THE CONTOURS OF FREE WILL SCEPTICISM



Simon-Pierre Chevarie-Cossette
University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Philosophy
Hilary Term 2019

The Contours of Free Will Scepticism

Simon-Pierre Chevarie-Cossette
University College, University of Oxford
A thesis submitted for the degree of DPhil in Philosophy
Hilary Term 2019
Primary supervisor: Prof. John Hyman
Secondary supervisor: Prof. Alison Hills

THESIS ABSTRACT

Free will sceptics claim that we lack free will, i.e. the command or control of our conduct that is required for moral responsibility. There are different conceptions of free will: it is sometimes understood as having the ability to choose between real options or alternatives; and sometimes as being the original or true source of our own conduct. Whether conceived in the first or in the second way, free will is subject to strong sceptical arguments. However, free will sceptics face a difficulty: doubting the existence of free will might itself be problematic, whether free will exists or not. Three main considerations are offered in support of this claim. Firstly, it is said that free will sceptics confuse contexts: even if they are right to doubt the existence of free will in the philosophy classroom, they are wrong to do so in other contexts, such as day-to-day interactions or a court of law. Secondly, it is said that free will scepticism is incompatible with inescapable and indispensable human practices such as practical deliberation or moral criticism and evaluation. For, to deliberate about what to do, we must presume that we have true options or alternatives. And to judge that some human practices are wrong, we need to assume the existence of moral responsibility. Thirdly, free will scepticism is sometimes said to be self-defeating, on the grounds that it cannot be rational to believe free will scepticism, because free will is required for rationality. This thesis critically assesses these arguments. None takes free will scepticism down, but all are instructive about how not to be a sceptic. They trace the 'contours' of free will scepticism by discrediting some conceptions of free will and some arguments for free will scepticism.

Word count: 74 230.

ACKNOWLEDGEMENTS

There are a number of people without whom writing this thesis would have been a much more difficult, lonely, and perhaps even futile enterprise. Instead of being forgetful, as I usually am, I will be selective.

I wish to thank my parents, Marc Chevarie and Christine Cossette, and my sister Justine Chevarie-Cossette, for being supportive and accepting of the fact that I was gone. Our weekly calls were invaluable.

I wish to thank all my philosophy colleagues and friends at Oxford, in particular Harry Alanen, James Kirkpatrick, Annina Loets, and Johanna Schnurr, for making me understand that writing philosophy is done best when it is recognised to be excruciatingly hard. I also wish to thank my philosophy friends from Québec, including Charles Côté-Bouchard, Marc-Kevin Daoust, Pierre-Luc Desjardins, Samuel Dishaw, Thiago Hutter, Thomas Mekhaël, and Charlotte Sabourin, who, on top of providing useful comments on my work, reminded me why I had chosen philosophy in the first place: because it is essential *and* profoundly pleasant.

I wish to thank my supervisors and mentors: Prof. Alison Hills, my secondary supervisor, whose incredibly welcoming supervision sessions were a great source of comfort and of fresh ideas; Prof. Karen Margrethe Nielsen, my teaching mentor, whose constant support helped me find my place in Oxford; Prof. Paul Lodge and Prof. Ian Philips, who gave me superb teaching opportunities.

Above all, I wish to express my gratitude to my primary supervisor, Prof. John Hyman, who has been so far my greatest living philosophical influence. I have immensely benefited from his demandingness and patience, from his attention to details and general insights.

Tanya Goodchild, my life partner, remains my best philosophical ally and fiercest critic, even though she is no longer my colleague. I will never be able to thank her enough.

CONTENTS

THESIS ABSTRACT III

ACKNOWLEDGEMENTS IV

CONTENTS V

INTRODUCTION 1

I. KNOWING ABOUT RESPONSIBILITY: A TRILEMMA 21

II. FREE WILL SCEPTICISM: CONTEXTS AND STANDARDS 50

III. SELF-DEFEATING BELIEFS AND MISLEADING REASONS 76

IV. IS FREE WILL SCEPTICISM SELF-DEFEATING? 100

V. FREE WILL SCEPTICISM AND THE MORAL LIFE 133

VI. SCEPTICAL DELIBERATIONS 163

VII. FROM COERCION TO MANIPULATION 207

VIII. EXCUSES AND ALTERNATIVES 246

BIBLIOGRAPHY 276

La première compagnie, dont l'origine remonte bien avant dans l'antiquité, est composée de gens qui vous disent nettement, qu'il n'y a ni allée, ni arbres, ni voyageurs ; que tout ce qu'on voit pourrait bien être quelque chose, et pourrait bien aussi n'être rien. Ils ont, dit-on, un merveilleux avantage au combat ; c'est que s'étant débarrassés du soin de se couvrir, ils ne sont occupés que de celui de frapper. Ils n'ont ni casque, ni bouclier, ni cuirasse ; mais seulement une épée courte, à deux tranchants, qu'ils manient avec une extrême dextérité. Ils attaquent tout le monde, même leurs propres camarades ; et quand ils vous ont fait de larges et profondes blessures, ou qu'eux-mêmes en sont couverts, ils soutiennent avec un sang-froid prodigieux que tout n'était qu'un jeu, qu'ils n'ont eu garde de vous porter des coups, puisqu'ils n'ont point d'épée, et que vous-même n'avez point de corps ; qu'après tout ils pourraient bien se tromper ; mais que le plus sûr pour eux et pour vous, c'est d'examiner si réellement ils sont armés, et si cette querelle, dont vous vous plaignez, n'est point une marque de leur amitié. […] Ces soldats sont bons pour les embuscades et les stratagèmes.

Une autre cohorte, non moins ancienne, quoique moins nombreuse, s'est formée des mutins de la précédente. Ils avouent qu'ils existent, qu'il y a une allée et des arbres; mais ils prétendent que les idées de régiment et de garnison sont ridicules, et même que le prince n'est qu'une chimère; que le bandeau est la livrée des sots, et que la crainte du châtiment actuel est la seule bonne raison qu'on ait de conserver sa robe sans tache. Ils s'avancent intrépidement vers le bout de l'allée, où ils s'attendent que le sable fondra sous leurs pieds, et qu'ils seront engloutis, ne tenant plus à rien, ni rien à eux. ('La Promenade du Sceptique', *Œuvres complètes*, Diderot 1875, 215–16)

***

The first troop, whose origin can be traced far back before Antiquity, is composed of people who tell you decidedly that there is neither path, nor trees, nor

> travellers; that everything one sees could well be something, and that it could also be nothing. They have, some say, a wonderful advantage in battle; it is that having rid themselves of the requirement to cover themselves, they have only experienced that of hitting. They have neither helm, nor shield, nor cuirass; but only a short double-edged sword which they wield with dexterity. They attack everyone, even their own comrades; and when they have dealt you large and deep wounds, or when they are themselves covered with them, they maintain with prodigious composure that this was all a game, that they had no intention to deliver you blows, since they have no sword and you have no body; that after all they could very well be wrong; but that what is safer for them and for you is to examine whether they are armed, and whether this quarrel, which you are complaining about, is actually a token of friendship. […] These soldiers are good for ambushes and ploys.
>
> Another cohort, just as ancient, although less numerous, has formed from mutineers of the former one. They confess that they exist, that there is a path with trees; but they pretend that the ideas of a regiment and a garrison are ridiculous and even that the prince is but an illusion; that the blindfold is the livery of fools, and that the fear of the actual punishment is the only good reason there is to keep their robe unstained. They walk daringly towards the end of the alley, where they expect the sand to melt before their feet, and to be submerged, neither caring about, nor cared for by, anything. ('The Sceptic's Walk', *Œuvres complètes*, Diderot 1875, 215–216; translation mine)*

*The 'regiment' refers to believers; the 'Prince' to God; the 'blindfold' to faith, the 'robe' to the sin ledger; and the 'alley' to life. Diderot is comparing two types of scepticism, scepticism about the external world and scepticism about God.

