shopped on Saturdays and wanted milk… then…
(c) It cannot be turned into something deterministic by *supplementation from a lower level*. E.g. factors as above **plus** being in neural states *N1*…*Nn*.

Why not? Because this sort of supplementation only works under the following conditions: an explanation at level *L1* can be made deterministic by supplementation from factors at level *L2* only if there is a deterministic explanation available in terms of *L2* alone—in the classic case, when the world is 'closed under physics', that is, when all micro entities follow the same basic physical laws, whether or not they are part of some complex entity: when there are, that is, no truly *emergent* laws. So the determinism does not come from the psychological *supplemented by* the physical, but from the physical alone. So one cannot have indeterministic or libertarian psychological explanations that are integrated into an overall deterministic scheme. By 'integrated' I mean being an essential part

of the process that is deterministic: so if the deterministic explanation need employ only sub-psychological concepts, the psychological is not 'integrated into' it.

This raises the question of what the relation is between an indeterminist psychology and the possibly mechanistic physical reality in which it is realized.

(b) Competition between causes, part (ii)

Does this not show that physical determinism (but not psychological determinism) and libertarian freedom are compatible?

I have just claimed, in line with Huoranszki, that psychological explanations have a libertarian logic—they are essentially non-deterministic, but not by the adding of a merely random element. But 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. It seems to me that a notion of freedom or agency that allows that it has no role in determining one's bodily location, movements, speech sounds, etc. can hardly be called a form of libertarianism. So what should we say about the psychological discourse that has this libertarian logic? There are two options, either attack or retreat.

To attack is to deny closure under physics for the behaviour of human bodies. This, in effect, commits one to some form of dualist interactionism, which fashion demands that one should avoid.

Retreat consists in admitting that the non-deterministic nature of psychological discourse does not reflect anything non-deterministic about the conduct of our bodies and that, therefore, it does not reflect the nature of any actual dynamic process. It is rather like Dennett's *intentional stance*: it is just a certain way of describing, for certain practical purposes, processes the outcome of which is decided on a wholly sub-personal level. If this is correct, then a realist libertarianism is inconsistent with physical closure. I have argued elsewhere that this Dennettian approach is self-refuting: it makes human psychology a matter of interpretation whilst denying that there are any real interpreters (Robinson 2010). But that takes us on to different issues. The relevant message is that one cannot reconcile a real libertarianism with compatibilism, as Huoranszki claims to do.[1]

[1] I am grateful for the comments of participants in the discussion at the conference on Huo's book, and for those of an anonymous reader, which greatly improved (I hope) a hurried initial draft.

REFERENCES

Frankfurt, Harry 1969. Alternative possibilities and moral responsibility. *Journal of Philosophy* 66, 829–839. [Reprinted in his *The Importance of what we Care About*, 1–10. Cambridge: Cambridge University Press, 1988.]

Huoranszki, Ferenc 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Lehrer, Keith 1966. An empirical disproof of determinism. In Lehrer, Keith (ed.), *Freedom and Determinism*, 175–202. New York: Random House.

Lehrer, Keith 1968/1982. Cans without Ifs. In Gary Watson (ed.), *Free Will*. Oxford: Oxford University Press, 41–45.

Martin, Charles Burton 1994. Dispositions and conditionals. *The Philosophical Quarterly* 44, 1–8.

Moore, G. E. 1912. *Ethics*. New York: Henry Holt.

Robinson, Howard 2010. Quality, thought and consciousness. P. Basile – J. Kiverstein – P. Phemester (eds.), *The Metaphysics of Consciousness*. *Royal Institute of Philosophy Supplement* 67. Cambridge: Cambridge University Press, 203–216.

van Inwagen, Peter 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

Judit Szalai

Rationality

In my comments, I will first lay out the argument of Chapter Seven, then pick out one particular theme within it: that of pathologically compulsive actions. I will outline Huoranszki's (2011a) understanding of compulsion, and point out some of its advantages over other philosophical accounts of the phenomenon. Finally, I will present part of my own view of obsessive-compulsive disorder, which in certain elements converges with Huoranszki's conception, while taking issue with others.

The main claim of the chapter entitled 'Rationality' is that free will and responsibility should not be understood in terms of agents' capacity for rational self-control. More specifically, the theses defended are as follows. First, an agent can be responsible for an action that he does not perform for a reason. Second, agents can be responsible even if they fail to exercise the capacity of *rational* self-control in specific contexts. Third, an agent can apprehend a reason, and her action can be caused by that apprehension, without her exercising rational control over the action.

How can an agent be responsible for something she does not do for a reason? First of all, in what sorts of cases do agents act without a reason? Apart from crossing our arms or scratching our heads voluntarily but for no reason, we may also perform more complex acts without a reason. A more straightforward case is when the action is done gratuitously, "for its own sake". An example is stealing a chocolate bar from a store, not for any reason (such as being unable to pay for it or having a particularly urgent desire for it). The agent is nonetheless thought to be responsible for what she does, for she could have avoided doing it had she chosen to. Another example is one related to the individuation of actions: we have a reason to buy *a* bottle of milk, but we do not have a reason to buy *that particular* bottle. Still, taking that particular one is an intentional and voluntary act, one for which the agent is responsible.

It could be argued against the possibility of action without a reason that, if a person intentionally performs an action, then there is at least this much that can be cited in the way of a reason: he wanted to *x*, wanted to perform the action.

Thus, any intentional action is done for a reason and must involve the exertion of the capacity of rational self-control. Huoranszki's answer is that if wanting to perform an action would count as a reason for it, then we could create reasons for our actions just by deciding that we shall perform them, and the agent's intention to perform the action would automatically rationalize the action.

Another reason why this idea should be rejected is provided by an analysis of compulsive action. The two premises are that compulsive agents are not responsible for their compulsive acts and that they perform them without having a reason to do so. If we do not exempt certain wants from making for reasons (a possibility addressed below), then we have to say that the compulsive agent did not have the want, either, otherwise he would have acted for a reason, which is assumed not to be the case.

Alternatively, it might be claimed that pathologically compulsive agents do something they do not want to do. What may give this view a semblance of plausibility, Huoranszki argues, is that in non-pathological cases of being compelled, we may do something that we don't want to do—induced, for instance, by a threat. But these cases are not relevantly similar to pathological compulsion, since here the agent does retain his responsibility. Such cases cannot support the claim that wants are reasons anyway, because in these (non-pathological) cases, the agent acts *for* a reason, but *against* his want. Therefore, such cases cannot ground the claim that wants are reasons.

Finally, concerning the idea that some wants are reasons while others are not: if we understand the issue in this way, then it would be responsibility for actions that would ground the classification of wants as reasons for those actions. Thus, rational capacities would not explain responsibility; rather, responsibility would explain what counts as an exertion of some rational capacity.

The above arguments, as I have mentioned, rely on the view that the compulsive agent does not have a reason for the compulsive act and is not responsible for performing it. I will offer an alternative account later on. Now I will move on to the second cluster of arguments of the chapter: that exercising the capacity of *rational* self-control is not necessary for responsibility.

Weak-willed agents do not exercise rational self-control: they act in an irrational manner. While their actions do not depend, or at least do not depend appropriately, on their reasons, but rather go against them, weak-willed agents are still responsible for what they do or fail to do. Thus, lack of rational self-control does not imply lack of responsibility.

Now why are we responsible for our akratic behavior? Because our will is free even when it is weak: the weak-willed agent *could* have acted upon her (best) reasons in the sense that she would have done so had she chosen to. Some people deny this and argue that akratic agents do not act of their own free will, because they do not choose their actions: this is Gary Watson's view in 'Skepticism about the Weakness of Will' (Watson 1977). Akratic agents cannot make

a choice about what to do; however, they are responsible for their action and omission because of their culpable lack of self-control. Huoranszki counters this by observing that it is not the actual exercise of self-control that should be made the condition of responsibility, but having the capacity in general—and akratic agents do not lose that. (In contrast, compulsive agents are thought to be incapable of making the relevant choices.) Also: why should we believe that akratic agents cannot make a choice about whether or not to perform an akratic action? It seems, for instance, that their behavior is, or at least can be, sensitive to positive incentives. A weak-willed person who intends to quit smoking, but cannot resist when a cigarette is offered to her, would most probably be able to give up were she offered a very large sum of money to resist this temptation. Thus, her ability to make a choice is revealed by the sensitivity of her behavior to certain incentives. (The pathologically compulsive agent's behavior, in contrast, is considered to be unaltered by new positive incentives.)

Another possible objection to the idea that akratic actions would prove that free will as a condition of responsibility is not the same as the ability of rational self-control is the following. Even if akratic agents act against their best judgment, they may not act against their reasons in general. Their action depends on a reason they have, if not the "best one." Huoranszki's reply to this issue is based on a rejection of Watson's and Davidson's (Davidson 1970) common premises. Davidson's understanding of the weakness of will is that agents act against their all-things-considered judgments (e.g., that they should come off a certain substance) but act upon their "unconditional" judgments concerning the value of their action.

The first answer to Davidson's suggestion is that the akratic agent's weakness is to be attributed not to some cognitive mistake but to a motivational or volitional one. Second, if she cannot form her all-things-considered judgment, she does what she sees most reason to do. Her reason and motive do not come apart. But then the problem is not that her will is weak.

That the weak-willed agent acts *sub ratione boni* is an assumption Watson and Davidson share. They both maintain that free will is a rational capacity. Huoranszki denies that the ability to make choices is an ability to exercise some kind of rational control.

The third main topic of the chapter is action as a result of an apprehension of reasons without exercising rational control. Huoranszki—rather helpfully, I believe—distinguishes between two different forms of weakness of the will. There are two sorts of control that the weak-willed agent may fail to exercise. In one sense, she does not control her choice by her reasons, by not choosing what she has most reason to do. This is a failure to exercise a *rational* capacity. For instance, someone might be aware of the fact that it is her duty to help an injured friend and does not lack the ability to help, but the sight of blood keeps her from doing so. What she fails to do is choose and intend to perform the ac-

tion despite her strong reason in favor of it. The other sort of failure is failing to execute the action planned and in this sense intended. This is a failure to exercise an *executive* capacity. A gambler's intention to keep away from casinos, for instance, might be thwarted by their lure on certain occasions. In these cases, the gambler fails to act on an intention he formed, rather than failed to form an intention based on his reasons.

