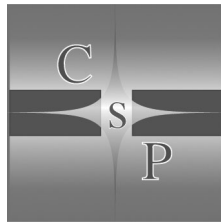


Essays on Free Will and Moral Responsibility

Essays on Free Will and Moral Responsibility

Edited by

Nick Trakakis and Daniel Cohen



Cambridge Scholars Publishing

Essays on Free Will and Moral Responsibility, Edited by Nick Trakakis and Daniel Cohen

This book first published 2008

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2008 by Nick Trakakis and Daniel Cohen and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-84718-867-2, ISBN (13): 9781847188670

TABLE OF CONTENTS

Acknowledgments	vii
Introduction	ix
Nick Trakakis and Daniel Cohen	
1. Derk Pereboom	
<i>Defending Hard Incompatibilism Again</i>	1
2. Nick Trakakis	
<i>Whither Morality in a Hard Determinist World?</i>	34
3. Trevor Pisciotta	
<i>Meaningfulness, Hard Determinism and Objectivity</i>	71
4. Manuel Vargas	
<i>Moral Influence, Moral Responsibility</i>	90
5. J.J.C. Smart	
<i>The Illusion of Libertarian Free Will</i>	123
6. Neil Levy	
<i>Restrictivism is a Covert Compatibilism</i>	129
7. Robert Kane	
<i>Three Freedoms, Free Will and Self-Formation: A Reply to Levy and Other Critics</i>	142
8. Ishtiyaque H. Haji	
<i>Obligation and Luck</i>	163
9. Michael McKenna	
<i>Ultimacy and Sweet Jane</i>	186

10. John Martin Fischer	
<i>The Direct Argument: You Say Goodbye, I Say Hello</i>	209
11. David Widerker	
<i>Some Further Thoughts on the Direct Argument</i>	224
12. Saul Smilansky	
<i>Free Will and Fairness</i>	234
13. Daniel Cohen and Lauren Saling	
<i>Addiction Is No Excuse</i>	247
Contributors.....	265
Index.....	269

ACKNOWLEDGMENTS

This collection of essays has its roots in a conference on free will and moral responsibility held at Monash University in November 2005, though only a few of the papers presented at the conference have made it into the current volume. We would like to thank both the participants at this conference and the contributors to this volume, as well as Cambridge Scholars Publishing for inviting us to put the collection together.

Grateful acknowledgement is also made to the editor of the online journal, *Sorites*, where Nick Trakakis' paper, "Whither Morality in a Hard Determinist World?", was originally published (in vol. 19, December 2007).

Finally, thanks to Shannon Weekes for her assistance in compiling the Index.

INTRODUCTION

NICK TRAKAKIS AND DANIEL COHEN

Much of the interest of the free will debate depends on the assumption that free will is necessary for moral responsibility. In particular, it is because responsibility seems so important for our practical lives that debates about the compatibility of free will and determinism seem so urgent. However, much of the discussion in this volume bypasses this link. Instead, questions are raised that *directly* concern responsibility, such as whether it is compatible with determinism (see, for example, the essays by Fischer, Widerker, and Pereboom) and whether it is compatible with indeterminism (for example, the exchange between Levy and Kane).

For the purposes of this introduction, we have not attempted to summarize the various ways in which the contributors construe the metaphysical foundations of moral responsibility. Instead, we wish to address a more preliminary matter. In the first part of this introduction, our aim is to say something about what we *mean* when we say that someone is morally responsible. It is surely important to clarify this before addressing any further substantive issues because, if we don't clarify the meaning of this key term, there remains a significant danger that different participants in the debate about the possibility of moral responsibility will simply 'talk past each other'. This suggests that in order to conduct a fruitful debate participants need firstly to agree on the nature of their subject-matter and, perhaps, to disambiguate different dimensions of the debate that arise if the term 'moral responsibility' has different connotations. In the second part of the introduction, we will discuss a neglected Wittgensteinian perspective on the notions of freedom and responsibility, a perspective that may help to clarify some of the confusion that arises when we ask what it means to say that a person is free or responsible.

1. The Meaning of Responsibility

Before proceeding to ask whether people are, in fact, ever morally responsible, it seems that an important preliminary matter needs to be settled. That is, we need to ask what we *mean* when we say that a person is

morally responsible. As will quickly become clear, settling this preliminary matter is, in fact, much harder than it first seems. Many of the controversies concerning the *possibility* of responsibility emerge even when we try to say just what ‘responsibility’ means.

Let’s start with a claim that seems relatively uncontroversial. We will simply assume that ‘person A is responsible’ is a normative claim. That is to say, it is a claim to the effect that it is *appropriate* to hold A responsible in certain circumstances (circumstances, for example, where A has acted wrongly and where no mitigating, justifying, or excusing factors are present). However, this account of the meaning of ‘A is responsible’ raises at least two further issues. Firstly, we need to say more about what it means to hold someone responsible and, secondly, we need to say more about the nature and basis of the norms that govern appropriate responsibility attribution.

It is important to keep in mind that our goal, in clarifying these two questions, should not be to settle any question regarding whether people are, in fact, ever morally responsible. Rather, we want to settle the terms of this debate before it begins. To this extent, we need sufficiently neutral accounts both of the nature of responsibility attribution and of its normative basis so that we don’t beg any substantive questions before debate has even begun. As we will try to show, however, this is a rather elusive goal. The only neutral account of the nature of responsibility attribution renders the normative question deeply controversial. And the only neutral account of the normative basis of responsibility attribution renders the nature of responsibility attribution deeply controversial.

Holding Responsible

There appear to be two plausible contending views regarding the nature of responsibility attribution. On *cognitivist* accounts, holding A responsible fundamentally involves believing something to be true of A, while on *non-cognitivist* accounts, holding A responsible essentially involves holding some conative attitude towards A. (Cognitivists may, of course, argue that responsibility attribution is also usually associated with some conative attitude. However, they will maintain that it is possible to hold someone responsible without holding such conative attitudes. Similarly for non-cognitivist accounts, *mutatis mutandis*.)