INTRODUCTION

1. Doubts about Control

Are we in command of our own conduct? There are two more precise ways to ask this question. The first concerns the degree of control that we exercise over our conduct; the second concerns whether we exercise control, or a certain kind of control, at all.

Regarding the first, studies in neuroscience and in psychology may give us a partial answer. They show for instance that many human actions are the result of implicit biases. We tend to be more lenient and permissive when the room we find ourselves in smells of lemon or of hot croissants. We tend to make negative judgements about those who are not part of our racial group, even if we strongly oppose racism. Research also shows that training against these tendencies or becoming aware of them is often not only inefficient—it can make things worse. There might also be certain domains where our control is limited. We often inherit the phobias and eating habits of our parents. From a few key socioeconomic features of someone's parents, we can predict their socioeconomical situation with non-negligible accuracy. These facts (and many others) might reveal that there are aspects of our lives over which we exercise very limited control. Alternatively, they might simply confirm

that our control over our conduct is more rarely *exercised*, and with greater difficulty, than we may imagine.[1]

This first way of posing the question about the control that we exercise over our conduct typically presumes or at least allows that we do in fact exercise some kind of control. It concedes that we make decisions; that our decisions and actions are guided by reasons or that they correspond to our values; it also allows that we choose between alternatives, and that when we have acted, or failed to act, or reasoned, or pursued certain values, we could have acted, omitted, reasoned, and valued differently from how we did. In a word, the scientific challenges are challenges not to the existence of our freedom or control, but to its extent or degree. Troubling as they might be, experimental results do not, at least most of the time,[2] constitute good grounds for a more general scepticism about control and freedom.

The second way of understanding the question concerns *whether* we exercise control, or a certain kind of control, or not. Of course, there *is* a kind of control that humans exercise: there is a difference between the adolescent who perpetually loses his temper and the stoic adult; between the animal who eats whenever he is fed and the peace activist who fasts for the greater cause. In this sense, adults can control their temper and

---

[1] On these discoveries and their philosophical import, see e.g. Nahmias (2014) and Walter (2011).

[2] An exception to this might be experiments of the sort that Libet conducted in the 1980's since they might suggest that our decisions are not causally linked to our actions. Interpreting these experiments is very difficult.

whether to give in to hunger, most of the time. But there is perhaps a troubling sense in which all humans are just like the adolescent or the animal. This is what the philosophical tradition—more than the scientific research—has suggested.

When we take seriously the fact that we are composed of matter that is governed by the laws of physics, laws over which we have no control, then it is tempting to be sceptical about our ability to choose between real alternatives. We can also come to think that every one of our actions or decisions traces back ultimately (even though this history remains to some extent a secret to us) to events which were not under our control and for which we were not responsible. And this gives us, again, great cause for concern about our freedom. For us to have been able to act differently and to be the ultimate source of our actions, it seems that the world would have needed to be different. Our conduct would have had to directly stem from our will; and our will, in turn would have had to be free in a sense that nothing besides the will itself explained its doings. But this picture is very difficult to make sense of. It is hard to distinguish a will that is genuinely free from a will that is governed by randomness. Thus, even if we exercise significant and important kinds of control over our conduct, it seems that we lack one crucial kind—the kind known as free will.

What is free will exactly? Philosophers sometimes conceive of free will as the possession of a form of 'leeway' or power—one has free will *just if* one has the ability and opportunity to do otherwise (see e.g. Kenny

1976; Vihvelin 2013; van Inwagen 2017). This is sometimes called 'liberty of indifference'. Free will is also sometimes conceived as a form of 'sourcehood' or causal relation—one has free will *just if* some of one's doings are the result of one's 'real' or 'deep' self (see e.g. Frankfurt 1971; Wolf 1993) or 'reason-responsive mechanisms' (see e.g. Fischer and Ravizza 1998; Sartorio 2016). This is sometimes called 'liberty of spontaneity'. To avoid verbal disputes, philosophers generally agree nowadays that leeway and sourcehood are two *conceptions* (or sometimes two aspects[3]) of free will.[4] They also generally admit that the *concept* of free will is that of the kind of command or control that *morally responsible* agents have over their conduct. So free will scepticism is the view that this type of command or control, and hence moral responsibility itself, does not exist, whether free will is regarded as a form of leeway or as a form of sourcehood.[5]

Why is free will scepticism interesting? It is a theoretical doctrine with practical and moral import. One might even say that it is a theoretical doctrine with a twofold practical programme. The first part of the programme concerns interpersonal relations. Since free will scepticism implies that no one is morally responsible for their conduct, it

[3] This view is held by Kenny (1976, 1978), who argues that liberty of indifference and liberty of spontaneity are two faces of the same coin.

[4] An exception to this is van Inwagen (2017), who prefers to use 'free will' to refer to the ability to do otherwise.

[5] These are the main conceptions of free will, but there can be others, e.g. the power to act for reasons.

also seems to imply that blame cannot be justified.[6] Other emotional reactions such as resentment and guilt might also be unjustified if we lack free will (see Pereboom 2001, 2014a). The second aspect of the sceptic's practical programme concerns institutions. Absent free will, punishing criminals cannot be justified, if justified punishment requires moral responsibility. Absent free will, protecting private property on the grounds that people *deserve* to enjoy the fruit of their labour may also be illegitimate.[7] There is great controversy about the extent of the human practices which are unjustified if free will scepticism is true. But as these specific practices are not the subject of this thesis, I will leave this stone unturned.

There are philosophers—so-called illusionists (see e.g. Smilansky 2000)—who argue that even if we discovered that we do not have free will, we should act as it if we *do* have it. But even illusionists accept that free will scepticism has some troubling consequences. In fact, illusionists accept that if we do not have free will, then we must live in a permanent state of make-believe. This surely is an important consequence; it is a cost of living in a world where free will does not exist.[8] The possibility

---

[6] This might also concern praise, although many believe that there is a difference between the conditions for justifiable blame and justifiable praise. See for instance Hyman (2016) and Wolf (1993).

[7] This is not to say that free will scepticism implies egalitarianism, as Hurley (2003) duly noted and argued.