Finally, an agent who lacks the ability of rational control may nevertheless perform a rational action. If a person is incapable of abiding by her self-destructive plan she has every intention of carrying out, she is not acting in an irrational manner, even though her actions are not controlled by her reasons.

Agents who lack the capacity of *rational self-control* do retain some capacity to control their actions—otherwise we could not hold them responsible. They have the ability to do otherwise: they would have done otherwise had they chosen to and retained the ability to make the relevant choice.

COMPULSIVE ACTION[1]

Huoranszki shares an understanding of the phenomenon of compulsion with most philosophers writing on the topic, while the explanation he offers is an innovative one. The standard view in the philosophical literature is that the compulsive agent cannot make a choice regarding her action. A couple of recent statements of this view will suffice.

> We understand that a person suffering from obsessive-compulsive disorder, spending all day washing his hands and checking dozens of times that he remembered to lock the front door, cannot be thought of as having free will. His actions are mechanically dictated by stereotyped scripts, from which he cannot escape. Thus, obsessive-compulsive disorder is a malady of free will, because it prevents normal strategic planning and meta-control of behavior from overcoming compulsions. (Levy 2003: 214.)

> OCD patients often indicate that they wish to be rid of hand-washing or footstep counting behavior, but cannot stop. Pharmacological interventions, such as Prozac, may enable the subject to have what we would all regard as normal, free choice about whether or not to wash his hands. (Churchland 2002: 208; emphasis added.)

[1] This section is in large part a summary of my paper 'Agency and Mental States in Obsessive-Compulsive Disorder' (in manuscript).

The received explanation for the compulsive's inability to make a choice regarding her behavior attributes it to some volitional deficiency. Before assessing the plausibility of this type of explanation, a preliminary distinction needs to be mentioned, one Huoranszki brings out in another of his writings, 'Weakness and Compulsion: The Essential Difference' (Huoranszki 2011b).

Writing about compulsion, most philosophers have obsessive-compulsive disorder in mind. This pathology is often not clearly distinguished from compulsion in another sense. We talk about "compulsive" eating, drinking and gambling, the mechanism of which is significantly different from that of OCD. Most importantly, the "compulsive" eater or gambler is attracted, at least initially, to the action itself or some aspect of it, while the OCD patient has no intrinsic interest in the compulsive act. The latter are apparently not performed, not even initially, for their hedonistic value.

In fact, it is hard to see how *any* value could be attached to the compulsive act itself, intrinsic or relational. Given the apparent pointlessness of, say, making idiosyncratic movements with one's hands, repeatedly emitting inarticulate sounds, or engaging in what under normal circumstances are goal-directed actions—such as washing hands or checking the gas stove—with unnecessary frequency, we can proceed in two different directions. One possibility is to take this pointlessness at face value and assume that, had the agent been capable of not performing the act, she would have avoided performing it. Conceived in this way, OCD would be similar to conditions with complete loss of action control, such as the anarchic hand syndrome (see below). The other possibility is to look for some property of the act that confers value or significance on it for the agent without readily manifesting itself from a third-person perspective, with the hypothesis that OCD-related acts might be voluntarily performed. My own account will take the second path. Before presenting it, I will sketch two versions of the volitional deficiency view and what I take to be advantages of Huoranszki's account over them.

Watson (1977) attempts to differentiate between compulsion and the weakness of the will, which also seems to involve a desire the agent is unable to resist. Watson suggests that the difference lies in the kind of irresistibility the respective desires exhibit: while persons are weak-willed in relation to desires a normal adult in our society can be expected to resist, "no degree of training and discipline would have enabled him or her to resist" a compulsive desire (Watson 1977: 332). This characterization, Watson admits, makes the very existence of compulsive desires doubtful.

Zaragosa's more recent account uses the idea of "ego-depletion", of temporarily losing the capacity of self-control after its overly straining exertion. "A compulsive is subjected to a nearly continuous stream of impulses to perform a specified behavior, which eventually overworks the will, producing a form of psychological stress" (Zaragosa 2006: 262). What explains OCD for Zaragosa is a

failure of inhibitory mechanisms that would prevent the agent from performing the compulsive act.

Both accounts seem to overlook the distinction between weakness and compulsion pointed out by Huoranszki and regard the agent as drawn to the compulsive act, with the difference that Zaragosa's agent struggles to retain control. The selectivity of OCD behavior—the fact that it only extends to specific activities, e.g. hand washing for some, cleaning for others—is not successfully explained by either account.

Though concurring with the idea that compulsives do not choose or control their behavior and that compulsive acts are performed without a reason, Huoranszki understands the compulsive's failure as ultimately cognitive. The pathologically compulsive agent cannot make a choice for the reason that she does not believe that she is able to perform (or not perform) certain kinds of action (even if, in fact, she is). This understanding is supported by self-reports of OCD patients who claim that they *had to* perform the action and could not have done otherwise.

Volitional deficiency accounts of OCD do not seem to do justice to the causal role of beliefs in this condition. In Huoranszki's view, however, cognitive factors are just as important as orectic ones. Compulsion cannot be understood in isolation from the beliefs that ground and maintain them. In the following, I will offer an alternative account of the cognitive background of OCD, one that is not wholly incompatible with the one offered by Huoranszki.

Compulsive acts are normally preceded by obsessive thoughts.[2] The latter come unbidden, often intrusively and thus are not experienced by the subject as "conjured up" by herself. The contents of such thoughts are typically threatening events: traffic accidents, illnesses, significant losses, even grossly inappropriate public behavior on the subject's part. An OCD patient might repeatedly think, for instance, that the next time he crosses the street he will be run over by a car. He has no way of making sure, through ordinary means, that this will not happen. As obsessive thoughts can be overwhelming and burdensome, their subject tries to "neutralize" them through actions he believes or hopes will ward off the threat. He gives himself an assignment, much like a vow, to carry out repeated acts meant to influence the future state of the world in a quasi-superstitious manner. The acts are voluntary,[3] although the manner of agency involved is somewhat out of the ordinary (see below).

[2] OCD may well be a heterogeneous phenomenon and the view I am trying to advance here is certainly not applicable to all of its forms. There can be, e.g., obsessions without compulsions (the "pure obsessional" type), and "pure" compulsions are also possible, although infrequent.

[3] The compulsive acts involved in OCD are explicitly stated to be voluntary by a number of psychologists. I will give a few examples. "The feature of compulsion that needs to be stressed here is that a compulsion is actively brought about by the patient: he is not happy

It might be objected that such obsessive thoughts cannot reach the status of beliefs, especially since many patients acknowledge them to be "unreasonable". This may be interpreted, however, as merely an admission that the average cognizer would not share the obsessive belief—which does not necessarily prove it false. The OCD patient might believe herself to be in a privileged epistemic position, with access to special evidence.

The OCD patient carries out the compulsive act in order to avert danger by a ritual. Such rituals are "rigid or stereotyped acts according to idiosyncratically elaborated rules" (DSM-IV 1994: 418). The content of the ritual task is not perceived by the subject as arbitrary or "made up" by herself. She may be uncomfortable with the task; its performance may feel like an imposition. This is compatible with the subject sincerely claiming that she "has no choice" but to carry out the act, meaning that what she takes to be her reasons in favour of doing so significantly outweigh the reasons against (those deriving from the unpleasantness, embarrassment, etc., of the act): she has no choice because she cannot run the risk of getting hit by a car or losing a close relative.

Thus, pathologically compulsive agents do apprehend their actions as done for reasons, even though we would evaluate those reasons as bad or peculiar. Evolutionary and developmental explanations of OCD equally tend to emphasize the fact that the compulsive agent perceives the compulsive act as a means of control (or as self-punishment). If this is right, the irrationality of compulsive behavior does not seem to come from the inability to resist a desire, but partly from a bizarre means-end reasoning motivated by other mental states (primarily fears) and partly from the irrationality of the obsessive thought itself.

What evidence can be adduced in support of this account, specifically the claim that OCD-related actions are voluntary and largely controlled by the agent? I will mention three considerations. First, OCD patients feel responsible for the anticipated outcome of their action or non-action: if the house gets robbed, it is because they had not checked on the lock enough times; if they are harmed in a traffic accident, it is because they did not perform the appropriate ritual to ward it off (Clark 2004: 94ff.; Shapiro and Stewart 2011). The fact that the patient has the sense that she influences such outcomes in the world is indirect evidence for her feeling responsible *for the compulsive act* as well. If the act was not in her power to initiate, the compulsive person could not blame herself if the negative event occurred. This argument is admittedly relatively weak,

about doing it, but it is essentially his voluntary action … and not an automatic behavior. Thus it is different from tics and muscle spasms that are found in some people, especially children, which are essentially involuntary motor responses. These are not actively, deliberately produced by the patient … Unlike compulsions, they are not purposeful." (De Silva and Rachman 2004: 9); "Although obsessions are felt as involuntary and are strongly linked to anxiety and distress, ritualizing (both overt and covert), is voluntary, controlled behaviour" (Arden and Linford 2008: 186); DSM-IV (1994: 120, 418).

though, for, as we will see, OCD phenomenology is not a sufficiently reliable indicator of actual agency.

Second, OCD-related actions take special effort. OCD patients are reported to have a "perfectionistic" tendency: they keep having the sense that they do not get the action "just right" and consequently feel a need to repeat it. Imperfectly executed actions may not ward off the event feared. The necessary number of repetitions is often determined along the way, according to when the performance "feels right". Thus, the OCD patient is capable of terminating her action at what appears to her the appropriate point.

Third, there is clear evidence that OCD patients can refrain from compulsive acts. That the OCD patient is in control of her action is most directly[4] shown by the way in which exposure and response prevention, the standard cognitive therapy treatment for OCD, is implemented. Physical prevention is no longer recommended practice (Foa and Franklin 2002: 100); rather, the patient is made voluntarily to refrain from performing her rituals.[5] The possibility of voluntary refraining shows the voluntariness of the action itself. When patients are exposed to the subject of their obsession and would otherwise perform the obsessive act, they can abstain from doing so, for therapeutic purposes. (The point of response prevention is to show that not performing the action does not lead to the feared event's occurrence.)