Non-cognitivism appears to provide the most successful neutral basis on which to premise the debate concerning the possibility of responsibility. This is because there seems little room for debate concerning the conative attitudes that characterize normal responsibility

attribution. In particular, few would disagree that responsibility attribution is strongly associated with the ‘reactive attitudes’ identified by P. F. Strawson, i.e., resentment, indignation, anger and so on.¹ If one wishes to argue, however, that the reactive attitudes, while prevalent, are inessential to responsibility attribution, it is much harder to locate any common ground concerning the beliefs that are essential to responsibility attribution.

One may suggest, for instance, that to hold A responsible is to believe that she is the *source* of some bad behaviour. Deep controversies quickly emerge on this view, however. One might take sourcehood to involve a psychological claim, for instance that A ‘really wanted’ to act wrongly.² However, others might object that any such glib psychological account fails to explain why it is *fair* to blame A for the wrongdoing (see Smilansky’s contribution). One might object, in this vein, that any such psychological story is unable to show that an agent really is the source of her having certain desires or values (see McKenna’s contribution), and that sourcehood thus requires some more obscure metaphysical basis (e.g., agent-causation). Alternatively, one may suggest that sourcehood involves some impossible requirement such as that an agent was *self-created*.³ On this view, holding someone responsible is essentially impossible.⁴

Our goal is to account for the meaning of responsibility in neutral terms so as to provide a basis for constructive debate about the conditions (and the very possibility) of responsibility. It appears, however, that the cognitivist view of responsibility attribution quickly leads to debates that already beg these important questions before debate has even begun! This suggests that the best theory-neutral account of the meaning of responsibility must explain holding responsible in non-cognitivist terms.

The Normative Basis of Responsibility Attribution

Recall that, for the purposes of this discussion, we have assumed a normative account of responsibility according to which ‘A is responsible’ means ‘it is appropriate to hold A responsible in certain conditions’. Having addressed how best to interpret what ‘holding A responsible’ might mean, without begging any important questions, we need now to turn to a second question raised by the normative account: when exactly is it appropriate to hold someone responsible? In other words, what are the norms that govern appropriate responsibility attribution? Again, there are two plausible contending views: appropriateness may be explained either in terms of practical norms (taking ‘holding responsible’ to be analogous

to the performance of an action) or by way of doxastic norms (taking 'holding responsible' to be analogous to the formation of a belief).

Again, only on one of these accounts—the doxastic view—is it possible to offer an appropriately uncontroversial explanation of the norms implicit in responsibility attribution. On the doxastic view, one ought to hold A responsible if and only if *it is true* that A is responsible. On this view, the normative basis of responsibility attribution straightforwardly derives from the normativity of belief. It is clear that the doxastic account presupposes the cognitivist view discussed earlier, according to which holding A responsible involves believing something about her. Given this view of the nature of responsibility attribution, the normative question—concerning when responsibility attribution is appropriate—has a straightforward answer.

Unfortunately, as we saw, there is no uncontroversial way to account for the truth-conditions of 'A is responsible', on the cognitivist assumption that it involves belief. So, despite the advantages of the doxastic view in providing a neutral account of the normative basis of responsibility attribution, this view at the same time precludes us from obtaining a neutral view regarding the nature of responsibility attribution (i.e., the truth-conditions for the belief that A is responsible.)

Might we find an account of the normative basis of responsibility attribution that is consistent with the preferable non-cognitivist view outlined earlier? This would have to involve an alternative view, according to which responsibility attribution is justified in virtue of practical norms. However, if responsibility attribution is governed by practical norms, then things are much less straightforward. One may suggest that the relevant practical norms are just moral norms, so that 'A is responsible' states something like: 'It is morally obligatory (or, perhaps, permissible) to hold A responsible'. This view may appear immediately problematic because the appropriateness of responsibility attribution will now depend on further questions that are deeply controversial (for instance, questions concerning the debate between consequentialism and non-consequentialism; see Vargas' contribution). A more fundamental worry arises concerning the methodological appropriateness of appealing to moral norms. One may argue, for instance, that the nature of moral obligation, itself, depends on the foundations of responsibility, which is, of course, the question at issue. Haji (in his contribution) argues that the best metaphysical basis of responsibility (i.e., event-causal libertarianism) renders moral obligation essentially lucky. This suggests that there would be something viciously circular in explaining the meaning of responsibility

in terms of some claim about our moral obligations. (See also Trakakis' contribution.)

To avoid these worries, one may appeal to practical norms that appear to be more fundamental than any particular moral system. For instance, R. J. Wallace offers a normative account of responsibility that appeals to fairness.⁵ This is meant to provide a normative basis for responsibility that remains neutral on more substantive moral issues. (See also Smilansky's contribution.) Clearly, however, appealing to practical norms launches us into further debates that already beg the question at issue. Again, such an account seems ill-suited for the purpose of setting up a neutral definitional framework on which to premise further debate.

A Dilemma

Our aim has been to find some neutral definition of responsibility to enable further non-question-begging debates about the possibility and conditions of responsibility. It seems that this goal gives rise to a tricky dilemma. The best theory-neutral account of holding responsible is the non-cognitivist account. However, this account appears incompatible with the best theory-neutral account of the norms that govern responsibility attribution—the doxastic account. The doxastic account, in turn, seems compatible only with the most problematic account of holding responsible—the cognitivist account.

This is a puzzling result. Even though responsibility clearly gives rise to very complex issues, it is surprising that it is not possible even to define the *terms* of the debate without deep controversy. The worry thus remains that debates about the possibility and conditions of responsibility are essentially question-begging, insofar as different participants to the debate conceive of its key terms differently.

Must we conclude, then, that different people and different theorists are indeed talking past each other when they debate about the possibility of responsible action? This, of course, would be a depressingly deflationist conclusion.