[8] The only ones who might think that free will scepticism, even if it is true, is of no practical and moral consequences are the free will contextualists (see chap. II).

that we lack free will, if it is a serious possibility, should leave no one indifferent.

2. Scepticism: Some Varieties

Why use the term 'scepticism'? It typically refers to doubt or absence of knowledge rather than to denial or rejection. Free will scepticism might sound like the doctrine that we do not know anything about free will. However, several philosophers have acknowledged both uses of the word.[9] 'Free will scepticism' is itself an expression prevalent in the literature, both amongst those who declare themselves free will sceptics (see e.g. Pereboom 2001, 2014a; Caruso 2013) and amongst their opponents (Hart 1962, 160, 176–79; P.F. Strawson 1985; Nagel 1986, 89, 125, 2012, 27; Dworkin 1996, 88, 2011). The examples selected are intentionally diverse: 'scepticism' can mean eliminativism and 'free will scepticism' is not simply a technical term of the recent specialised literature on free will.

There are two philosophical reasons to make this semantic choice. The first is that one might recognise that whether we *doubt* or *deny* that free will exists, the consequences are troubling and significant. The word 'scepticism' reminds us of this fact. One might think, for instance, that if we entertain serious doubts about the existence of moral responsibility,

[9] Some have even gone a step further. Take Nagel: 'I think a view deserves to be called skeptical if it offers an account of ordinary thoughts which cannot be incorporated into those thoughts without destroying them. One may be a skeptic about *x* no matter how sincerely one protests that one is not denying the existence of *x* but merely explaining what *x* really amounts to.' (1986, 89)

then we would be imprudent to leave our practices intact. For it might seem that it is worse to wrongly assume that moral responsibility exists and hence retain a criminal system that embodies this assumption than to wrongly assume that moral responsibility does *not* exist and hence set in place a different criminal system. (This is not to say that since it is worse to condemn the innocent than to acquit the guilty, we should only prosecute people if we are sure not to *ever* make mistakes.) Therefore, scepticism about free will in both senses of 'scepticism' is a challenge to our system of moral responsibility.

There is a second and more interesting reason to use the expression 'free will scepticism', at least for my purposes. This links back to the epigraph, where Diderot compares two groups of sceptics. The first group mentioned by Diderot doubt the existence of the external world—they are *epistemic sceptics*, those who suspect that we do not know anything, or at least anything about the so-called 'external world'. The second group mentioned in Diderot's story doubt the existence of God. Diderot clearly links the two groups together: they travel the same life path ('*l'Allée des Marronniers*'); and the second group is composed of the mutineers of the first. Diderot's thought seems to be that people of a certain intellectual temperament first enter the philosophical life by applying their doubt excessively; only later do they become reasonable and healthily sceptical. In this way, they eliminate superstition, whether about God or about human freedom.

So, the second philosophical reason for using the word 'scepticism' is that this underscores a similarity between different forms of scepticism, as Diderot suggested. More precisely, I think that free will scepticism and epistemic scepticism share important features, which have been recognised on several occasions in the philosophical literature,[10] but which still hold out the promise of some undiscovered intellectual gains. My particular interest in this parallel is due to the fact that both scepticisms have been contested in similar ways. I shall mention four.

First, scepticism has been described as obviously false. This was illustrated by G.E. Moore (1939) when he raised his hands to prove that scepticism about the external world was false. What he claimed was that for any valid argument, philosophers have a choice. They can choose to accept the premises and the conclusion; or they can deny the conclusion and reject at least one of the premises. For scepticism you might think that the choice is easy; so easy in fact that it can be made before seeing any sceptical argument. We should simply reject scepticism and accept that any valid sceptical argument contains at least one false premise. No wonder then that epistemic scepticism was described as a pathology, one for which there was a prevention but no cures (Williamson 2000). Call this the *dogmatic* response to scepticism. Regarding free will scepticism, Peter van Inwagen (1983, 219–20), one of the most persuasive and

[10] See Strawson (1985), Nagel (1986), and Hawthorne (2001). The name of this section comes from Strawson's own book title, 'Skepticism and Naturalism: Some Varieties'.

adamant proponents of incompatibilism, claimed that if he discovered that determinism was true, he would endorse compatibilism.[11] This is a way of making a similar point to Moore: any argument that leads to scepticism must contain a false premise, whether it be 'determinism is true' or 'incompatibilism is true'. Our knowledge that we are responsible is more stable than our knowledge that any particular theory of free will is true.

Second, it has been argued that scepticism confuses contexts. This point is made by contextualists. *Contextualists about knowledge* claim that the propositions expressed by utterances of the form '*S* knows that *p*' may vary depending on the context of utterance (Lewis 1996). In the philosophy classroom, when we consider whether we have hands, it is true that knowledge is unattainable; but in our day-to-day life, we can know things. Similarly, *contextualists about free will* claim that the propositions expressed by utterances of the form '*S* freely φes' may vary depending on the context of utterance (Hawthorne 2001; Rieber 2006; Willaschek 2010). In the philosophy classroom, when we consider the free will problem, it is true that we lack free will. But outside of the philosophy classroom, for instance in a court of law, it would be confused to embrace either form of scepticism. If contextualism is right, then sceptics are like geometricians who claim stubbornly that tables are never flat. They are excessively rigorous.

[11] As Kenny noted (1976, 145), it is precisely because they wanted to keep freedom that some have endorsed indeterminism.

Third, scepticism has been considered self-defeating. Ludwig Wittgenstein (1974) claimed that a doubt that was meant to doubt everything would not be a doubt since the possibility of doubt presupposed the possibility of certainty, while Douglas Long (1992) argued that scepticism about the external world is self-defeating because if perception is generally questioned, then we lose self-knowledge and self-reference and therefore the possibility to doubt that we know. Similarly, free will scepticism stands accused of being self-defeating. Some (e.g. Epicurus, Popper 2012; Lockie 2018) have argued that rationality or reason requires the existence of free will. So, it is self-defeating to be a free will sceptic.

Fourth, scepticism is sometimes considered impractical, in the sense that it is incompatible with human action, planning, and practical or ethical life. In particular, it would first be impossible to deliberate and make decisions if we were true sceptics of either kind. Thus, we are told baffling anecdotes by Diogenes of Laertius about the sceptic Pyrrho of Elis who left his colleague Anaxarchus for dead in a swamp because of his sceptical doubts (D.L. IX 63). Diderot's sarcastic description of the sceptic[12] illustrates the alleged impractical character of scepticism. On the free will side, several philosophers (starting with Alexander of Aphrodisias) have argued that scepticism is incompatible with deliberation, or at least rational deliberation.[13] Scepticism is also

[12] This description of Pyrrhonism might be unfair. See Burnyeat (1984).

[13] See also Taylor (1966) and van Inwagen (1983).

considered incompatible with the moral life. More precisely, it is alleged that free will sceptics cannot fairly or coherently criticise others (Epicurus, *The Extant Remains*), that free will sceptics cannot judge that anyone's conduct is morally wrong (van Inwagen 1983) and that free will sceptics cannot explain the ethical significance of excuses (such as ignorance and compulsion) and exemptions (such as being a child). If any of this is true, free will scepticism is not practical.