The second and third considerations suggest that the patient has actual, if perhaps somewhat limited (see below) control over her act. The reason why the claims made here are partly compatible with Huoranszki's account is that self-reports do not seem to be a reliable guide to OCD patients' agency and control over their actions. 30–40% of patients have poor insight into their condition; many of them, for instance, mistakenly deny that their compulsive acts are driven by obsessions (Kalra and Swedo 2009: 737–738). It seems reasonable to assume that the self-assessment of those who claim they are "unable to do otherwise" may also be mistaken. The possibility that the patient could have acted otherwise, had it not been for the sense that she was unable to, seems to be left open. The sense of a lack of control may be compatible with actual control.

While the acts OCD is manifested in are here argued to be voluntary, initiated and terminated by the agent, what can be labeled the "manner" or "style" of agency is somewhat out of the ordinary (Balconi 2010: 136ff). As was men-

[4] Compulsive agents also tend to refrain from compulsive acts in public—aware as they remain of how disturbing others would find their behavior—and have a tendency to confine their compulsive behavior to their home environment. (Many people have compulsions and we very rarely see compulsive acts.) Reportedly, it is also possible to replace a compulsive act with a socially less intrusive one—for instance, one that does not involve emitting noises. This also seems to indicate that compulsive agents do not lose their control capacities.

[5] "To maximize improvement, the patient needs to voluntarily refrain from ritualizing while engaging in programmatic exposure exercises" (Foa and Franklin 2002: 100).

tioned, OCD patients tend to have a sense of imperfectness, of "not getting it right" in executing their action plan. They also tend not to focus on that action plan, a phenomenon that may have to do with the repetitions involved. The action is likely to become highly habitualized, and part of its execution is often automatic. This might also contribute to the sense of being "unable to do otherwise" in some of the self-reported cases.

If my interpretation is along the right lines, obsessive-compulsive disorder is not as interesting to the philosophy of action and free will as it is generally thought to be (although the low-level agency mentioned in the previous paragraph might be). There are other mental pathologies, however, which seem to be more pertinent to the issue of the loss of control and a sense thereof.

One such disorder is the so-called "anarchic hand" syndrome. "Patients with Anarchic Hand syndrome sometimes find one of their hands performing complex, apparently goal-directed movements they are unable to suppress (except by using their 'good' hand)" (Eilan and Roessler 2003: 2).[6] Sometimes the anarchic hand interferes with what the other hand does; at other times it does something that has nothing to do with the subject's intentions. The act is outside the patient's control and is experienced as such. The answer to the question of whether these are actions at all, even if their subjects are not in control of, or responsible for, them, seems to be "yes", for "…the activities of the anarchic hand are skillfully controlled: they are not pure reflexes, but clearly devoted to a particular goal, and relative to the goal, well-executed" (Eilan and Roessler 2003: 2).

Action control appears to break down in other pathological conditions as well. For instance, in "utilization behavior", a certain type of stimulus seems to "force" the agent to act in a certain way; in "perseveration", they manifest an inability to stop a sequence of actions. In these cases it does seem that the agent has no choice, could not have done otherwise, and therefore is not free with regard to the action. There are also non-pathological failures of control, such as the 'Double Capture Error', in the case of which "attention is captured by some internal preoccupation, allowing the action to be captured by a stimulus associated with a strong habit". For instance, we find ourselves proceeding in the same direction as we regularly do, despite our intention to do otherwise this time. (This seems similar to the failure of an execution capacity, as described in Huoranszki's chapter.) Consideration of these phenomena may be relevant to the free will issue in more ways than just whether the agent could have done otherwise. Pathological conditions like the anarchic hand syndrome and the everyday experience of double capture may indicate the possibility of there being different control mechanisms, or control mechanisms at different levels, rather than one unified, central form of control.

[6] The last two passages rely on examples from Elian and Roessler (2003).

REFERENCES

American Psychiatric Association 1994. *Diagnostic and Statistical Manual of Mental Disorders*. Fourth Edition. Washington, DC: American Psychiatric Association.

Arden, John B. and Linford, Lloyd 2008. *Brain-Based Therapy with Adults: Evidence-based treatment for everyday practice*. London: John Wiley & Sons.

Balconi, Michela 2010. Disruption of the Sense of Agency: From perception to self-knowledge. In Balconi, Michela (ed.), *Neuropsychology of the Sense of Agency*. Milan: Springer, 125–144.

Churchland, Patricia 2002. *Brain-Wise: Studies in neurophilosophy*. Cambridge: The MIT Press.

Clark, David 2004. *Cognitive-Behavioral Therapy for OCD*. New York: Guilford Press.

Davidson, Donald 1970/1980. How Weakness of the Will is Possible? In his *Essays on Actions and Events*, 21–42. Oxford: Clarendon Press.

De Silva, Padmal and Rachman, Stanley 2004. *Obsessive-Compulsive Disorder: The facts*. New York: Oxford University Press.

Eilan, Naomi and Roessler, Johannes 2003. *Agency and Self-Awareness: Issues in philosophy and psychology*. New York: Oxford University Press.

Foa, Edna B. and Franklin, Martin E. 2002. Psychotherapies for Obsessive-Compulsive Disorder: A Review. In Maj, Mario et al. (eds.), *Obsessive-Compulsive Disorder*. London: John Wiley & Sons, 93–115.

Huoranszki, Ferenc 2011a. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Huoranszki, Ferenc 2011b. Weakness and Compulsion: The Essential Difference. *Philosophical Explorations* 14, 81–97.

Kalra, Simran K. and Swedo, Susan E. 2009. Children With Obsessive-Compulsive Disorder: Are They Just "Little Adults"? *Journal of Clinical Investigation* 119, 737–746.

Levy, Daniel 2003. Neural holism and free will. *Philosophical Psychology* 16, 205–228.

Rector, Neil A. et al. 2001. *Obsessive-Compulsive Disorder: An information guide*. Toronto: Centre for Addiction and Mental Health.

Shapiro, Leslie J. and Stewart, Evelyn S. 2011. Pathological Guilt: A Persistent Yet Overlooked Treatment Factor In Obsessive-Compulsive Disorder. *Annals of Clinical Psychiatry* 23, 63–70.

Watson, Gary 1977. Skepticism about the Weakness of Will. *Philosophical Review* 86, 316–339.

Zaragosa, Kevin 2006. What Happens When Someone Acts Compulsively? *Philosophical Studies* 131, 251–268.

Dániel Corsano

Spontaneity and Self-Determination

The title of the eighth chapter of *Freedom of the Will* by Ferenc Huoranszki is "Spontaneity", which is the ability to determine one's action in a particular situation by choice (Huoranszki 2011: 141). But the topic is really the correct interpretation of self-determination. Huoranszki argues that the two concepts should be separated, as demonstrated by negligence, in the case of which we are unable to do what we ought to (and hence lack spontaneity), but are nevertheless responsible for this inability. He also differentiates his interpretation of self-determination from that of his incompatibilist adversaries, which requires the agent to be able to determine herself in the sense of being responsible for her current character, motives and reasons.

According to some philosophers, self-determination is utterly impossible, since it would require us to be *causa sui*.[1] Others think that it is only possible in physically indeterministic worlds, since the kind of 'ultimate' or 'true' responsibility that is required for determination of the self is only possible in these.[2] Huoranszki claims, however, that neither of these two claims is correct. He argues that if we follow through with this strong interpretation of self-determination, we will find it to be impossible in indeterministic worlds as well, and therefore the concept should be interpreted in a more permissive way in order to be a condition of responsibility. He proceeds to provide his own interpretation, which is evidently possible in both deterministic and indeterministic worlds, and measures it against two libertarian alternatives, finding it superior to both of them. In the following paper I am going to explicate Huoranszki's views expressed in this chapter of his book, following his line of thought. Then I will present my objections to some of the views discussed or endorsed by the author.

[1] See Strawson (1994) for such an argument.
[2] See Kane (1996).

8.1 REASONS, ABILITIES AND SPONTANEITY

Huoranszki agrees with his adversaries that self-determination is a condition of moral responsibility and thus important for the analysis of free will. An agent can only be morally responsible if she is able to determine herself; the question is what exactly is meant by this. Huoranszki introduces his interpretation by saying that by something being self-determined, he only means that the future state of the thing is only determined by its inner workings. If a clock works properly, its future state can be explained only by its current internal states and powers, and there is nothing metaphysically problematic in saying that an event identified as a behavior of a substance is caused by an internal change in the substance. Of course, the situation is more complicated in the case of moral agents, but Huoranszki wants to apply the same general principle.

If our motives and character causally determined our actions, we would only be responsible if we could determine them; hence the strong interpretation of self-determination. There is no sense in debating the fact that in a deterministic world 'being determined by something' is transitive (if B is determined by A and C is determined by B, then C is determined by A), and the consequence argument[3] tries to use exactly this fact to make it seem unfeasible for an agent to determine herself in any freedom-relevant way in deterministic worlds. The previous states of the world and the laws of physics determine its current state, including our reasons and motives, so if our actions are determined by our reasons and motives, they are ultimately determined by external causes.

But if Huoranszki is right that self-determination in this sense is impossible in indeterministic worlds as well, we have to choose between two conclusions: we could claim that the strong interpretation of self-determination is a condition of moral responsibility, which is in turn impossible. Conversely, we could argue that the strong interpretation is incorrect, because, according to it, moral responsibility is impossible. This choice actually consists in determining the relative strengths of two incompatible assumptions: that ultimate responsibility is necessary for self-determination, and that moral responsibility is possible.

Huoranszki chooses the latter path, of course, and proceeds to delineate his interpretation of self-determination as a condition of moral responsibility. According to this, the kind of control required for moral responsibility is grounded in our ability to perform an actually unperformed action, in the dispositional sense employed by his conditional analysis of free will as the central thesis of the book. He defines this control in the following way.

[3] For the argument, see van Inwagen (1983); for Huoranszki's treatment of it, see Chapter 2 in Huoranszki (2011).

> What responsibility requires is that we can control our actions in the sense that we could have done otherwise. And we could have done otherwise in the responsibility-relevant sense provided we would have done otherwise if we had chosen so and retained our ability to choose and to perform the relevant sort of action. (Huoranszki 2011: 143.)