There is a possible way out, however, that is rarely canvassed. If the question concerns the meaning of 'responsibility', one might suggest that there are, perhaps, other ways to settle things. In particular, isn't the meaning of a term determined by our *use* of the term? (Or, at least, isn't use a good *guide* to meaning?) Thus, appeal to real-world attributions of responsibility may illuminate the meaning of the term 'responsible'. Such a methodology is sometimes rejected on the grounds that real-world attributions are rife with internal inconsistency (see, for example, Cohen

and Saling's contribution). But perhaps a closer look at the Wittgensteinian 'solution' to the problem of freedom and responsibility will throw new light on the matter.

2. Making Sense of Free Will: A Wittgensteinian Account

Wittgenstein published very little during his lifetime, and even less on the topic of free will. He does, however, make some pertinent remarks in his *Tractatus Logico-Philosophicus* that mirror in some respects the contemporary debates on free will. As is well known, the early Wittgenstein claims to have definitively solved the central problems of philosophy, and by implication this would include the perennial problem of free will. Wittgenstein's strategy proceeds by separating sense from nonsense. The realm of sense is delimited in the light of his picture theory of meaning, according to which a proposition is meaningful (or capable of expressing a fact) only if it can represent or picture a contingent state of affairs. What cannot thus be represented is consigned to silence, or as Wittgenstein famously put it at the conclusion of his book, "What we cannot speak about we must pass over in silence" (proposition 7).⁶ In the course of the book, however, it soon becomes clear that what can be meaningfully said are only the propositions of natural science, thereby leaving out of the realm of sense a daunting number of statements which are regularly made and used in language, including the propositions of logic, aesthetics, religion, and (most relevant for our purposes here) traditional metaphysics and ethics. But unlike the positivists, Wittgenstein does not assume that what is nonsensical is of no value. As he stated in a letter to a prospective publisher, Ludwig von Ficker: "My work consists of two parts: the one presented here plus all that I have *not* written. And it is precisely this second part that is the important one."⁷ Further, ethical and metaphysical truths that cannot be 'said' or formulated in sayable (sensical) propositions can nonetheless be 'shown': "There are, indeed, things that cannot be put into words. They *make themselves manifest*. They are what is mystical" (6.522).

In line with this austere outlook, Wittgenstein hints in the *Tractatus* that the very concepts at issue in the free will debate—concepts such as 'freedom' and 'responsibility'—cannot be meaningfully expressed. Although in propositions 5.1361 ("We *cannot* infer the events of the future from those of the present") and 5.1362 ("The freedom of the will consists in the impossibility of knowing actions that still lie in the future") he seems to reject determinism as false and to accept freedom of the will as true, rather than rejecting both as nonsensical, here we arguably have a

minimalist conception of free will as nothing more than ignorance or uncertainty regarding the future, as opposed to a substantive commitment to the metaphysical idea of a ‘will’ that could be free or unfree. This is confirmed in later passages where Wittgenstein takes the law of causality—the principle that every event has a cause—to be “not a law but the form of a law” (6.32), adding a few propositions later that, “If there were a law of causality, it might be put in the following way: There are laws of nature. But of course that cannot be said: it makes itself manifest” (6.36). The law of causality, in other words, is not itself a law of logic nor a law of nature (or an empirical generalization), nor a synthetic *a priori* proposition, but rather “something purely logical” (6.3211), a vacuous principle that tells us, not something about the world, but only something about the form our thinking about the world must take. But what is formal, according to the *Tractatus*, can only be shown, not said. On this view, then, the law of causality, and by extension any substantial or metaphysical doctrine of determinism, cannot be affirmed or denied, but must be placed in the category of ineffability or nonsense. Similarly, the denial of determinism—viz., indeterminism—is bound to result in nonsense. At one stroke, then, Wittgenstein seems to have dissolved the free will problem.

Contemporary discussions of free will often take a similar turn. For example, concepts such as ‘free will’ and ‘moral responsibility’ are routinely rejected as internally incoherent or contradictory, or as incompatible with determinism or indeterminism (or both), and like the early Wittgenstein this result is achieved simply through a kind of armchair or *a priori* reflection on the conditions of possibility of free will and responsibility. A glaring instance of this is Galen Strawson’s ‘pessimist’ conclusion that free will, of the sort that is necessary for genuine moral responsibility, is impossible, for in order to have that kind of free will (according to Strawson) one would *per impossible* have to be the ultimate cause or origin of oneself, a sort of *causa sui*. But what is neglected in this procedure is attention to particulars, to the variety of ways in which concepts such as free will and responsibility function in different discourses and social practices.

This, of course, is the message of the ‘later Wittgenstein’, the Wittgenstein of the *Philosophical Investigations*⁸, which effects a fundamental change of perspective: from the realm of an idealized logical language with rigorous definitions and analyses to the vagaries of everyday life and action out of which arise the multifarious ‘language games’. The earlier reduction of language to representation is now seen as incapable of doing justice to the rich fabric of human language, and so

Wittgenstein famously moves from a conception of meaning as representation to a view of meaning as use: language as a kind of doing rather than a kind of picturing. We are therefore exhorted to “look and see” (*PI* 66) whether there is anything in common in the variety of uses to which a word is put. We cannot simply assume that words like ‘freedom’ and ‘responsibility’ must have a hidden essence, or a universally applicable meaning that can be formulated in terms of necessary and sufficient conditions. Rather, we need to look to the complicated network of overlapping and criss-crossing similarities—what Wittgenstein calls ‘family resemblances’—between various words as these find expression in ordinary life and in various language games.

Each language game, however, has its own unique ‘grammar’ (or network of rules which determine what linguistic or conceptual moves are allowed as making sense) and its own criteria of truth, rationality and intelligibility which may or may not be shared by other language games. What counts as freedom of the will may therefore differ widely depending on which language game is being played: freedom, for some religious believers, requires the extinction of one’s will, while for purposes of legislation freedom may be conceived of as requiring a significant degree of self-determination. Underlying this view is the rejection of the mathematical ideal of the *Tractatus*, typified by statements that are put forward as impersonal, unambiguous and impervious to context, and whose truth is intended to be timeless and without qualification. Language, for the later Wittgenstein, is a much more dynamic, diversified and activity-oriented phenomenon. And to be faithful to the richness and complexity of this phenomenon demands an appreciation of the bewildering and sometimes conflicting ways in which words and concepts—‘free will’ and ‘responsibility’ included—are used.