3. A Lesson from Moore

The comparison between scepticism about knowledge and scepticism about free will is useful in identifying the challenges that await the latter. It is also useful in understanding my general project: what is the link between the accusation that free will scepticism is obviously false, that it is raised out of context, that it is self-defeating, or that it is impractical? I want to offer a putative answer here. It is controversial and difficult to give an argument for, but it is helpful to understand my overall approach.

To see this, let me briefly go back to Moore's dogmatic answer to epistemic scepticism. By showing his hands, Moore reminds us that any argument has 'two sides'. One can accept the premises of an argument and therefore its conclusion; or one can take the argument as a reason—and perhaps a *new* reason—against one of the premises of the argument. The second path is the better one here, according to Moore. He takes

the sceptical argument as a reason for thinking that he knows that there is an external world.[14]

But how can he claim this? His (externalist) posterity has provided a general answer: Moore can claim this because he is reliably or safely connected with the fact that he has hands. This answer seems dogmatic. For the question that is up for grabs is whether there is an external world. 'How do you know that there is an external world?', asks the sceptic. 'Because I am connected with the external world', responds the post-Moorean externalist.

The lesson that this interaction invites us to draw is that a dogmatic answer is the best that the sceptic can hope for. To refute scepticism, we do not have much choice but to raise our hands. This is where my more general point lies: proponents of certain doctrines cannot expect to systematically receive non-dogmatic answers. Sometimes, a doctrine is so radical that it is taken as reason—a new reason—to reject any argument in its favour (or to stop theorising[15]). This attitude might seem dogmatic, but, again, it is unclear that there is a way around it.

Interestingly, by being so blunt, Moore forces us to ask: to whom do we owe a non-dogmatic response? The answer cannot be 'everyone', otherwise the epistemic sceptic wins (this is controversial, but very plausible). The answer to the question might be 'no one'. Or it might be

[14] Obviously, some have made a different choice; see e.g. Nozick (1981) and Dretske (1970).

[15] For instance, if we had to choose between our knowledge of the external world and the validity of *modus ponens*, we might want to stop theorising.

'some, namely those who offer serious philosophical disagreement'. This is my preferred answer.

Who, then, offers serious philosophical disagreement? Not the epistemic sceptic, we might want to claim. Diderot's story says as much: we do not feel the need to refute the sceptic's arguments in order to decide against leading the sceptic's life. We may think—this is merely one possibility[16]—that those who offer serious disagreement satisfy three criteria. First, their doctrine is not contextually misplaced. Second, it is compatible with a range of fundamental and inescapable activities or goals of human life. Third, it is possible to believe in or act on it rationally or reasonably—i.e. it is not self-defeating. In this, the Moorean argument is a gateway to the three other challenges to scepticism about knowledge and about free will, namely the contextualist challenge, the impracticality challenge, and the self-defeat challenge.

Well, what about the free will sceptic? Does she offer serious philosophical opposition? This is the main question of my thesis.

4. The Contours of Free Will Scepticism

Working out how the free will sceptic can avoid the flaws (whether real or alleged) of epistemic scepticism is a recurrent theme in my

[16] I say that this is a mere proposal because I have neither the space nor the arguments to solve the dispute that dogmatists have with anti-dogmatists in epistemology. Dogmatists (see e.g. Lasonen-Aarnio 2014b) believe that dogmatism is warranted so long as we have knowledge (knowledge that our view is true or that our opponent's is false). Anti-dogmatists (see e.g. Wright 2007) try to show that certain arguments are such that even if we know their premises, we cannot put two and two together.

dissertation; it is the key to showing that free will scepticism is a serious theory. As Diderot maintained regarding atheism, scepticism about free will is revisionary and stems from sceptical tradition, but it is not the theoretical foil that epistemic scepticism is.

My dissertation is composed of eight chapters. Six directly address the four accusations which afflict both epistemic and free will scepticism. The two others are concerned with common reasons for being a free will sceptic.

The first chapter briefly discusses a question related to Moore's dogmatic response to scepticism. The dogmatic response consists in claiming that, since we know what the epistemic sceptic casts doubt on (e.g. that we have hands), we know that we are not in a bad case (that we are not brains in vats). Turning to the typical free will sceptical challenge, a Moorean response consists in claiming that since we know what the free will sceptic casts doubt on (e.g. that we are responsible agents), we know that we are not in the bad case (that we are not determined). In other words, I know that I am responsible, therefore I know that I am not determined. This chapter assesses critically this Moorean response.

The second chapter discusses the contextualist response to free will scepticism. This short chapter has two aims. The first is to show that contextualism fails to respond to free will scepticism for the same reason that it fails to respond to epistemic scepticism. The second is to argue that this teaches a lesson to the free will sceptic. It teaches the sceptic

that the sceptical challenge is not captured by the trivial claim that we are never 'completely free'. Differently put, the debate over free will scepticism is not about degrees of freedom.

Chapters three and four concern free will scepticism and self-defeat. Very roughly, a belief is self-defeating if it implies that the believer is irrational or unreasonable in believing it. For instance, it is self-defeating to believe that one is incapable of rational belief. The third chapter is not mainly concerned with free will scepticism. Instead, it tackles the error theory about reason, which claims that reasons do not exist. This theory is, if any, a good candidate for self-defeat. In arguing that it is in fact self-defeating to believe the error theory, this chapter is a good test case for my account of self-defeat, an account which I put to work in the following three chapters.

The fourth chapter tackles the view that it is self-defeating to believe in free will scepticism because free will is necessary for rationality or reason, and therefore for rational or reasonable doubt. If that is true, then the presumed truth of free will scepticism ensures that it cannot be believed rationally or reasonably. Why would rationality or reason require free will? I consider three arguments. They turn on considerations about reasons for action, obligations to act, and the relation between obligations and abilities. This long chapter critically assesses three more precise arguments. It shows that the sceptic might have to abandon some intuitive principles (such as "'ought" implies "can"') in order to avoid self-defeat.

Chapters five and six tackle the view that free will scepticism is impractical. The fifth chapter focusses on the idea that free will sceptics cannot participate in the moral life: they cannot criticise others; they cannot judge that some conduct in the world is morally wrong; and they cannot explain the ethical significance of excuses and exemptions. This chapter argues that in order to respond to such accusations, free will sceptics must admit the existence of responsibility in the sense of obligation or duty.

The sixth chapter discusses a second theme related to the impracticality response to scepticism: the question of whether free will scepticism precludes practical deliberation. This longer chapter refines the case that was made for the view that free will sceptics cannot deliberate; or at least that they cannot deliberate well or rationally, by investigating the aim of practical deliberation, namely making a decision.