The detailed explication of Huoranszki's conditional analysis belongs to another chapter. As for his concept of self-determination as a condition of responsibility, the first thing he needs to do with it is to differentiate it from spontaneity. The latter only applies to current choices, but there are many cases where it seems natural to say that someone is responsible for not having a choice in a particular situation. The author quotes an example originally described by A. S. Kaufman:

> Suppose that a lifeguard who has lied about her qualifications is unable to swim. Assume now that a child drowns whose life it was the lifeguard's duty to save. We would certainly hold the lifeguard responsible and yet, being unable to swim, she could not have saved the child's life. (Lamb 1993: 525.)

In this case most people would say that the lifeguard is responsible, despite the fact that she didn't have the choice of saving or not saving the child. Huoranszki claims that this is because she is responsible for having got herself in this situation in the first place. She is responsible not for her inability to save the child—not being able to swim could just be a genetic disability—but for making a prior choice, the foreseeable consequence of which is the current situation, in which she does not have a choice.

Self-determination as conceived by Huoranszki rests on spontaneity in the sense of our actions depending on our choices, and it also requires the ability to perceive moral reasons. But a third condition needs to be added as well, so we can accommodate the intuition of being morally responsible for negligence. He claims that self-determination also requires "… that we could have done something, which we have actually failed to do, and *the foreseeable consequence* of which is our present inability", or "… that we could have avoided being in the circumstances in which our lack of ability cannot exempt us." (Huoranszki 2011: 145.)

Of course, both articulations of this condition are rather vague, since they place the burden of actually foreseeing the possible outcomes of actions on the agent. I would say that the lifeguard in the above example would be responsible even if she did not realize that she would need to be able to swim in order to be an effective lifeguard, similarly to someone who would cause an accident by driving down the wrong lane of a highway and thinking that no danger would come of it. I imagine that most cases of negligence are the consequence of someone not foreseeing the consequences of her actions, even if they seem quite

foreseeable to someone else. But I would not say that for example somebody was responsible for going down with an airplane and hitting someone's house, even though such accidents are obviously foreseeable consequences of travelling by airplane, and some of the passengers—who were afraid of flying—might even have actively visualized it before takeoff.

A consequence of an action can only be sensibly deemed foreseeable from the perspective of a particular agent. And the above examples show that this concept does not afford an exact condition of moral responsibility for negligence in this sense either. No human is capable of foreseeing all the consequences of every possible action (not even those that follow logically from the known facts), nor are we responsible for every consequence of our actions that we do foresee. To claim that we are only responsible for the foreseeable consequences of our actions might seem intuitively correct, because intuitively we more or less agree on which cases to count as negligence (at least with extreme examples), but it does not clarify the exact conditions at work.

Huoranszki claims that we can determine our reasons and abilities without ultimate responsibility. The case of the lifeguard illustrates that sometimes the current situation which provides our reasons and abilities is a foreseeable consequence of our prior choices, and is thus determined by ourselves. This kind of self-determination is definitely possible, even in deterministic worlds, and it is also a necessary condition of moral responsibility. Thus the relationship between self-determination (in this sense), moral responsibility and freedom of will is as follows. Self-determination is a necessary condition of moral responsibility and freedom of will is a necessary condition of self-determination. We are only responsible for our present inabilities "if they are the *foreseeable consequences* of our prior choices, and thus could have been avoided." (Huoranszki 2011: 146–147.)

8.2 ULTIMATE RESPONSIBILITY

After delineating his concept of self-determination, Huoranszki presents two alternative libertarian interpretations. According to these, the kind of self-determination required for moral responsibility consists in being able to control our motives and character, which is impossible in a deterministic world, but possible in an indeterministic one. So moral responsibility implies the falsity of physical determinism. If the libertarian wants this view to seem plausible, she must explain how physically undetermined events can make ultimate responsibility for actions possible. The falsity of physical determinism implies the nomological possibility of an actually non-occurrent event occurring at any time t, and thus a particular situation can have more than one possible outcome. The real difficulty lies in placing this nomological contingency at the fundamental physical level somewhere in the causal chain leading to an action, where it could sensibly

guarantee the possibility of agential control. But Huoranszki (2011: 147) argues that there is no such place for this contingency.

If physical indeterminacy is to be relevant for free will, it should take place before the intention for an action was formed, since the indeterminacy of the outcome of an already willful action should not concern free will. At this point Huoranszki quotes Dennett (1978: 295), who, though not himself a libertarian, recommends the process of deliberation preceding choice and action as a place for this indeterminism to occur, but this account is also highly problematic. Huoranszki (2011: 150) admits that it is quite possible that the origin of some intellectual achievements might be explained by such an indeterminacy, but not moral responsibility, since when we find someone morally responsible for neglecting an action, we do so based on their inaction, and not if and because they have indeterministically forgotten the action.

Huoranszki concludes that the indeterminism relevant to libertarian ultimate responsibility must occur after the process of deliberation and before an intention is formed. Most libertarians believe that choice is only possible if the formulation of intentions is an indeterminate physical process or event, thus they take practical deliberation not to determine the preferences resulting from it. But if choices are indeterminate only in this way, this means that the actions of agents are actually not controlled by themselves, but by pure chance. To avoid this contradiction between preferences and chance, some libertarians argue that agents can only choose in some exceptional situations. This comes down to two lines of thought: one can either go the way of plural rationality, as Kane (1996) did, and hold that our will is only free if our rational deliberation does not determine which actions we judge to be the best; or one can choose what Huoranszki calls the indifference strategy, holding that our will is only free if our motives do not determine what we do. The essential difference between these two accounts is that the former specifies rational indifference and the latter psychological indifference. In the remainder of the chapter, Huoranszki argues that neither of these views captures the sense in which self-determination is a condition of free will and moral responsibility.

8.3 MOTIVES, CHOICES AND RESTRICTIVISM

According to Peter van Inwagen (1989), there are three kinds of situation in which we can do otherwise: first, so called Buridan's Ass cases, where there is no qualitative difference among the possible alternatives; second, when our duty is in conflict with our inclinations; third, when we have to choose between incommensurable values. On all other occasions we cannot do otherwise, which

implies that in most cases, even when we are morally responsible for our actions, we do not act out of our free will. Hence this view is called restrictivism.

Huoranszki examines the indifference strategy first. Though the three abovementioned cases differ significantly, he calls all of them indifference, and the advocation of the view that we can only act freely in such cases the indifference strategy. By indifference he means that there is no strong ordering of the possible outcomes (regardless of this resulting from lack of or equal motivations), which results in agents not needing to act against their own preferences. He finds it implausible that we do not act out of our free will when we act according to our preference. In fact he sees this as a *reductio ad absurdum* of the indifference strategy and maintains the opposite view, that "only those who *can* choose and act according to their preference (even if they may not actually do) act of their own free will." (Huoranszki 2011: 152.) Accordingly he shifts the burden of proof, claiming that "unless the restrictivist has a strong argument to the contrary, we have no grounds to think that our will is free only if and when we are indifferent." (Huoranszki 2011: 152.)

The proposed argument for restrictivism comes from van Inwagen (1989: 406), who uses an example from Dennett, in which he is asked to torture someone in return for a small sum. Dennett claims that he—in his present state—is simply unable to do this or indeed anything else he finds morally reprehensible. Van Inwagen argues that this means that the ability to do otherwise is unnecessary for moral responsibility, since we would obviously take Dennett's inability to comply with the request as proof of his morality, with which it would conflict to do so.

Huoranszki finds this argument inconclusive. It would only be sound if "from the fact that *S would never choose to do A* we could infer that *S cannot make a choice about whether or not to perform A*" (Huoranszki 2011: 154), which he thinks is fallacious, since not being motivated to do something and not being able to do it is simply not the same thing. So it comes down to how we think about abilities: we could either agree with Huoranszki that there are unactualized abilities, or disagree with him, as incompatibilists usually do. Notice that this is the same difference in presuppositions that leads Huoranszki to reject the consequence argument, and van Inwagen to endorse it. Thus the only thing this line of thought proves is that compatibilists and incompatibilists tend to disagree on this matter, but this we already knew. Huoranszki also mentions an argument from van Inwagen, which uses the No Choice principle.

> Consider an action A (like torturing someone in return for a small sum) which *S* would regard morally indefensible. According to van Inwagen's argument, *S* cannot make a choice about whether or not he finds A indefensible. And he cannot make a choice about whether or not he performs an action that he finds indefensible.

> So he cannot make a choice about whether or not to perform A. And as everyone agrees, if *S* has no choice about whether or not to perform A, then his will is not free. (Huoranszki 2011: 154.)[4]

Huoranszki disagrees with the premise that agents can have no choice about whether or not to perform an action they find indefensible. He thinks that restrictivism is based on the mistaken view that "[m]otives are motives for actions and not choices" (Huoranszki 2011: 155) and offers the following example as a *reductio ad absurdum* of this. There are many situations where we only slightly prefer one choice to another, for example chocolate cake to cheesecake. In these cases Huoranszki would say that he chooses to order a chocolate cake *because* he prefers it slightly, but the restrictivist would have to say that he did not have a choice because he had a preference. If in this situation someone would ask the restrictivist why he ordered the chocolate cake instead of the cheesecake, for which the restaurant is famous, he would have to reply 'because I could not have done otherwise'. Huoranszki finds this pretty much in conflict with how we normally think of preferences and choices, and proposes the use of the expression 'I could not have done otherwise' for when the agent has a pathological aversion to cheese, or when there is no cheesecake on the menu, and other similar cases. I will examine Huoranszki's treatment of pathological aversions later. As for the above argument, apart from showing that the restrictivist is inclined to use some rather uncommon language,[5] it seems to come down once again to the question of unactualized abilities, and thus merely restates the well-known difference between his and his adversaries' views on them.

Huoranszki closes the discussion of the indifference strategy by examining the notion of self-determination that it implies. According to the advocates of this strategy, responsibility is rooted in the undetermined choices that occur when we are indifferent towards multiple outcomes. We are not able to do otherwise when we have a preference, but we can still be responsible for our actions in these cases, because we have shaped our present selves (and motives) by our prior undetermined (indifferent) actions. So, in most cases, we are not directly responsible for our actions, but for being as we became because of our prior undetermined choices.[6] Thus the ultimate ground for our present moral responsibility is our prior indifference, which Huoranszki finds absurd. He thinks

[4] For van Inwagen's original argument see Inwagen (1989: 408–409).