It may be instructive to briefly compare (the later) Wittgenstein’s treatment of religious belief with his remarks on free will and voluntary action. Consider, for example, how Wittgenstein, in his “Remarks on Frazer’s *Golden Bough*”, responds to the phenomenon of religious diversity, to the fact that different religions seem to say different and incompatible things about (e.g.) the nature of divine or ultimate reality and the nature and destiny of the human race:

Was Augustine in error, then, when he called upon God on every page of the *Confessions*?

But—one might say—if he was not in error, surely the Buddhist holy man was—or anyone else—whose religion gives expression to completely different views. But *none* of them was in error, except when he set forth a theory.⁹

Against philosophers and anthropologists like James George Frazer, who construe religious doctrines as hypotheses or theories that can be confirmed or disconfirmed in light of empirical evidence, Wittgenstein views each religion as embodying a unique form of life that finds expression in language games whose ‘rules’ (relating to truth, rationality, intelligibility, and so on) may diverge quite dramatically from those of science. On this view, the various religions of the world are not in the business of constructing hypotheses and searching for evidence, and so they are not in competition with one another, at least in the way that scientific theories may vie for the mantle of verisimilitude. The problem of religious diversity is therefore dissolved. No language game, religious or otherwise, has a monopoly on truth and on the meaning of ‘truth’. Wittgenstein is thus opposed to both the religious exclusivist, who maintains that there is one religion which is privileged with respect to truth and soteriological effectiveness, and the scientific philosopher who argues that the methods and techniques of science (perhaps construed broadly enough to include philosophy and logic) are our only reliable guide to truth. Instead, Wittgenstein places all language games on an equal footing, allowing a thousand flowers to bloom. Thus, as David Pears has perceptively pointed out, Wittgenstein’s later work has

...an extraordinary levelling effect. It does not assimilate one kind of discourse to another: on the contrary, it is always the differences between them that are emphasized, and particularly the difference between factual discourse and the other kinds. But it does bring all the great philosophical questions which arise within them back to the same level, ordinary human life, from which philosophy started. Philosophy is the voyage out, and the voyage back, both of which are necessary if the logical space of our ordinary linguistic practices is to be understood.¹⁰

One of these ‘great philosophical questions’ is, of course, the question of free will and moral responsibility. On this matter, as with questions to do with religious faith, Wittgenstein refused the dominance of science on modern thinking: rather than constructing systematic theories that dictate from ‘on high’—inevitably from a position that holds up science as the model for all other discourses—how the phenomena in question are to be understood, we are brought back time and again to particular facts and examples rooted in everyday language and practices.

Taking such an approach to free will can produce startling results. For one thing, the belief in free will begins to look more like a religious commitment than a theoretical or scientific belief. Wittgenstein’s Kierkegaardian (or, more pejoratively, fideistic) account of religious belief is well known:

It strikes me that a religious belief could only be something like a passionate commitment to a system of reference. Hence, although it's *belief*, it's really a way of living, or a way of assessing life. It's passionately seizing hold of *this* interpretation.¹¹

But his somewhat similar account of free will has received less attention, despite the ever-expanding publishing industry on free will:

In the sense in which asking a question and insisting on an answer is expressive of a different attitude, a different mode of life, from not asking it, the *same* can be said of utterances like "It is God's will" or "We are not masters of our fate". The work done by this sentence, or at any rate something like it, could also be done by a command! Including one which you give yourself.¹²

Life is like a path along a mountain ridge; to left and right are slippery slopes down which you slide without being able to stop yourself, in one direction or the other. I keep seeing people slip like this and I say "How could a man help himself in such a situation!" And *that* is what "denying free will" comes to. That is the attitude expressed in this 'belief'. But it is not a *scientific* belief and has nothing to do with scientific convictions.¹³

Thus, belief in free will, much like religious belief, does not purport to express an empirical fact, but is rather expressive of an attitude, a mode of life, an imperative to live in a certain way. In the two lectures he delivered at Cambridge on freedom of the will, Wittgenstein went on to characterize belief in free will as 'groundless', as not supported by evidence or arguments¹⁴, again indicating parallels with religious faith.

To better appreciate this view of free will, one might introduce certain ideas from the *Philosophical Investigations* and, especially, *On Certainty*.¹⁵ In the former work, Wittgenstein speaks of our beliefs as founded upon a 'bedrock' certainty: "If I have exhausted the justifications [for following a rule] I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I do'" (*PI* 217). Similarly, in *OC* 341 Wittgenstein states, "the *questions* that we raise and our *doubts* depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn" (cf. *OC* 88). Although Wittgenstein refers here to *propositions* as acting as hinges, it is commonly thought that in the more mature phase of his epistemological work Wittgenstein thought of *practices* rather than propositions as primary. This, then, is no traditional foundationalism, where an inferential relationship is thought to obtain between the set of beliefs that comprise the foundations and the other beliefs we hold, with the former justifying the latter. Instead, for Wittgenstein what lies at the bottom or at the foundations of our language

games are not specific beliefs or propositions, but ‘ungrounded ways of acting’ (*OC* 110, 204), ‘communal practices’ (*OC* 128, 298), and ‘forms of life’ (*OC* 7, 358). It is in this spirit that Wittgenstein quotes from Goethe’s *Faust*: “In the beginning was the deed” (*OC* 402). Nonetheless, our practices and forms of life can be said to show or manifest the beliefs (or quasi-beliefs) and assumptions upon which we base our lives, including such beliefs as ‘I have two hands’ and ‘The world has existed for more than 10 minutes’. However, in opposition to G. E. Moore, Wittgenstein describes these as ‘certainties’ rather than ‘knowledge-claims’, for they are not grounded in evidence or open to verification, but express an attitude and a way of acting, and so are not true or false, reasonable or unreasonable, but simply “there—like our life” (*OC* 559, cf. 162, 205).