All in all, I conclude that free will scepticism is a reasonable doctrine—much more so than its epistemic counterpart seems to be. It passes the test of contextualism (chap. II) and of self-defeat (chap. III–IV). But to do so, it needs either to abandon the view that we cannot do otherwise *or* the view that "'ought" implies "can'". As for the test of practicality (chap. V–VI), it depends on what kind of free will scepticism is at stake. I claimed earlier that there were two kinds of conceptions of free will, free will as leeway or indifference, and free will as sourcehood or spontaneity. If the free will sceptic understands free will as leeway or indifference, it fails the test of practicality. For, in deliberating, we must

presume that we have leeway or liberty of indifference. I suspect however that if the free will sceptic understands free will as sourcehood or spontaneity, then scepticism is compatible with practical deliberation. As for the moral part of the practicality test, free will sceptics can only pass it if they admit the existence of obligations.

If this general assessment is right, it appears that the non-sceptic cannot legitimately respond to scepticism dogmatically, *à la* Moore. At least this follows if dogmatic responses should be reserved for doctrines which fail to offer serious philosophical opposition, i.e. ones which pass the tests of contextualism, self-defeat, and impracticality.

Free will scepticism is alive and kicking. But it can only live within certain limits. The first six chapters of my thesis, even though they examine arguments which mostly fail, show that free will sceptics must accept and reject a number of important principles or notions to pass the three tests. Differently put, my thesis traces the *contours* of free will scepticism: it shows that certain forms of free will scepticism (e.g. leeway scepticism) and certain reasons for being a sceptic (e.g. that we never act for reasons) are too problematic to be maintained.

5. Further thoughts on Moral Responsibility

As I said, free will is typically defined as the command or control that morally responsible agents have over their conduct. And moral responsibility is often understood in terms of praiseworthiness and

blameworthiness. In the last part of my dissertation, I set aside scepticism and focus on blameworthiness. What does it require?

In the eighth chapter, I argue that blameworthiness is fundamentally interpersonal. I argue that considerations about coercion and manipulation show that we cannot always determine whether someone is blameworthy by looking only at the agent. Since blameworthiness is excluded by certain forms of coercion and manipulation, blameworthiness has some fundamentally interpersonal requirements.

In the ninth chapter, I argue that blameworthiness does not require leeway or the ability to choose amongst real alternatives. This claim became very popular after Harry Frankfurt published his seminal paper on the topic (1969). But it is also a claim that many have disputed, since many have rejected Frankfurt's counterexamples to the principle that moral responsibility requires alternate possibilities (PAP). My approach in the paper is to try to bypass the dispute over whether Frankfurt cases are counterexamples to PAP. Instead of addressing this question, I challenge the *reasons* we are thought to have for believing that PAP is true.

The content of the second, and more modest, part of this dissertation is thus more traditional. It is more traditional because it tackles two of the most discussed questions in the literature on free will: Frankfurt cases and manipulation cases. As such, it is quite different from the main part of my thesis.

Yet this second part is continuous with the rest of the dissertation in two ways. First, the considerations that it raises are important for free will sceptics. PAP is an important principle for free will sceptics because it allows sceptics to infer that no one is responsible (and blameworthy) for their conduct from the premise that no one has leeway, and manipulation cases are often used to support free will scepticism. Sceptics also sometimes argue that we are all structurally similar to manipulated agents. The second continuity between these two last chapters and the rest of my thesis lies in a common method. In assessing free will scepticism, I do not ask directly whether it is true, but I examine whether we can reasonably believe it and act on it. In assessing whether blameworthiness requires leeway, I do not ask directly whether it does, but whether we can believe that it does for good reasons. In assessing whether blameworthiness requires the absence of manipulation, I examine the methodological or epistemological assumptions made by those who say no. In a word, my thesis takes on free will with an epistemological twist.

I have not tried to settle the question of whether we have free will. But I think that most of what I have written suggests that free will scepticism is a reasonable philosophical stance. On the way, I have suggested lessons about a variety of philosophical topics: on Moorean arguments (chap. I); on the limits of contextualism (chap. II); on self-defeating beliefs and misleading reasons (chap. III); on the relationship between reason and free will (chap. IV); on the relationship between

responsibility and judgement (chap. V); on the nature and rules of rational deliberation (chap. VI); on manipulation and the interpersonal character of blameworthiness (chap. VII); finally, on our reason for believing that alternatives are required for responsibility (chap. VIII). The variety of these considerations proves yet again that scepticism is a fruitful philosophical project, whether we ultimately embrace it or reject it.

i. KNOWING ABOUT RESPONSIBILITY: A TRILEMMA

Forthcoming, in a revised version, in the *American Philosophical Quarterly*. (Chevarie-Cossette, Forthcoming)

Abstract: Here is a sceptical trilemma. We know that we are responsible. We do not know that determinism is false. Yet, if we do not know that determinism is false, we do not know that we are responsible. If we want to avoid the sceptical conclusion of rejecting the first claim, we ought to reject the third. It does not follow that compatibilism is false. But it follows that libertarianism, the view that we are responsible despite the fact that responsibility requires indeterminism, is unknowable, at least for now.

> *The best argument for compatibilism is that we know better that we are sometimes free than that we ever escape predetermination; wherefore it may be for all we know that we are free but predetermined.* (Lewis 2000, 109)

1. Introduction

Free will sceptics claim that moral responsibility does not exist, in the sense of basic desert or liability for praise or blame (see e.g. Pereboom 2014a).[1] As a consequence, they claim that we do not *know* that moral responsibility exists, since knowledge is factive. It is a rarer[2] thing to see them argue for this second claim directly, without invoking the factivity

[1] "Free will" and "moral responsibility" are different concepts, although it is common nowadays to define "free will" as the power necessary for morally responsible conduct (see Pereboom 2014a; cf. Fischer 1994). I prefer to talk in terms of moral responsibility to include in the discussion philosophers who suspect free will is unimportant.

[2] But see Byrd (2010, 412-13) and Kearns (2015) who argue explicitly in favour of this view.

of knowledge. But they could. After all, they are *sceptics* and sceptics like to doubt what others take to be knowledge.[3] They could say:

> I don't know that determinism is false. (A)
>
> If I don't know that determinism is false, I don't know that I am ever morally responsible for my conduct. (B)
>
> *Therefore*, I don't know that I am ever responsible for my conduct.[4] (~C)

Here, determinism refers to the view that every event is necessitated or determined by antecedent (or future) conditions together with the laws of nature (see Hoefer 2016).

This argument echoes the standard sceptical argument in epistemology: I don't know that I am not in the bad case (being systematically deceived by a Cartesian Demon, living in a deterministic world); if I don't know this, I don't know something I took for granted (that I have hands, that I am ever morally responsible for my conduct); therefore, I don't know that thing I took for granted. And, like the standard sceptical argument in epistemology, it suggests an inescapable trilemma: reject (A), reject (B), or accept (~C), that is reject (C).

In the debate about moral responsibility and determinism, there are three clans: the libertarians, who claim that moral responsibility exists and is incompatible with determinism; the non-sceptical compatibilists

---

[3] This is not a mere wordplay. Free will sceptics share a certain methodology with other sceptics, that is they do not revise their premises when these lead to revisionist conclusions.