[5] This is also true of Huoranszki in other cases. Think of Dennett's example, in which he similarly tricks the compatibilist into using the word 'able' in an uncommon way. The premise is that Dennett is unable to do anything he finds morally reprehensible, but the compatibilist cannot allow this to be true for his sense of 'being able to do something' (which allows for unactualized abilities, unlike the one Dennett uses in the premise), and is thus led to argue for Dennett being able to do something he is unable to do. Language makes this seem like a contradiction, but in fact the two instances of the word 'able' have different senses.

[6] See van Inwagen (1989: 418–421).

that the indifference strategy is doomed to failure because of this conclusion and because "we are able to choose and perform actions, in the relevant sense, which we are not motivated to do" (Huoranszki 2011: 159), by which, in my understanding, he means that the arguments for restrictivism are not as convincing, as the use of words they commit to is counterintuitive.

8.4 PLURAL RATIONALITY

After deeming what he calls the indifference strategy unsuccessful, Huoranszki moves on to discuss the other possible libertarian strategy: plural rationality. This strategy excludes Buridan's Ass-type cases, leaving only two kinds of choices we can make out of our free will. The main reason for this is that in Buridan's Ass-type cases we do not have any reason to perform a specific action, while in the other two cases it could be said that whatever we do, we do it with a reason. So the ground for freedom of will and moral responsibility is that the relative strengths of conflicting reasons prior to a decision do not determine which reason the agent will choose to act on.

This indeterminacy of preferences can be interpreted in two ways: either by saying that reasons cause actions only probabilistically, or by saying that it is not determined whether or not a reason causes a particular action. The first option is not very plausible: probabilistic causation means that it is not up to us that a specific reason succeeds in causing an action, and thus would be quite problematic as a ground for moral responsibility. Would Dennett really be morally responsible for taking the small sum of money for torture in his thought experiment if the cause of this action would in fact be merely that one of his conflicting reasons lost in a game of chance? Sure, if the history of the world was replayed, the outcome might be different, but only by chance and not because of any effort on my part. So it is understandable that Huoranszki ignores this option and goes on to quote Robert Nozick (1980: 295), who suggested that the indeterminacy relevant to libertarian free will should be interpreted the second way.

According to Nozick, decision consists in choosing one of the competing reasons as the one that will determine the subsequent action. So, at the point of decision, all the possibilities are open, and, instead of choosing an action, we actually choose a reason which causes an action. This account has the advantage of not requiring indifference: we need to have conflicting reasons to exert volitional control, but they do not need to be of equal strength. Thus the plural rationality strategy seems to allow much more free choices. Nozick (1980: 295) also claims that this allows that if the history of the world was replayed until the point of decision, "it could have continued with a different action". I will propose some objections to this at a later point.

Huoranszki quotes an objection to the plural rationality strategy from Richard Double (1991: 204), according to which it is impossible to explain an agent's actions rationally, unless the conflicting sets of reasons have equal prior chances of causing an action. Double assumes that probabilistic explanations require that, given the cause, the chance of the occurrence of an event be higher than the chance of its non-occurrence. If the initially weaker (less probable) reason prevails, we cannot explain the choice rationally. According to Huoranszki (2011: 161–162), if Double is right, the plural rationality strategy collapses into the indifference strategy, but he finds Double's argument unconvincing and offers three counterarguments, of which I will only explicate the first (and in my opinion the strongest), which he illustrates with an example derived from van Fraassen (1980: 105):

> Consider the case when someone contracts paresis as a result of untreated syphilis. On the one hand, paresis is only contracted by those who have untreated tertiary syphilis. On the other hand, only a very low percentage of those who have untreated tertiary syphilis contract paresis. (Huoranszki 2011: 161.)

Examples like this show that probabilistic explanations only require that, given the cause, the chance of the occurrence of an event be higher than in the absence of the cause. After dismantling Double's argument against the plural rationality strategy, Huoranszki sets the grounds for his own, which will be put forward in the next chapter. He concludes that the plural rationality strategy has to accept reasons dependence (while his compatibilist account does not) and—supporting this view with a quote from Kane—that it has to assume some form of restrictivism.[7] Kane claims—following Dennett and van Inwagen—that we are not responsible in any situation where we are unable to do otherwise. Since many of our motivated actions fall into this category, Kane must adhere to the concept of ultimate responsibility, in the sense that when agents' actions are determined by their character and motives, they are responsible to the extent that "they are responsible for being the sort of person they had become by that time" (Kane 1996: 39). Huoranszki sees a fundamental difficulty at this point. While his compatibilist view conceives an action being self-determined as the agent determining what she does, ultimate responsibility means that agents are responsible only if they can determine their selves. Thus, in the case of both libertarian views discussed, responsibility for our actions is rooted in our responsibility for being the sort of persons we are, and it is this view he will argue against in the next chapter.

[7] For the quote see Huoranszki (2011: 163); for the original see Kane (1996: 120).

OBJECTIONS

I will now present my objections to some of the claims discussed in this chapter. The first of these will be to the probabilistic treatment of the relative strengths of reasons, and the second to Huoranszki's interpretation of pathological aversions.

Probability and the strength of reasons: I find the probabilistic treatment of competing reasons employed by Double—and to some degree by Nozick—quite problematic. Double's argument against the plural rationality strategy presupposes that reasons have a relative probability of successfully causing their respective actions prior to one of them actually doing so. On this line of thought, having reason R_A for choosing action A with a 50% probability of being effective, and reason R_B for action B with the same probability, would mean that the two reasons are of equal strength, while probabilities of 70% and 30% would indicate this numerical proportion of their strengths.

To see the difficulties with this view, the concept of probability it employs needs further investigation. There are many rival theories of probability,[8] but in this case we have two distinct ways to conceive it. We can either think of genuine probability, which means that our probability attributions reflect a genuine openness of future (indeterminacy) in the physical world, and if we kept rewinding the history of the universe, then two possible futures, with 50% chances of actualization would occur with roughly the same frequency. Or we could think of subjective probability, in which case the probability is not 'in the world', but in our lack of full physical knowledge of it. This means that saying that an event has a 50% chance of occurring has nothing to do with rewinding the history of the universe—it merely means that, given the known facts, we have equal reason to expect its occurrence and its non-occurrence. Obviously the latter conception is compatible with physical determinism while the former is not.

Now if we examine the treatment of the strength of reasons as the proportion of the probabilities of them becoming effective with this distinction in mind, it is evident that neither kind of probability is able to play this role. Let's say that in Dennett's torture for money example, the amount of money offered amounts to a reason of 10% strength, while the opposing reason is of 90% strength. If we conceive this in terms of genuine probability, it would mean that if we continually kept rewinding the history of the universe, the agent would take the money in roughly one tenth of the instances of the decision. But this is absurd; if the sum of money amounts to a weaker reason than Dennett's dedication to human rights (or whatever stands on the opposing side), then he will never take the money. If I am strongly opposed to torturing anyone for the given amount of money, then no matter how many times the history of the universe is rewound,

[8] For an extensive discussion of the interpretations of probability, see Sklar (1993: 90–127).

I will not do it. Opposition to something can only be manifest as a disposition to decline it, and if I would take the money, I could hardly claim to be opposed to doing so. It can truthfully be said that sugar cubes have a disposition to melt when placed in warm beverages, but if I found one that would not do so, that would mean that it lacks such a disposition. In this sense, Dennett's original interpretation of the example is correct.

My conclusion is that since a reason being the strongest can only mean that it is effective, the prior genuine probability of a stronger reason being effective is always 100%, just as the heavier of two objects is always heavier. That the mass of an object is 25% of another one does not mean that the lighter object is heavier roughly one fifth of the times if we rewind a physically indeterministic universe. Therefore the correct way to conceive the relative strengths of competing reasons is the one employed by Huoranszki (2011: 156): by the amount of change in circumstances needed for a change of preference. Just as the difference between the weights of two objects is understood as the amount that needs to be added to one of them in order for them to become equal.

This conclusion also poses a problem for Nozick's claim that the plural rationality strategy would allow different outcomes for decisions if we rewound the history of the universe, while maintaining the agent's volitional control. If we assign a genuine chance to outcome B of a process of deliberation that has actually resulted in A, we have to assign a genuine probability to it. And this only takes us back to the above problem, but at a meta-decision level: if genuine probability (and not meta-reasons, for example) governs which reason we choose to act on, the process of decision is a coin flip, and not the outcome of a volitional effort. If it is governed by reasons for choosing reasons (meta-reasons), we have to assign a genuine probability of 100% to the strongest reason, or temporarily avoid gauging their strengths by adding a further level of rationality, where we would encounter the same problem. Of course we could also assign subjective probabilities, but once again, this does not result in any genuine openness of future. The only way we could plausibly say that an agent who has made a (plurally rational) choice could have done otherwise is in Huoranszki's dispositional sense, as having an unactualized ability. But if we are willing to accept this as a valid interpretation of the ability to do otherwise, we have no reason not to accept his conditional analysis as well, and abandon the plural rationality strategy and libertarianism.

Pathological aversions: Huoranszki (2011: 155) claims that having a pathological aversion to something—unlike simply being opposed to it—is sufficient to deprive us of our free choice. So, according to Huoranszki, if I prefer chocolate cake to cheesecake, this does not mean that I do not have a choice between them, simply that I am determined to choose chocolate cake, which is compatible with making a choice. But if there was no cheesecake on the menu, or if I had a pathological aversion to it, I would be unable to order it (in the strong

sense that I wouldn't even have an unactualized ability to do so) and thus be deprived of my choice in this matter. But I find this interpretation of pathological aversion highly problematic.

Pathological aversion is a psychological term, used to describe an irrational opposition to something. But notice that this conception of rationality is purely subjective; it only means that the agent's behavior does not conform to the observer's expectations. If someone never chooses cheesecake when offered a choice between a variety of desserts, we can say that she dislikes cheesecake. When someone does not eat the cheesecake when it is the only food available to her, we may speak of an aversion. When someone does this for days, the psychologist would be compelled to call it a pathological aversion. But this only means that the case is extreme and interesting in a medical sense. The meaning of the word pathological in this psychological sense derives from an earlier medical term meaning that a case is unexpected and thus worthy of pathological examination.