It would be arguably in keeping with this epistemological account, in conjunction with the later Wittgenstein’s remarks on free will, to say that belief in free will (and moral responsibility) may also function, at least in some contexts, as one of the bedrock certainties, as one of the things that ‘stand fast for us’ in our actions and practices (cf. *OC* 116), or as the framework within which our ethical practices operate and are made intelligible. There are close parallels here with existentialist philosophy, where to exist as a human being and to be free (almost) come to the same thing. Jean-Paul Sartre, for example, famously stated that we are ‘condemned’ to freedom, not free not to be free. In a similar vein, the Russian religious existentialist Nikolai Berdiaev, dubbed ‘the philosopher of freedom’, eschewed traditional accounts of freedom, which treat free will as an object that could somehow be perceived, investigated and proved or disproved from the outside, and adopted instead the Kantian position that freedom is a postulate of action: it is something we must presuppose to even think of a world in which human life and human agency are possible.

Wittgenstein would have been sympathetic to this outlook, for like the existentialists he is primarily concerned with concrete social and linguistic practices and seeks to provide a philosophical understanding of human existence that is not restricted to the explanatory framework of science (or even that of much traditional philosophy). For Wittgenstein, therefore, belief in free will, just as much as belief in God, is not threatened by scientific discoveries: “we couldn’t say now ‘If they discover so and so, then I’ll say I am free’.”¹⁶ In line with this view, Wittgenstein spends some time in his lectures on freedom of the will in attempting to show that even if a deterministic account of the world were demanded by our best scientific theories, belief in free will need not be affected at all.¹⁷ But he is not thereby putting forward a case for compatibilism: “All these arguments

might look as if I wanted to argue for the freedom of the will or against it. But I don't want to."¹⁸ Wittgenstein does not follow the traditional course of attempting to resolve the free will problem by proving one position or refuting another. His aim, as with other traditional philosophical problems, is to expose the problem as some kind of deep muddle or confusion arising largely from misunderstandings of the workings of language.

One recurrent criticism of this view is that it appears to render the language games of science, religion, and ethics entirely self-contained and cut off from each other, if not also immune from criticism from 'without'. This, indeed, is a common misconception of the Wittgensteinian account of religion, and in response Wittgensteinians such as D. Z. Phillips have emphasized the many important connections between religious and non-religious forms of life which (it is held) must be recognized if religious belief is not to degenerate into superstition. Similarly, Wittgenstein points out that scientific discoveries may have a bearing on ascriptions of free will: "A discovery might influence what you say on the freedom of the will. If only by directing your attention in a particular way."¹⁹ But despite these connections between the scientific and non-scientific domains, Wittgenstein insists that the distinctiveness of each language game must not be overlooked. In particular, the languages of ethics and religion must not be assimilated to the kind of fact-stating discourse and fact-finding activities that characterize the empirical sciences.

On this Wittgensteinian picture, conflicts in the free will debate arise at least in part because the central notions involved in these debates have been divorced from their original contexts where they serve a primarily practical purpose, and have instead been made the basis for philosophical theories such as libertarianism and compatibilism. One, indeed, wonders how this craving for generality and theory-construction would cope with some of the discourse on free will and responsibility found in religious contexts. Consider, for example, what Father Zosima has to say in Dostoevsky's *The Brothers Karamazov* (part 2, book 6): "In truth we are each responsible to all for all, only people don't realize it, but if they did, we should all instantly be in paradise!" This conception of responsibility seems worlds away from the analyses of responsibility usually found in the philosophical literature of the analytic tradition, although it is a prevalent theme in religious life and in some Continental writers, particularly Emmanuel Levinas (who develops an ethical form of phenomenology centred on the notion of 'responsibility for the Other', a responsibility which is taken to be both infinite and asymmetrical, thus making me responsible for the Other without the Other being responsible for me in turn).

But one does not need to stray too far from so-called ‘ordinary’ contexts to appreciate the diverse ways in which words like ‘freedom’ and ‘responsibility’ are used. And this, of course, is Wittgenstein’s point: ascriptions of free will are always embedded within a particular context, and it is the context that accords meaning to these ascriptions. Wittgenstein, for example, talks of voluntary movements as situated within a “normal *surrounding* of intention, learning, trying, acting”²⁰, and he notes that “there is a particular interplay of movements, words, expressions of face, as of manifestations of reluctance or readiness, which are characteristic of the voluntary movements of a normal human being.”²¹ In calling attention to these contextual features, and in comparing and contrasting a range of cases or examples drawn from daily life involving attributions of free will and responsibility, Wittgenstein’s purpose (as he describes it in *PI* 116) is “to bring words back from their metaphysical to their everyday use.”

No doubt much more could be said both in defence and against this Wittgensteinian account, but perhaps this will suffice to draw attention to a view whose promise and importance are not evidenced by the relative neglect it has been shown in contemporary discussions.²²

Notes

¹ P. F. Strawson, “Freedom and Resentment,” *Proceedings of the British Academy* 48 (1962): 1-25.

² See Harry Frankfurt, “Freedom of the Will and the Concept of a Person,” *The Journal of Philosophy* 68 (1971): 5-20.

³ See Galen Strawson, “The Impossibility of Moral Responsibility,” *Philosophical Studies* 75 (1994): 5-24.

⁴ See Susan Hurley, “Is Responsibility Essentially Impossible?” *Philosophical Studies* 99 (2000): 229-68.

⁵ R. J. Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1996).

⁶ All translations from the *Tractatus* are taken from the translation by D.F. Pears and B.F. McGuinness, published by Routledge in 1974.

⁷ Quoted in Georg Henrik von Wright, *Wittgenstein* (Oxford: Basil Blackwell, 1982), p.83.