[4] See Sehon (2013, 371) for the same argument about free will.

(henceforth compatibilists), who claim that moral responsibility exists and is compatible with determinism; and the sceptics, who claim that moral responsibility does not exist.[5] Rejecting (A) is the natural home of the libertarians. For if they know that moral responsibility is incompatible with determinism and if they know that they are responsible, they should be able to know that determinism is false. Rejecting (B) is the natural home of the compatibilists. For if determinism and responsibility have nothing to do with one another, it is natural to think that knowledge of the one will be unrelated to knowledge of the other. (As we shall see later, things are more complicated.) Rejecting (C) is, as we said above, the sceptic's take.

This argument has an advantage and a disadvantage for the free will sceptic. The advantage is that it bypasses difficult questions related to *indeterminism*. Sceptics do not need to assess whether indeterminism is compatible with responsibility; in particular, they do not need to show that there is no middle ground between chance and determination. The disadvantage is that it does not show that we are *not* responsible. Still, its conclusion is serious. If we do not know that we are responsible, can we really just act *as if* we were? Shouldn't we be more prudent regarding matters as important as responsibility?

---

[5] Again, these views could be defined in terms of free will rather than moral responsibility. For the first, see for example Chisholm (1964) and van Inwagen (1983); for the second, Strawson (1974) and Kenny (1976); for the third, Waller (2011) and Pereboom (2014a).

Some have responded[6] that we had better just act as if we were responsible because of practical considerations; because, without a belief in responsibility, our lives would be miserable.[7] But this option is not obviously right. The sceptical life is not unappealing;[8] and the thought of potentially living under a permanent and pervasive illusion is gloomy.[9] So, it is important to look elsewhere, that is to respond to the sceptic by rejecting (A) or (B).

It is worth emphasising why. We not only care about being responsible, but about knowing that we are. And we care not only about whether responsibility is compatible with determinism, but about whether we know this. In fact, some philosophers have argued that in general we should act in accordance with what we know (see e.g. Hawthorne 2004; Hawthorne & Stanley 2008, 578). Others have maintained that knowledge matters because it connects us with the facts (see e.g. Hyman 2015, chap. 8). At least, if we think that we are *ignorant* about responsibility, this should modify our conduct and attitudes (see Vilhauer 2009). So even though most of the literature on free will and moral responsibility focusses on moral or metaphysical questions such as, 'which theories of moral responsibility allow its existence and are these theories true?' we should start asking ourselves, 'which theories of

[6] See for instance, Smilansky (2000) and Nichols (2007).

[7] See Strawson (1974).

[8] See for instance, Pereboom (2014a), Waller (2011), Nadelhoffer and Tocchetto (2013).

[9] See Nozick (1974) and Wolf (1981).

moral responsibility allow knowledge of its existence and can we know these theories?'

The trilemma we just considered offers a humble but significant start to this project. In this paper, I consider in turn rejecting (A) and (B) and argue in favour of the second. For reasons of space but not of antipathy, I leave free will scepticism aside here—and thus the option of rejecting (C). Every non-sceptic, including the libertarians, is likely to accept that we do not know that determinism is false and nevertheless insist that we know that we are responsible. But this comes at a cost for libertarians, because it shows that their doctrine is unknowable given the actual state of science. If it were, we could come to know determinism without the help of physics. And that is absurd. So, by mapping out the available terrain for non-scepticism about free will, we discover that it is to the compatibilist's advantage. In § 2 and § 3, I consider the two options. In § 4 and § 5, I give extra arguments in favour of the second or against the first. In § 6, I come back to the parallel between the two scepticisms and consider an objection from van Inwagen (1983).

2. First Way: Knowledge of Determinism

The first way out of the trilemma is to reject (A): that is, to claim that we know determinism to be false. This, I hope to show, is impossible for us at the moment. As I have noted, this is a strategy which is mainly appealing for the libertarians, who claim that we are responsible even though incompatibilism is true. It is mainly appealing to them because

rejecting (C) constraints seriously the relevance of their doctrine; and rejecting (B) has bad consequences for them, which we will explore in § 3.

On what grounds can we reject (A)? Physics or other empirical sciences do not yet give us a clear answer about whether determinism is true. Some interpretations of quantum mechanics are indeterministic, some are not (see Hoefer 2016; section 4). Crucially, even if in the next few years, a clearer consensus emerges in favour of indeterminism, it will not follow that determinism is false *in the right way* for the libertarian. Let me explain. If indeterminism is required for responsibility, this is because at least some of our mental events are undetermined. Further, if indeterminism is required for responsibility, this is because *the right kinds of mental events* are undetermined (perhaps self-forming actions[10] or agent-caused actions[11]). If, for instance, emotions are undetermined, this does not seem to give us any reason to admit the existence of responsibility. Clearly, physics, biology, and psychology have not—and will not soon—reveal that undetermined events occur when these specific actions are performed. What this means is that *if* physics reveals that determinism is false, the sceptic can rephrase the trilemma by replacing 'determinism' with something like 'mental determinism'. So, let us ask again: on what grounds should we reject (A)? 'Science' is not an answer for now. Perhaps *one day* philosophers will be able to claim that they know that the

[10] Self-forming actions are at the centre of Kane's (1996) libertarian conception of free will and responsibility.

[11] See Chisholm (1964) and Pereboom (2014a).

right kind of indeterminism is true; and that this allows[12] them to solve our trilemma by granting (B) and (C) and rejecting (A).

Until then, the non-sceptics (in particular the libertarians) might be tempted to try something more audacious:

> I know that I am sometimes responsible. (C)
>
> If I don't know that determinism is false, I don't know that I am ever morally responsible for my conduct. (B)
>
> *Therefore*, I know that determinism is false. (~A)

'The libertarians may know their Moore',[13] we should reply, 'but they can't know that determinism is false without the physics, which, we have just pointed out, does not disprove (A) yet. Without scientific research, they might have *some reason* to believe in indeterminism; and perhaps it is even *metaphysically possible* for them to acquire what some have called "lucky" or "unreasonable" knowledge.[14] But they can't "go Moore" on this one. They lack sufficient scientific evidence that determinism is false.'[15] We should note that this response does not turn on the claim that we cannot discover an *a posteriori* truth like determinism from *a priori*

---

[12] Kearns (2015, p. 242) declares that he sees no reason for thinking that if I know I have free will (responsibility), discovering that incompatibilism is true could remove this piece of knowledge. I disagree: incompatibilism could be a defeater.

[13] I am referring here to G.E. Moore's (1939) infamous response to scepticism.

[14] See Lasonen-Aarnio (2010).

[15] Maier (2014) and Kearns (2015) have argued for this view. Coffman (2016) has replied, it seems, by exploiting the fact that sometimes we come to know things because we have not paid attention to defeaters. I don't think that this possibility is a cause for comfort for the libertarian.

considerations. For (C) is a good candidate for an *a posteriori* truth: we discover that we are responsible through our moral experience, whether emotional or not. (In § 6, I come back to the parallel between the Moorean argument against scepticism and this Moorean argument.)