Many people whom we might describe as having a pathological aversion to cheese would still eat it after some time if it was the only food available and thus their only chance of survival. And even those who wouldn't, could probably be persuaded by, for example, an evil extradimensional alien threatening to force-feed them five kilograms of gorgonzola and then destroy the whole universe unless they eat one very thin slice of a cheese of their choosing. And even if there was an incredibly weak-willed person who would refuse the slice of cheese in this extreme case, and we would name this as the measure of pathological aversion, it would still only differ in a quantitative and not a qualitative way from strong dislike, and so it would hardly constitute a different case in a metaphysical sense.

Because of the above reasons I conclude that there is no objective demarcation between dislike and pathological aversion; the latter is merely a subjective term applied to an extremely strong dislike, similarly to saying someone is extremely tall instead of just tall. Thus it is fallacious to assume that pathological aversion is metaphysically different from a strong dislike in the sense that the former deprives us of our free choice, while the latter does not. In fact either both of them deprive us of free choice, or neither does, and since I believe Huoranszki's arguments against the former view are conclusive and also essential to his account of free will, I cannot see how he can maintain that pathological aversion deprives us of our free choice.

REFERENCES

Dennett, Daniel 1978. On Giving Libertarians What They Say They Want. In his *Brainstorms*, 286–299. Cambridge, MA: MIT Press.

Double, Richard 1991. *The Non-Reality of Free Will*. Oxford: Oxford University Press.

Huoranszki, Ferenc 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Kane, Robert 1996. *The Significance of Free Will*. Oxford: Oxford University Press.

Lamb, James W. 1993. Evaluative Compatibilism and the Principle of Alternative Possibilities. *Journal of Philosophy* 90, 517–527.

Nozick, Robert 1980. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Sklar, Lawrence 1993. *Physics and Chance: philosophical issues in the foundations of statistical mechanics*. Cambridge: Cambridge University Press.

Strawson, Galen 1994. The Impossibility of Moral Responsibility. *Philosophical Studies* 75, 5–24.

van Fraassen, Bastiaan Cornelis 1980. *The Scientific Image*. Oxford: Clarendon Press.

van Inwagen, Peter 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

van Inwagen, Peter 1989. When is the Will Free? *Philosophical Perspectives* 3, 399–422.

László Bernáth

Self-Forming Acts and Other Miracles*

INTRODUCTION

Ferenc Huoranszki argues for two main claims in the ninth chapter of *Freedom of the Will: A Conditional Analysis* (Huoranszki 2011). First, Huoranszki tries to show that libertarian restrictivism is false because self-determination in the libertarian sense is not necessary for our responsibility, even if motives, reasons or psychological characteristics can influence us relatively strongly to choose one or the other alternative. Second, Huoranszki rejects the so-called manipulation argument.[1] This is an argument for the conclusion that unless physical indeterminism is true, nobody can be morally responsible because our behavior is never independent enough of our environment.

Therefore, according to Huoranszki, neither libertarian self-determination nor physical indeterminism is required for moral responsibility. In my view, Huoranszki's counterarguments do not defeat libertarian restrictivism. However, they can force philosophers who defend this theory to modify or refine it. I analyze Huoranszki's arguments against libertarian self-determination in the first part of my paper. In the second part, I briefly argue for one supervenience argument inspired by a similar objection made earlier (Bács 2012). According to this modified argument, Huoranszki's theory about abilities and responsibility would entail that if physical determinism is true then we are responsible for our ordinary actions only because we are able to do miraculous acts as well. If this objection is correct, Huoranszki's compatibilism is unsuccessful.

* I would like to thank János Tőzsér, Dávid Such and the reviewer who drew my attention to many important details.

[1] Huoranszki discusses Pereboom's (2001: 112–117) version of the argument.

1. OBJECTIONS AGAINST THE THEORY OF SELF-FORMATION AND A POSSIBLE SOLUTION

1.1 Restrictivist libertarianism and reason-dependence

Before I reconstruct Huoranszki's argumentation, I summarize briefly why many libertarians think that we can only be responsible if it is possible to form our own character by choices and actions. Robert Kane called these self-, character- or motive-forming acts (Kane 1996). According to restrictivist libertarianism, *many times* agents are not able to choose and do otherwise, since their actual character and their motives/reasons determine the way they can choose and act in a particular situation. Nevertheless, they are *frequently* morally responsible even in these cases. This is so if their character and their motives/reasons are the consequences of former choices and intentional actions. If an agent has an irresistibly strong motive, and if the strength of one's motive is impossible to derive from former choices, the agent is not morally responsible.

The main idea behind this theory of responsibility is that the impossibility of acting otherwise, at least in many cases, has different source than the physical infeasibility of the alternative action. Rather, the alternative action is impossible because there are not psychologically sufficient grounds to act otherwise. It is plausible that a sadistic serial killer might be unable to show mercy for her next victim because she lacks emphatic motivation. Even if she can perceive moral reasons, their motivational power is too low compared to her selfish sadistic desires. Why would she act suddenly in a more humanistic way if nothing inclines her to do so? Still, she is responsible, because her former choices made her conscience too weak. According to this kind of libertarianism, the agent is *ultimately responsible*[2] despite the fact that she cannot choose otherwise just before the murder, if there was, somewhere in the past, at least one key situation (i) where sadistic motivations have significant motivational rivals and (ii) the decision made in that situation is the very origin of the weakness of humanistic motivations. One can be free in a direct way only if one has at least two significant motivational tendencies.

Since Robert Kane has elaborated the theory of self-forming acts in most detailed fashion, my suggestions about moral development will mostly be based on Kane's theory. Huoranszki has a special objection to Kane's restrictivism. He claims that Kane's theory presupposes the reason-dependence of free choices.

[2] According to Kane, an agent is ultimately responsible for an act if their act's ultimate source is the agent herself, and not the environment, the past, education and so on. (Kane 1996: 33–35.)

In Kane's view, we can use our free will directly if the agent perceives at least two sets of reasons suggesting different choices about the particular situation (Kane 1996: 114). Kane holds that character-forming acts are based on rational choices, because this kind of choice can ensure that the agent is in control.[3] Desires and other possible irrational motives only increase the probability of alternatives. This is because they force the will to make a greater effort in so far as the agent tries to choose the other reason, which has less motivational support (Kane 2007: 36).

Huoranszki's problem with this understanding of reason-dependence is the following. We frequently have responsibility for acts which are not based on reasons. For instance, weak-willed or negligent actions, *actes gratuits*,[4] and so on. It is beyond the scope of this paper to analyze this claim. However, I should mention that restrictivist libertarians have an answer to this problem of self-formation even if they would not reject Huoranszki's claims about irrational acts. A restrictivist libertarian can claim agents are responsible for acts not based upon reasons as long as the strong irrational compulsions are consequences of a clearly reason-dependent self-forming choice. This answer is not doomed to failure *if* self-forming actions can be the ground of moral responsibility.

Beyond this issue, Huoranszki has other counterarguments against libertarian self-determination.

> A) The consequence of libertarian self-determination theory is either that the agents have a point in their life after which they just act quasi-mechanically or there is a stage in the agent's life when she does not have the relevant trait and hence cannot adequately perceive (moral) reasons. These consequences are very implausible (Huoranszki 2011: 170). Moreover, it is not clear how someone who cannot perceive reasons in relevant situations can be responsible.

> B) Libertarian self-determination is not the grounds of moral responsibility since agents are not able to control how their acts form their characters and motives. This is because agents cannot acquire the desired moral motives and character traits by conscious habituation (or at least this cannot be typical). Why? (b1) Conscious habituation often has undesired moral effects. (b2) Conscious habituation is not effective enough. (b3) Effects of regular actions on character traits cannot be foreseen and controlled by the agent. (Huoranszki 2011: 170-175.)

[3] Since, according to Kane, the responsible self's will is fundamentally rational. (Kane 1996: 21–28.)

[4] These are acts agents perform intentionally but for no particular reason.

C) The relative strength of some motives, reasons and character traits cannot deprive the ability to choose and to do otherwise in non-pathological cases (Huoranszki 2011: 174-175). This is because the ability to choose otherwise requires first and foremost that the agent can represent herself as somebody who can do more than one thing in a particular situation.[5]

I answer objection A) in section 1.2, (b1) in 1.3, (b2) and (b3) in 1.4, and C) in 1.5.

1.2 Character-forming acts and automatism

According to Huoranszki, the problem described in A) is the least worrying one. However, he thinks that it does pose a challenge for libertarian self-determination. The source of this problem is that restrictivist libertarians say there are strong opposing reasons/motivations only in the case of self-determining acts. It would imply that agents who have carried out self-forming acts in one way or another and acquired a new character trait behave quasi-mechanistically according to these characteristics.

Huoranszki suggests that the only alternative to avoid this consequence is a dead end. If libertarians claim that positive character traits enable agents to perceive moral reasons, this then would explain why she always acts rightly. But this solution does not help. If the agent cannot perceive moral reasons adequately before acquiring the proper character traits, how could the self-forming actions when the agent is not able to perceive these reasons be the grounds for her moral responsibility?

In my view, the first option of quasi-mechanistic acting is actually not that problematic or implausible provided we respect the complexity of the human motivational system. I think the human motivational system requires psychological and natural scientific investigation. In other words, philosophers can only make sketchy remarks about it. Nevertheless, it seems that agents have many different dispositions which can ground opposite motivations and reasons in particular situations.

Let us suppose that somebody has three relatively strong motivational dispositions: irascibility, respect for authority, and altruism. Furthermore, imagine that she resists her irascibility when she speaks with her boss, as she would like to respect her. This might strengthen her self-control so that she will always resist her irascibility when it clashes with respect for authority. However, this does not mean that she can resist her bad temper in every case. For example,

[5] This last claim is made before chapter nine (Huoranszki 2011: 41). Nonetheless, it is important because it is meant to explain why character traits and ordinary psychological states cannot, according to Huoranszki, deprive us of our ability to do otherwise.

in a situation where her irascibility clashes with her altruism, it is questionable which disposition will win. Or if her other dispositions militate against her respectfulness, she may not obey her superior. However, I believe it is probable that her former choices for respectful acts raise the probability of obeying.

The point is that, even if somebody acquires a new character trait, she may not act accordingly in a different type of situation where this trait has a different motivational "rival". Moreover, it is also possible that she loses her character trait. If the employee acts irascibly in other situations where the temptation has different motivational rivals, her irascibility may become stronger. As a result, she may later not be able to show respect toward her superior. In my opinion, all this is not incompatible with Kane's or Aristotle's conception of self-formation.