⁸ Quotations from Wittgenstein’s *Philosophical Investigations* (hereafter ‘*PI*’) are taken from G.E.M. Anscombe’s translation, published by Basil Blackwell in 1972.

⁹ Wittgenstein, “Remarks on Frazer’s *Golden Bough*,” in Wittgenstein, *Philosophical Occasions: 1912-1951*, edited by James C. Klagge and Alfred Nordmann (Indianapolis: Hackett Publishing Company, 1993), p.119.

¹⁰ David Pears, *Wittgenstein* (London: Fontana/Collins, 1971), p.173.

¹¹ Wittgenstein, *Culture and Value*, edited by G.H. von Wright, translated by Peter Winch (Chicago: The University of Chicago Press, 1984), 64e.

¹² *Ibid.*, 61e.

¹³ *Ibid.*, 63e.

¹⁴ Wittgenstein, "Lectures on Freedom of the Will" (notes by Yorick Smythies), in *Philosophical Occasions*, p.438.

¹⁵ Quotations from *On Certainty* (hereafter "OC") are taken from the translation by Denis Paul and G.E.M. Anscombe (published by Harper Torchbooks in 1972).

¹⁶ Wittgenstein, "Lectures on Freedom of the Will," p.440.

¹⁷ *Ibid.*, pp.429-30, 436.

¹⁸ *Ibid.*, p.436.

¹⁹ *Ibid.*, p.441.

²⁰ Wittgenstein, *Zettel*, edited by G.E.M. Anscombe and G.H. von Wright, translated by G.E.M. Anscombe (Oxford: Basil Blackwell, 1967), §577.

²¹ *Ibid.*, §594.

²² For the record, it may be added that one of the editors, Nick Trakakis, has recently moved in this Wittgensteinian direction, and is therefore no longer entirely satisfied with the approach taken in the paper he has contributed to this volume.

DEFENDING HARD INCOMPATIBILISM AGAIN

DERK PEREBOOM

1. Hard Incompatibilism Characterized

According to the hard incompatibilist position I advocate, we would not have the sort of free will required for moral responsibility if determinism were true. We would also lack this sort of free will if indeterminism were true and the causes of our actions were exclusively states or events. If the causes of our actions were exclusively states or events, indeterministic causal histories of actions would be as threatening to this kind of free will as deterministic histories are. However, it might well be that if we were undetermined agent-causes—if we as substances had the power to cause decisions without being causally determined to cause them—we would then have this sort of free will. But although our being undetermined agent-causes has not been ruled out as a coherent possibility, it is not credible given our best physical theories. Thus we need to take seriously the prospect that we are not free in the sense required for moral responsibility.¹

I oppose a type of incompatibilism according to which the availability of alternative possibilities is the most important factor for explaining moral responsibility, and accept instead a variety that ascribes the most significant explanatory role to the way in which the agent actually produces the action. In metaphysical terms, the sort of free will required for moral responsibility does not consist most fundamentally in the availability of alternative possibilities, but rather in the agent's being the causal source of her action in a specific way. Accordingly, I advocate *source* as opposed to *leeway* incompatibilism. Agent-causal libertarianism is typically conceived as an incompatibilism according to which an agent can be the causal source of her action in the way required for moral responsibility, and thus proponents of this view are typically source incompatibilists. But a source incompatibilist might seriously doubt that we have the sort of free will required for moral responsibility, and this is

the position I defend.

But, in addition, I contend that a conception of life without this type of free will would not be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial. The type of free will that is undermined according to the hard incompatibilism I advocate is the kind required for moral responsibility in the following specific sense: for an agent to be morally responsible for an action is for it to belong to her in such a way that she would deserve blame if the action were morally wrong, and she would deserve credit or perhaps praise if it were morally exemplary. The desert at issue here is basic in the sense that the agent, to be morally responsible, would deserve the blame or credit just because she has performed the action, given an understanding of its moral status, and not, for example, by virtue of consequentialist considerations, or solely by way of a contractualist account. This is the sense of moral responsibility that has been at issue in the debate about whether the sort of free will required for moral responsibility is compatible with determinism. Other notions of moral responsibility have not been at issue: for example, the legitimacy of calling agents to moral account, that is, the legitimacy of demanding that an agent explain how an action might be in accord with moral principles, and if this fails, of demanding that the agent take steps to avoid similar behaviour in the future.² The hard incompatibilism I advocate takes no issue with this notion of moral responsibility, or with the characteristics of agency required for it.

Philosophers not infrequently take on the task of rescuing ordinary beliefs and practices from threats that result from scientific or naturalistic conceptions of reality. Such conceptions have posed a challenge to belief in the sort of free will required for moral responsibility and to the attendant practice of holding people morally responsible; and also, for example, to belief in God, in an immaterial soul, in immortality, and to theistic religious practice. While naturalistic philosophers have often given up God, the soul, immortality, and religious practice, they have typically not come to deny moral responsibility in the sense at issue in the debate, or its attendant practice. In the phrasing of Wilfred Sellars, they have not conceived of our moral responsibility, and the legitimacy of treating people as morally responsible, as a feature of the manifest image that has been undermined by the scientific image.³

I argue that, although denying that we are morally responsible in this sense has its cost to our ordinary self-conception, this cost is not as high as is often thought. We would need to reject the rationality of basic desert, of the reactive attitudes that presuppose basic desert, of retributive

justification of criminal punishment and personal recrimination, since all of this presupposes that we have the sort of free will required for moral responsibility in the sense at issue. What would survive untainted is the practice of calling each other to moral account, attitudes such as joy and sadness about what people do, justification for detaining criminals analogous to our rationale for quarantining carriers of dangerous diseases, and enjoyment of our achievements on a par with our enjoyment of our natural gifts. If we are careful to separate what in our conception of morality and meaning in life is undercut and what is not, we will see that we can live with what remains.⁴