The following reply might be made to the above argument. It is true that if the argument is valid, we can come to know that the *universe* (or its laws) is indeterministic without relying on any premises from physics. But we would discover this fact about the universe, the non-sceptic might argue, simply by identifying a counterexample to a generalisation. Similarly, if we discovered a British person who dislikes tea, we could infer something about the universe as a whole, namely, that it is a universe which contains a British person who dislikes tea (or perhaps that it is a 'disorderly' universe!). Or, more dramatically, when Europe woke up after the 1755 Lisbon Earthquake, they perhaps discovered that the universe was *imperfect* in virtue of discovering something about one of its remote corners. The illusion would be to think that one can't make a discovery about a whole because one has only made a discovery about a part, where the discovery about the whole is precisely that its parts do not *all* have a certain property.

Yet, there seems to be an important difference between the case of determinism and these other cases. The latter merely disproves a *generalisation*. But if the laws of nature are indeterministic, this is not merely to say that determinism is a mere generalisation, which turns out to be false. Instead, it tells against a *conception* of laws of nature, as simple,

unified, and necessarily binding or governing. And *this* seems to be a discovery that we cannot make simply by reflecting on the fact that we are morally responsible. That I am responsible for having forgotten the milk on my way home just cannot be a Galilean crux experiment (an experiment which disproves a system) against deterministic interpretations of quantum mechanics. If it was, it seems that scientists should start compiling cases where people have formed knowledge of responsibility to mount a case against determinism in the same way that Galileo mounted a case against Aristotelean physics by registering the spots on the Sun. So, we need to consider a different way to respond to our trilemma. (We will come back to rejecting (A) in § 5.)

The libertarian might reply that determinism is not a scientific hypothesis but a philosophical interpretation (and a generalisation of) some scientific theories or discoveries.[16] The idea then is that the libertarian indeterministic hypothesis does not falsify any scientific discovery but circumscribe how these discoveries can be interpreted (and generalised).[17]

This is an important point, but one which I believe overplays the separation of science and philosophy. First, in general, it is very strange that reflection on our own moral agency could give us any strong clues as to whether determinism is true or not, whether it is characterised as a

[16] This is suggested by Helen Steward.

[17] I am grateful to an anonymous reviewer of the *American Philosophical Quarterly* for raising this issue.

scientific theoretical claim or as a philosophical interpretation of a scientific thesis. Philosophy of science, just like science, should be independent of considerations about moral agency.[18]

Secondly, I am inclined to believe that determinism is *not* just a philosophical interpretation of scientific discoveries. This is because whether determinism is true or false seems to have some implications, although not straightforward ones, on what 'first-order' scientific facts we can observe. More precisely, whether determinism is true or false should in principle allow for *some prediction of human behaviour that indeterminism (of the libertarian kind) does not allow for*. Of course, indeterminism is compatible with *some* prediction of human behaviour and determinism does not imply predictability. Yet, determinism combined with certain further plausible assumptions allows that we could predict the behaviour of certain agents in equipoise, that is who are just as tempted/likely to Φ as to refrain from Φing. But this is not allowed in principle by the kind of indeterminism relevant to the free will debate.

3. Second Way: Ignorance of Incompatibilism

'Very well', responds the non-sceptic, 'give me a second shot':

I don't know that determinism is false. (A)

I know that I am sometimes responsible. (C)

[18] Even if determinism were not a thesis *in the philosophy of science*, it would still have important implications for it; but it would be troubling if moral considerations had important implications for philosophy of science.

> *Therefore,* if I know that I am sometimes responsible, it does not follow that determinism is false. (~B)

That is, (A) and (C) are compatible. This path is the only acceptable one for non-sceptics (those who want to preserve knowledge of responsibility). In this section, I will argue that, although it is open to libertarians, it implies that libertarianism is unknowable.

As I claimed in the introduction, this path is the home of compatibilists. Suppose that compatibilism is true. Then, since determinism and responsibility have nothing to do with each other, my knowledge that I am responsible does not stand in the way of my ignorance of determinism and vice versa. Is accepting this argument in fact only possible for compatibilists? Now, we touch on an important point. Some might want to argue that compatibilism *is* in fact the only way to explain how I know that I am responsible despite not knowing that determinism is false. This, they will argue, is clear in the lights of what we said earlier: we *cannot* know independently of science that determinism is false. They argue as follows. I know that I am sometimes responsible. But if I know this, either compatibilism is true or I *know* that determinism is false. I do not know that determinism is false; and I cannot come to know that determinism is false from non-scientific premises (like that I am sometimes responsible, and that compatibilism is false), as we argued in § 2. Therefore, compatibilism is true, or at least it is very unlikely that it is false. Call this the *Blunt Argument*. For clarity, let's restate it (asterisks explained below).

> I know that I am sometimes responsible, and I cannot* know that determinism is false.
>
> If I know that I am sometimes responsible and I cannot* know that determinism is false, then compatibilism is true.
>
> *Therefore*, compatibilism is true.[19]

The asterisks remind us of the fact that we cannot know that determinism is false, *so long as science does not reveal that it is*. The Blunt Argument seems to be what Lewis had in mind when he said, 'The best argument for compatibilism is that we know better that we are sometimes free than that we ever escape predetermination [that determinism is false]; wherefore it may be for all we know that we are free but predetermined.' (2000, 109)

The Blunt Argument fails because its second premise is false. From the fact that I know that I am responsible and that I cannot know that determinism is false, it only follows that *I cannot know that* (non-sceptical) compatibilism is false. Differently put, if I know that I am responsible and that indeterminism is unknowable*, then libertarianism is unknowable* to me.

This can be illustrated by two cases where determinism is false, although I am ignorant of this fact. In the first case, I have never heard of the problem of free will; I come to know that I am responsible for insulting my sibling, simply in virtue of noticing that I did this

[19] Kearns (2015, 242) suggests a similar argument. He claims that if incompatibilism is true, no one knows there is free will. He claims later that if he is wrong to claim that we cannot know that free will exists, then the compatibilist can use the Blunt Argument, in terms of free will (2015, 246).

voluntarily, out of choice—and perhaps in virtue of feeling guilty.[20] In the second case, I have encountered the problem of free will; but the possibility of determinism is too remote to prevent me from knowing that I am responsible; it is no obstacle to satisfying the safety condition on knowledge.[21] In each case, I know that I am responsible, I don't know that determinism is false and yet incompatibilism is true. We have the confirmation that the Blunt Argument fails.

But a subtle counterpart of the Blunt Argument is successful. Here it is. I know that I am sometimes responsible, and I cannot know that determinism is false independently of science. But I know that if I am sometimes responsible, either compatibilism is true or determinism is false. Suppose, for *reductio*, that I knew that compatibilism was false. Then I could come to know that determinism was false from these (non-scientific) premises. But I cannot. Therefore, it is implausible that I know that compatibilism is false, provided that I know I am responsible. Call this the *Subtle Argument*. Let's restate it:

> I know that I am sometimes responsible, and I cannot* know that determinism is false.
>
> I know that <If I am sometimes responsible and compatibilism is false, then determinism is false.>
>
> *Therefore*, I don't know that compatibilism is false.