There is another reason why the charge of quasi-automatism is not an essential problem for restrictivist libertarianism. Granted, actions determined by strong character traits are automatic in the sense that the agent always acts in the same way in similar situations. In addition, in these cases the agent's habituation is so strong that it is possible to predict how she is going to act. But this need not prevent her from perceiving opposing reasons.

Kane distinguishes between the notion of external and internal reason (Kane 1996: 29-30). Usually, philosophers talk about internal reasons in the free will debate because only internal reasons have significant motivational power in the agent's deliberation process, while external reasons are not supposed to have such influence. They are mere facts, which theoretically could have been a reason for someone, but actually they do not motivate the agent. In Kane's example, someone who does not know that her friend will go to some party will not be motivated by this fact. Nevertheless, this fact is a reason for her to go to the party in the externalist sense.

In my opinion, the case of a perfectly virtuous agent would be similar to this in some respects. Granted, she could perceive that the morally wrong options have some benefits. So this example does differ in one way. Still, this knowledge did not really pose a danger that she would choose the morally right alternative every time. In short, she perceives morally wrong reasons, but these would be reason for her only in the externalist sense. Therefore, acquiring new character traits changes not the cognitive capacity and reason-perceiving ability (I do not exclude that this can sometimes happen during character-formation), but the motivational background, the motivational power of reasons, desires, and so on.

It is important to note that if we accept this view of self-formation suggested by Kane, ordinary habituation and self-formation will be similar in some respects. Nevertheless, there remain important differences. It may well be that the Aristotelian theory of self-formation criticized by Huoranszki does indeed put too much stress on the similarity. So I agree with Huoranszki that (b1) agents cannot acquire moral motives and character traits by conscious habituation (or at least this is not a paradigmatic case). Furthermore, I do not deny that (b2) the way

regular actions impact on character traits cannot constitute control in the ordinary sense of 'control'. Still, I claim that agents are responsible for the morally relevant consequences of character-forming acts.

1.3 The invisible hand of character-formation

Why is Huoranszki so skeptical regarding agents' ability to acquire motives and character traits by conscious habituation? Huoranszki does not deny that sometimes it is possible to gain these character traits in this way. Nonetheless, he has the problem that conscious habituation is not reliable, because trying consciously to develop a character trait can produce the opposite result (b1). His example is the following:

> Let us suppose that someone does not have the disposition to behave kindly or respectfully with others. But she does judge in many situations that the best thing for her to do would be to behave kindly and respectfully. Thus, she makes an effort and, if she is continent enough, then she can regularly behave as if she was naturally kind and respectful.
>
> The result of such kind of behavior may be disastrous. It is all too easy to imagine that instead of acquiring kindness and respectfulness, the person becomes a hypocrite. … There is no guarantee that the recognition of what behavioral patterns would manifest such traits and the attempt to conform one's behavior to such recognition will necessarily improve her 'moral self'. (Huoranszki 2011: 171.)

There is something odd about this example though. A hypocrite who shows respectful behavior to somebody else does not really think that this other person has earned her respect. She behaves this way only because she believes that this hypocritical behavior will somehow pay off. Consequently, a hypocrite does not really want to be respectful with other people if she really is a hypocrite. So if the agent would like to be a kind and respectful person because she thought that this is morally good and other people deserve this kind of treatment, she was obviously not a hypocrite to begin with.

I cannot imagine a scenario in which somebody, who has such good intentions, loses her respect towards somebody else because she tries to behave respectfully. Maybe Huoranszki is thinking here of someone who is misanthropic but does not like herself just because of her misanthropy. But again, if somebody hates her own misanthropy because she believes that people deserve better, she will not be a hypocrite if she behaves not as her inclinations dictate but as her rational considerations do. This remains true even if her misanthropic feelings will never change. Suppose the misanthrope learns how to behave differently from what her irrational dispositions dictate. Suppose, moreover, that

she behaves kindly, even in cases in which she rationally thinks that it would be good if she had not behaved so nicely. The problem is not that she learned to be kind to people who deserve it. The real problem is that she does not have sufficient self-control to use her skills rightly. But this is a different problem.

The source of disagreement with Huoranszki here is that I attribute great importance to motivations to behave morally rightly. If somebody acts appropriately because she attempts to gain values which are distant from the territory of the morally right, the behavior and the choice which is behind the act might not have had a positive character-forming effect. Moreover, in my view, the character-forming choices which have a morally good effect have a different motivational center than the desire forming the agent's character.

This is the point where the Aristotelian model needs to be refined. Aristotle does not pay enough attention to the fact that motivations are indifferent in the case of ordinary skills but very important if we try to develop our moral virtues. If somebody paints frequently because she would like to learn painting, it is not relevant why she wants to be a good painter. But this is an important aspect if we investigate the problem of moral development.

It is at the very least suspicious if somebody acts morally rightly just for the reason that she would like to acquire a better character trait. But it is an entirely different case if the agent desires better character traits because, for example, she would like to help other people. Also, more commonly, in the case of positive character-forming actions the agent does not think about the action's character-forming power at all. The agent concentrates only on the action's potential good effects on other people and on its moral rightness. By contrast, if somebody wants to act morally rightly because she desires to gain new character traits, she uses people as a means. It may well be that the main motivation is only vanity or ambition.[6]

There is no reason to suppose that suitably motivated unselfish choices could have any morally problematic side-effects on one's personality. A long series of morally impeccable choices ensures the development of moral character unintentionally and invisibly, just as selfish choices ensure economic growth in Adam Smith's theory.

1.4 Pre-established harmony

Huoranszki criticized Aristotelian self-formation theory for claiming that the main source of moral responsibility is the conscious and direct practice of virtue. I granted that conscious and direct practice is not the paradigmatic form of gain-

[6] Robert Kane (1996: 126–127) thinks also that the typical examples of self-forming choices are not outrightly directed at self-formation.

ing new morally important character traits. So I also rejected the Aristotelian picture, at least in part.

But my solution made it less clear how agents can control their character development. One problem is this. If character-forming choices have limited consequences, they cannot totally guarantee our virtuous acting in situations which are different from the original character-forming situation (b2). Their effects are too partial. This is Huoranszki's second problem about self-forming actions.

His third objection is the following. If conscious practice is not the best way to gain character traits, how we can ensure that we will be morally good people? If we cannot foresee what the consequences of our acts will be, why would we be responsible for our morally wrong acts (b3)? After all, their origin was a seemingly harmless choice, the effects of which were not predictable. First, I attempt to answer to (b3) objection. Subsequently, I will try to answer and offer a possible solution to problem of the limited efficacy of self-forming acts.

I claimed previously that actions and choices based on morally perfect motivations had no harmful influence on character development. In addition, it seems that these choices have frequently good effects. By the same token, self-forming choices based on inappropriate motivations almost always have a negative impact. I also suppose that morally neutral choices, based on neither positive nor negative motivations, have no morally relevant outcomes. My point is that if such 'pre-established harmony' actually exists between choices/acts and character-forming results, there is no need to foresee or control the character-forming effects of choices to be responsible for them. It is enough if the agent knows which acts and choices are morally good and morally bad. Or, at least, we can say that the agent should have known this.

For instance, *if somebody knew* from a reliable person that it would be *morally wrong* to pour water on a box with a bright red 'dangerous' label on it, she is morally responsible for the explosion if she does so. This will be so even if she did not know that it was a necessary consequence of the fact that this box contained sodium explosive. Another example is the following. Suppose that somebody knows that using strong drugs for hedonistic aims is morally bad. However, she has not heard that strong drugs turn people into addicts. If she does not care about her moral knowledge, she is not only morally responsible for using drugs. She is also responsible for becoming an addict.[7] Similarly, if a person knew that it would be blameworthy if she acted in a particular way, and she chose this possibility anyway, she is responsible both for the choice and its bad influence on her character, even if she could not have known anything about the character-forming effect of her choice. On the one hand, if we had a strong notion of control, agents did not control the results—neither the destructive explosion nor

[7] Nevertheless, this kind of ignorance about the character-forming effects can slightly moderate the agent's blameworthiness.

the changing of moral character. On the other hand, if we use a weaker notion of control, we can say that she had control over what will happen in the future. In my opinion, this latter degree of control is enough to be morally responsible for a character-forming effect, if the agent knew or should have known the *moral value*[8] of the possible choices, *since there is harmony between character-forming choices and their results*.

But how is such harmony possible? I cannot present a full theory of character-formation here. My purpose is only to prove that Huoranszki's arguments are not conclusive against restrictivist libertarianism. But to make this defense more plausible, I briefly have to say something about the general issue as well.

So, first, I think that the presupposition of "harmony" fits best with an "intentionalist morality" in which wrong acts are based directly or indirectly on morally problematic intentions. An intention of a morally responsible agent is morally problematic if it is directed at some option of inferior value compared to other alternatives also accessible in the particular situation (provided the agent knew or should have known that this option is less valuable). If an intention is based on such inferior reasons, it will be less and less probable that the agent will form intentions based on morally superior values later. This is because she becomes accustomed to choosing in this way. Moreover, she is likely to identify increasingly with the value set compatible with her former choices.

Consequently, people who are motivated by selfishness where other people's interests would demand that they tell the truth will be more likely to lie in similar future situations. The probability of immoral action by such agents is increased in different situations as well. For instance, in a situation where the question for the agent is whether to embezzle some money or not.

If altruist intentions are indeed so central to morality and moral development, we can answer Huoranszki's second objection about self-forming action's partial effects. Altruism has many different manifestations in different virtues. Nonetheless, these moral virtues are not totally independent from each other. Each one is linked in one way or another to the willingness to undertake selfless actions. If this is true, every altruist choice can be considered a "preparation" for other moral challenges. So I claim that if one does everything to be an altruist person, that is, if it is a real possibility that one unselfishly chooses the morally good option, then one can form their character effectively. Therefore, one can be morally responsible for a morally wrong action even if one is not able to do otherwise, provided this inability is a consequence of a selfish and morally problematic choice in the first place.

[8] If she knows about the moral value of an act, it does not mean that she knows about the self-forming effects of the action.