2. A Defence of the ‘Tax Evasion’ Frankfurt-Style Case

Why opt for a source as opposed to a leeway position? I argue that examples of the kind devised by Frankfurt yield an effective challenge to the leeway position.⁵ In those examples, an agent considers performing some action, but an intervener is concerned that she will not come through. So if she were to show some sign that she will not or might not perform the action, the intervener would arrange matters so that she would perform it anyway. Here is one of John Fischer’s examples: Jones will decide to kill Smith only if Jones blushes beforehand. Jones’ failure to blush (by a certain time) can then function as the prior sign that would trigger the intervention that would cause her to kill Smith. Suppose that Jones acts without intervention. Here we might well have the intuition that she is morally responsible for killing Smith, even though she could not have done otherwise than to kill Smith, and even though she could not even have formed an alternative intention. She could have failed to blush, but Fischer argues that such a flicker of freedom is of no use to the libertarian, since it is not sufficiently *robust* to have a role in grounding the agent’s moral responsibility.⁶

Here is my earlier proposal on what it is for an alternative possibility to be robust⁷:

Robustness (1): For an alternative possibility to be relevant *per se* to explaining an agent’s moral responsibility for an action it must satisfy the following characterization: the agent could have willed something other than what she actually willed such that she understood that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action.

The intuition that underlies the proposal to ground moral responsibility in the accessibility of alternative possibilities is of the following sort: to be

blameworthy for an action, the agent must have been able to do something that would have precluded her from being blameworthy, at least to the degree she's blameworthy.⁸ Accordingly, for an alternative possibility to be robust, it must first of all satisfy this condition: the agent could have willed something other than what she actually willed such that by willing it she would thereby have been precluded from the moral responsibility she actually has for the action.⁹ Secondly, the epistemic element of Robustness (1)—that the agent must have *understood* that by willing otherwise she would have been precluded from the responsibility she actually has—is motivated by the following sort of consideration. Suppose that the only way Joe could have avoided deciding to take an illegal deduction on his tax form—a choice he does in fact make—is by voluntarily taking a sip from his coffee cup, for unbeknownst to him the coffee was laced with the drug that induces compliance with the tax code. In this situation, he could have behaved voluntarily in such a manner that would have precluded the choice for which he was in fact blameworthy, as a result of which he would not have been morally responsible for it. But whether he could have voluntarily taken the sip from the coffee cup, having no understanding that it would render him blameless in this way, is intuitively irrelevant to explaining whether he is morally responsible for his choice.

But here are two concerns for Robustness (1):

(a) One might imagine an agent who has alternative possibilities, where so acting would preclude the responsibility she has for the option she selects, but due to some epistemic failing on her part, she does not believe that she has an alternative possibility that meets this specification. Dana Nelkin (in correspondence) suggests a case in which an agent mistakenly believes that the alternative possibility does not preclude the responsibility she has for the option she selects, but she does recognize significant morally salient differences between the two options. One might propose that the agent has a robust alternative possibility partly because there are good reasons available to her for believing that she has an alternative in which her responsibility is different in the relevant way, even though she does not appreciate those reasons adequately, but only partially. But, first, imagine that Joe should have known what effect drinking the coffee would have, because he should have been paying attention when this fact about the coffee was revealed at his Tax Evaders Anonymous class. Does he, as a result, now have a robust alternative possibility? Not clearly, and I would say not. Note that denying that he has a robust alternative possibility does not preclude the advocate of a principle of alternative possibilities from assessing him as derivatively

responsible for evading taxes, for the reason that he may have met a relevant epistemic condition on derivative moral responsibility when he neglected to pay attention in the class. In addition, I am inclined to deny that such an epistemic failing supplemented by a mere partial understanding of morally salient differences between accessible alternatives is enough for robustness, given that the partial understanding does not amount to an understanding that availing himself of an alternative possibility would preclude the responsibility he actually turns out to have. Suppose that Suzy could have saved Billy from a painful death by giving him an additional injection, but that she has no understanding of this since she wasn't paying enough attention to the instructions when she should have been. But she was paying enough attention to understand that Billy would have been more comfortable had she given him the injection. My sense is that she does not have a robust alternative possibility in this case that would ground moral responsibility for allowing Billy to die, but still that she is perhaps responsible for allowing Billy to die derivatively from her not paying attention when she should have been—depending on the details of the case.

(b) In this example, is having a non-occurrent or even occurrent belief that taking a sip from the coffee cup *might* result in not evading taxes enough for robustness?¹⁰ It seems not. For, if asked, Joe might well agree that the probability of this connection is non-zero—he might admit, for instance, that it's at least .000001, and if he's taken a class in epistemology or probability, something like this might well be his response. But, intuitively, this is not sufficient to generate robustness. Should it be required for robustness that Joe understood that taking the sip of coffee would, with a probability of 1.0, result in his not evading taxes? This is clearly too strong, for it would intuitively be enough for robustness if he understood that the probability was, say, .95.¹¹ But the threshold probability, as one would expect, is difficult or impossible to determine. So here is my new proposal:

Robustness (2): For an alternative possibility to be relevant to explaining why an agent is morally responsible for an action it must satisfy the following characterization: the agent could have willed something different from what she actually willed such that she understood that by willing it she would be, or at least would likely to be, precluded from the responsibility she actually has.

Perhaps the most significant objection that has been raised against the earlier kinds of Frankfurt-style arguments was initially suggested by Robert Kane and then systematically developed by David Widerker and

Carl Ginet.¹² The general form of the Kane/Widerker/Ginet objection is this: for any Frankfurt-style example, if universal causal determinism is assumed to hold in the actual causal sequence that results in the action, the libertarian will not have and cannot be expected to have the intuition that the agent is morally responsible. If, on the other hand, libertarian indeterminism in this actual sequence is presupposed, the scenario will not serve the Frankfurt-defender's purpose, for any such case will fall to a dilemma. In Frankfurt-style cases the actual situation will feature a prior sign that signals the fact that intervention is not required. If in the proposed case the prior sign causally determined the action, or if it were associated with some factor that did, the intervener's predictive ability could be explained. However, then the libertarian would not and could not be expected to have the intuition that the agent is morally responsible. But if the relationship between the prior sign and the action were not causally deterministic in such ways, then it will be the case that the agent could have done otherwise despite the occurrence of the prior sign. Either way, an alternative-possibilities condition on moral responsibility emerges unscathed.

I have proposed a Frankfurt-style scenario that avoids this objection.¹³ Its distinguishing features are these: the cue for intervention must be a *necessary* rather than a sufficient condition, not for the action that the agent actually performs, but for the agent's availing herself of any robust alternative possibility (without the intervener's device in place), while the cue for intervention itself is not a robust alternative possibility, and the absence of the cue for intervention in no sense causally determines the action the agent actually performs. Here is the example:

Tax Evasion (2): Joe is considering claiming a tax deduction for the registration fee that he paid when he bought a house. He knows that claiming this deduction is illegal, but that he probably won't be caught, and that if he were, he could convincingly plead ignorance. Suppose he has a strong but not always overriding desire to advance his self-interest regardless of its cost to others and even if it involves illegal activity. In addition, the only way that in this situation he could fail to choose to evade taxes is for moral reasons of which he is aware. He could not, for example, choose to evade taxes for no reason or simply on a whim. Moreover, it is causally necessary for his failing to choose to evade taxes in this situation that he attain a certain level of attentiveness to moral reasons. Joe can secure this level of attentiveness voluntarily. However, his attaining this level of attentiveness is not causally sufficient for his failing to choose to evade taxes. If he were to attain this level of attentiveness, he could, exercising his libertarian free will, either choose to

evade taxes or refrain from so choosing (without the intervener's device in place). However, to ensure that he will choose to evade taxes, a neuroscientist has, unbeknownst to Joe, implanted a device in his brain, which, were it to sense the requisite level of attentiveness, would electronically stimulate the right neural centres so as to inevitably result in his making this choice. As it happens, Joe does not attain this level of attentiveness to his moral reasons, and he chooses to evade taxes on his own, while the device remains idle.¹⁴

In this situation, Joe could be morally responsible for choosing to evade taxes despite the fact that he could not have chosen otherwise.

The example does feature alternative possibilities that are available to the agent—his achieving higher levels of attentiveness to moral reasons. Indeed, at this point one might object that given that the intervener's device is in place, by voluntarily achieving the specified higher level of attentiveness Joe would have voluntarily done something whereby he would have avoided the blameworthiness he actually incurs.¹⁵ For had he voluntarily achieved the requisite level of attentiveness, the intervention would have taken place, whereupon he would not have been blameworthy for deciding to evade taxes. But this alternative possibility is not robust. Joe does not understand, and, moreover, he has no reason to believe, that voluntarily achieving the requisite level of attentiveness would or would likely preclude him from responsibility for choosing to evade taxes. True, were he voluntarily to achieve this attentiveness, the intervention would take place, and he would then not have been responsible for this choice. Still, Joe has no inkling, and has no reason to believe, that the intervention would then take place, as a result of which he would be precluded from responsibility for this choice. In fact, one might imagine that he believes that achieving this level of attentiveness is compatible with his freely deciding to evade taxes anyway, and that he has no reason to suspect otherwise. Nevertheless, Joe is morally responsible for deciding to evade taxes.

3. Robert Kane's Challenge to 'Tax Evasion'

Kane's reply to Tax Evasion crucially features the claim that the controller "is not going to let Joe make the undetermined choice between A and B,"¹⁶ where A is the choice to evade taxes and B is doing otherwise, and from this Kane concludes that Joe will not be (non-derivatively) morally responsible for the choice to evade taxes. His argument is this: if the cue for intervention, Joe's attaining the requisite level of attentiveness to moral reasons, does not occur, and he thus chooses A since the necessary

condition for choosing B is not in place, Joe's decision "will not be a 'will-setting' SFA [self-forming action]...because he will only have reasons to 'set his will' on A and will not have attended to any good reasons to set his will on B."¹⁷ If he does attain the level of attentiveness, the controller will intervene and make him choose A, and "so Joe will not get a chance to make a true SFA *either way* once the controller is in the picture."¹⁸

Thus the reason Kane cites for Joe's not being non-derivatively morally responsible is that he will not have the undetermined choice between A and B. Notice that he is contending that Joe is not morally responsible because he cannot do otherwise. More precisely, Kane is claiming that Joe is not responsible because he lacks plural voluntary control, and in the sense specified by this notion, a robust alternative possibility. One might protest that this is just the issue the leeway and the source theorist are arguing about, i.e., whether robust alternative possibilities are required for moral responsibility. In order to advance the debate, the source theorist devises a Frankfurt-style case in which the agent lacks robust alternative possibilities, but which is intended to elicit the intuition that he is morally responsible. What are we then to say of the response that the agent is not responsible because he lacks robust alternative possibilities?

It would be mistaken to say that Kane's response actually begs the question against the Frankfurt-defender.¹⁹ For the success of a Frankfurt-style argument depends on whether the audience finds it intuitive that the agent is morally responsible. As it turns out, Kane does not find it intuitive that Joe is morally responsible. For him, the ultimate reason is that Joe lacks alternative possibilities, and this view may, in the last analysis, be correct. Still, there is a respect in which this response to a Frankfurt-style case is unsatisfying, since it explicitly cites the leeway position on what is at issue as the reason why Joe is not morally responsible. To be sure, one can run a principle of alternative possibilities through any example, and then tally the results. But this procedure stands to miss the force of what might be a counterexample, and thus runs a serious risk of failing to engage an objection. Accordingly, it is at least *prima facie* dialectically unsatisfying. Moreover, this procedure precludes the possibility of discussing the issue at hand by way of Frankfurt-style cases. For we know in advance what, ultimately, the response to any such case will be: the agent is not responsible because he lacks robust alternative possibilities.

Like many philosophical discussions, the correct outcome of the debate about the principle of alternative possibilities should be viewed as a matter of reflective equilibrium.²⁰ On Frankfurt's side, we have the