1.5 Representations without motivations—calculable failure

Huoranszki could complain that such a theory of character formation is not required. According to Huoranszki, character traits and the motivational background cannot undermine the agent's ability to do otherwise. Furthermore, the ability to do and choose otherwise presupposes only that the agent is able to recognize moral reasons and represent herself as somebody who can act in more than one way (Huoranszki 2011: 41). And it is beyond dispute that a brave man can represent himself as somebody who runs away from battle. Also, a coward is able to see himself as somebody who dies for his country.

I do not agree with Huoranszki on this point either. Sometimes ordinary people who have no pathological psychological problems can misrepresent their abilities in such a way that they have a false belief about what they are able to do psychologically. Let us suppose a football player in his thirties thinks about whether to retire. He is not particularly self-aware and does not know that he is a very ambitious person. But his mother is wiser, and knows that her son would be unable to keep his promise to spend more time with his family after his retirement, even if he did perceive that it would be the morally right choice. The reason for this is that the football player has no real desire to act in the morally right way. Nor would he really like to spend much more time with his family. He made his promise just because he wanted to put an end to a quarrel with his wife. In fact, he deceived himself about his real motivations.

I think that such self-deception is not pathological. The football player is responsible if he breaks his promise to his family. This is because the lack of appropriate motivation is explained by his former selfish choices.

Huoranszki would disagree. He would say that this football player could have chosen otherwise in this particular situation even if he had a selfish character.[9] He perceives the right course of action and he thinks that he is able to choose it. He just does not choose this alternative, as it turns out. But why does he not choose this alternative? Huoranszki suggests that there is no full explanation. Nor is such an explanation possible in principle (Huoranszki 2011: 162). Contrary to restrictivist libertarianism, Huoranszki (2011: 118) denies that psychological states of non-pathological agents can determine how they choose and act.

Firstly, I believe that this is empirically improbable, because not only psychological experts but also ordinary people who are good judges of character can predict decisions. Take the football player's mother in the previous example. Secondly, sometimes we feel that we can give a *perfectly exhaustive* answer to the

[9] Moreover, Huoranszki (2011: 175) claims that the only actions that reveal our character are ones we could have avoided. I beg to differ. I agree that actions that are physically impossible to avoid cannot reveal our character. However, I would argue that actions unavoidable due to psychological states are the most revealing as to our character traits.

question of why the agent decides in one way and not another, by referring to the agent's reasons, desires and other psychological states.

To summarize, in the first part of my paper I provided a possible defense of restrictivist libertarianism against Huoranszki's arguments. Moreover, I have argued that the hypothesis of self-formation is not all that implausible. Although the right theory of self-formation has to diverge to some extent from Aristotle's approach, restrictivist libertarianism is a better theoretical option to save moral responsibility.

2.0 Are we responsible for not doing miracles?

In the second part of chapter nine, Huoranszki returns to the question whether free will and moral responsibility are compatible with physical determinism (Huoranszki 2011: 175-182). Many libertarians claim that even if self-formation is not necessary for exercising free will, physical indeterminism is an indispensible condition so that agents can be independent enough of their social and physical environment.

Contrary to some compatibilists, Huoranszki accepts that some degree of independence is needed for moral responsibility. But he claims that social and physical laws do not endanger our independence.

First of all, he denies the possibility of strong social determinism. There is no social training which could deprive us of our ability to do otherwise. Social training mainly determines our character and motives. But, according to Huoranszki, our character and motives cannot determine how we choose or act. Therefore, social training cannot determine how we act in a particular situation. As already noted, I do not agree with this because I think motives and character can determine our action in some cases.

Huoranszki thinks the only possible way that social determinism can be a problem for a compatibilist is if libertarians can prove that social determinism necessarily follows from physical determinism. I agree with Huoranszki that this is a difficult, perhaps impossible, task. However, the libertarian does not have to show this in order to refute Huoranszki.

I believe Huoranszki accepts all of the following claims. First, that free actions need more independence than physical particles with regard to physical laws and past events. It is important for Huoranszki that social and psychological phenomena cannot reduce to physical ones. Second, that our actions are physical events (at least partly). Third, that there are free actions. But given these claims it is hard to see how they can simultaneously be true. The argument to show this runs as follows:

(1) Two alternative actions related to the same agent and the same situation are connected to different movements of particles constituting the body. So the agent could have acted otherwise in every case only if her body's particles could have moved otherwise.
(2) 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.
(3) Every movement of a set of particles is dependent on physical laws and past physical events to such a high degree that, according to physical laws, if determinism is true, any moving sets of particles could have moved otherwise only by some miracle (i.e. due to an event violating physical laws).
[2+3] (4) Every action is dependent on physical laws and past physical events to such a high degree that, if determinism is true, any action carried out according to physical laws could have been done otherwise only at the cost of a miracle.
(5) (At least) most actions are carried out according to physical laws and past physical events in the actual world.
(6) Nobody can be responsible for an action which could only have been done otherwise at the cost of a miracle.
(C) No agent can be responsible for (at least) most of her actions in the actual world.

This argument modifies Bács's (2012) supervenience argument,[10] and the independence argument of incompatibilists such as Pereboom.[11] I attempted to preserve the main intuitions underlying these arguments while directing them specifically against Huoranszki's compatibilism.

The first two premises establish the consequences of the fact that the execution of every action supervenes on movements of our particles. We cannot act differently from action 'x' if our body's particles do not move differently from how they would move if we carried out action 'x'. Thus if the conjunction of physical laws and the remote past determines which movement of the particles would be a miracle at a particular time *t*, this also determines which act would be found in this "miraculous" category. Furthermore, if physical determinism is true, only one kind of movement for every set of particles is not a special movement at any particular time. But since every different action implies different

[10] Bács's argument relies on the supposition that mental states supervene on brain processes. I utilize only the supervenience-relation between the movement of action and particles. I do so because the former relation is unclear. Huoranszki (2013) takes advantage of this in his answer to Bács.

[11] Pereboom (2001) argues that if determinism is true, we are not responsible because the environment determines how we choose and act. This is because there is no important difference between manipulation and ordinary causal chains. By contrast, my point here is that the world determines only whether an action would be a miracle or not. I think this is sufficient to reject Huoranszki's special version of compatibilism.

movements by the particles, at any time *t* an agent can execute only one action that is not a miracle.

Huoranszki holds that if we are morally responsible, we are able to act in more than one way. He also claims that an agent morally responsible because she is able to act in more than one way. Together with the claims above it would follow that if determinism is true, we are able to perform at least one miraculous action in every situation when we are morally responsible. Moreover, it would follow that we are morally responsible just *because* we could have performed a miracle (viz. other than how the laws of physics dictate it).

This would be an absurd conclusion. Even if agents were able to perform miracles in some sense of the word, this ability could not be grounds for moral responsibility. For instance, this would have the unacceptable consequence that if the agent failed to meet her obligation, fulfilling her obligation would have been a miracle.

SUMMARY

I have tried to show that restrictivist libertarianism is a defensible theory. I also pointed out that Huoranszki gives us too much freedom when he argues that an agent's character and her psychological background can never determine choices and actions of psychologically healthy persons. Furthermore, his compatibilism cannot handle the apparent implication of physical determinism that only one possible action of the agent is not a miracle at any given time *t*.

Huoranszki's objections against the theory of self-forming actions can force libertarians to develop a self-formation theory less directly based on the Aristotelian analogy between character-development and the acquisition of practical skills. I have sketched such a theory here, but of course there remain many open questions about character development.

REFERENCES

Bács, Gábor 2012. Huoranszki Ferenc: Freedom of the Will. A Conditional Analysis. *BUKSZ* 24, 307-313. [In Hungarian.]

Huoranszki, Ferenc 2011. *Freedom of the Will: A Conditional Analysis*. New York: Routledge.

Huoranszki, Ferenc 2013. Válasz Bács Gábor ellenvetéseire. *BUKSZ* 25, 6-9. [In Hungarian.]

Kane, Robert 1996. *The Significance of Free Will*. Oxford: Oxford University Press.

Kane, Robert 2007. Libertarianism. In John Martin Fischer, Robert Kane, Derk Pereboom, Manuel Vargas, *Four Views on Free Will*. Oxford: Blackwell Publishing, 5-43.

Pereboom, Derk 2001. *Living Without Free Will*. Cambridge, UK: Cambridge University Press.

Ferenc Huoranszki

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 none-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 form 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 orectic 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 subpersonal 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.

Contributors

Gábor Bács • assistant professor, Department of Social Science, University of Kaposvár, Hungary. Areas of specialisation: metaphysics.
E-mail: bacs.gabor2@gmail.com

Tibor Bárány • postdoctoral researcher, "Human Project", Central European University, Hungary. Assistant professor, Department of Sociology and Communication, Budapest University of Technology and Economics. Areas of specialisation: philosophy of language, philosophy of art.
E-mail: barany.tibor@gmail.com

László Bernáth • PhD student, Department of Philosophy, Eötvös Loránd University, Hungary. Areas of specialisation: metaphysics.
E-mail: bernathlaszlo11@gmail.com

Dániel Corsano • graduate student, Department of Philosophy, Eötvös Loránd University, Hungary. Areas of specialisation: metaphysics.
E-mail: daniel.corsano@gmail.com

Ferenc Huoranszki • full professor, Department of Philosophy, Central European University, Hungary. Areas of specialisation: metaphysics, philosophy of action.
E-mail: huoransz@ceu.hu

Anna Réz • postdoctoral researcher, „Human Project", Central European University, Hungary. Areas of specialisation: moral philosophy, moral psychology.
E-mail: rez_anna@ceu-budapest.edu

Howard Robinson • full professor, Department of Philosophy, Central European University, Hungary. Areas of specialisation: philosophy of mind, metaphysics.
E-mail: robinson@ceu.hu

Judit Szalai • associate professor, Department of Philosophy, Eötvös Loránd University, Hungary. Areas of specialisation: philosophical psychology, 17th-century philosophy.
E-mail: szalai@elte.hu

András Szigeti • senior lecturer, Department of Philosophy, Linköping University, Sweden. Postdoctoral research fellow, Department of Philosophy, The Arctic University of Norway, Tromsø. Areas of specialisation: moral responsibility, collective agency, philosophy of emotions.
E-mail: andras.szigeti@uit.no

Zsófia Zvolenszky • associate professor, Department of Philosophy, Eötvös Loránd University, Hungary. Areas of specialisation: philosophy of language, logic, metaphysics, formal semantics.
E-mail: zvolenszky@nyu.edu