---

[20] This example, but not the next, is compatible with an internalist framework in epistemology.

[21] *Safety* claims that *S* knows *p* only if there is no close possible world where *S* believes *p* falsely.

This argument proceeds by two applications of the principle of *single premise closure under known entailment,* that is, if I know that *p* and that *p* → *q*, and from these premises I competently infer that *q*, then I know that *q*.[22] (I will come back to this assumption in § 4.)

If we accept the Subtle Argument, we must accept that libertarianism is *unknowable**, that is unknowable provided that we make two restrictions. The first restriction is important, but unsurprising: we cannot know that compatibilism is false so long as physics has not shown that determinism is false. (So, we cannot* know that incompatibilism is false.) The second restriction is perhaps surprising, but unimportant: the conclusion is perhaps restricted to those who have made the relevant inferences. I say that this is unimportant because it is no comfort to libertarians that they could only know their theory by refraining to read their rivals' papers.

That libertarianism *might be* knowable* to those who fail to make the relevant inferences should be no consolation for the libertarian. To make this point clearer, consider the following:

> I cannot* know that determinism is false.
>
> I *must* know that <If libertarianism is true, determinism is false.>
>
> *Therefore*, I cannot* know that libertarianism is true.

---

[22] PROOF. From the two premises, I first infer that <if compatibilism is false, determinism is false>, which I come to know because of closure. But then, since I do not know that determinism is false—and could not come to know it in the process—I can infer that I do not know that compatibilism is false.

Suppose for *reductio* that I can know that libertarianism is true. Then, by this and the second premise, I can know that determinism is false. But I cannot know this, *qua* the first premise.

It is good to pause for a moment to see the epistemic difference between the proponent of the Blunt Argument and of the Subtle Argument.[23] The proponent of the Blunt Argument argues that compatibilism is true; the proponent of the Subtle Argument argues that libertarianism is unknowable*. The difference looks obvious, but it is not. It is not obvious, because the premises of each argument narrow the difference between their friends. Even though they agree with its conclusion, the free will sceptic and the free will agnostic are not friends of the Subtle Argument since they reject its first premise.

Who, then, is the friend of the Subtle Argument but not of the Blunt Argument? First, this friend might be a compatibilist. She believes, *qua* the argument, that libertarianism is unknowable*; and she also believes (perhaps because there is no converse argument) that compatibilism is knowable*. And yet, having realised that the Blunt Argument fails, this friend of the Subtle Argument is aware that the Subtle Argument in itself gives her insufficient reasons to believe in compatibilism. Second, she might be an incompatibilist. She would then believe, *qua* the Subtle Argument, that libertarianism is unknowable*; and she might believe also that compatibilism *would be knowable* if true*. But she would take it that

[23] I am grateful to an anonymous reviewer of the *American Philosophical Quarterly* for raising this issue.

incompatibilism is true, though unknowable*, and so that compatibilism is also unknowable*.

These views are both coherent and appealing—especially the first. From the facts that libertarianism is unknowable* and that it is incompatible with compatibilism, we cannot infer that compatibilism is true, even if we grant that compatibilism is also knowable* if true. To see this, take something close to Russell's Teapot proposition:

> TEAPOT: There is a teapot revolving around a Star in some other galaxy than ours.

If true, TEAPOT is knowable*: if we found such a teapot, we could know it exists. But propositions which entail the falsehood of TEAPOT such as <there are no tea-related objects in other galaxies> are also unknowable*—we would need to scan the whole universe to know them. And yet, these facts provide no good reason to believe TEAPOT. Now, both the compatibilist and incompatibilist friends of the Subtle Argument believe that compatibilism is just like TEAPOT. Compatibilism is knowable* if true; yet libertarianism, a proposition which entails that compatibilism is false, is unknowable* even if true. But neither these remarks nor the Subtle Argument provide good reason to believe in compatibilism.[24]

Let's recap. We considered in this section whether we could respond to the sceptic's trilemma by accepting that I can know that I am

[24] I am very grateful to Kolja Keller for this parallel.

responsible even if I do not know that determinism is false. Having shown that the Blunt Argument is unsound, we can claim that this path is opened to all non-sceptics: not just to non-sceptical compatibilists but to some libertarians as well. But having shown that the Subtle Argument is sound, we can claim that this path is costly for libertarians: for they must accept that their view is unknowable*. This conclusion is serious and thus we should be ready to examine all the alternatives, even if they are unpopular.

4. Third Way: Rejection of Closure

In this section, I want to show that rejecting single premise closure under known entailment (henceforth 'closure') is no third way for the libertarians. In particular, libertarians cannot claim that they know that libertarianism is true and yet that they do not know that determinism is false, on the grounds that knowledge is not closed under known entailment. We may note that even though rejecting closure is at times tempting, there are strong philosophical sentiments on both sides of the internalist-externalist aisle in favour of it. But more importantly for our purpose, there is a distinct problem for libertarians in rejecting closure. If they abandon closure, they need to replace it with something else. But the possible replacements—the two anti-closure accounts of knowledge—are not libertarian-friendly.

Fred Dretske (1970) offered a first anti-closure account of knowledge, *error ruling-out*. It features a principle according to which I

know a proposition only when I rule out the relevant possibilities of error. Consider the following classic scenario. I am in a zoo and observe a zebra in a pen. I know that this is a zebra. And I know that if this is a zebra, this is not a cleverly disguised mule. Yet, I can't know that this is not a cleverly disguised mule. If, by contrast, I had been in a dodgy zoo in Cairo,[25] I would have perhaps not even been able to know that the animal in front of me was a zebra, even if it were. For, then, the possibility that it be a cleverly disguised mule would be a relevant possibility of error.

Many have rejected Dretske's anti-closure account of knowledge, but we do not need to go there for our purpose. All we need is to see that his account cannot be of any help for our third-way libertarian. Our third-way libertarian claims that she knows that she is responsible and yet that she does not know that determinism is false. On Drestske's account of knowledge, this is possible only if determinism is not a relevant possibility of error. Well, determinism is perhaps not normally a relevant possibility of error: on Dretske's account, if I break a wine glass while carelessly rinsing it, I can perhaps know that I am responsible without ruling out the possibility that my action was determined. But when I form the belief that *libertarianism* is true, determinism *is* a relevant possibility of error. This is simply because determinism is part of the *definition* of libertarianism. By contrast, not being a cleverly disguised mule is not part of the *definition* of being a zebra (and not being

[25] For real cases, see https://www.bbc.co.uk/news/world-africa-44968509.

determined is not a part of the *definition* of being responsible).[26] Therefore, on Dretske's account, no one can come to know that libertarianism is true without having ruled out determinism; and since this is impossible independently of science, on Dretske's account libertarianism is unknowable independently of science. So Dretske's account leads the 'third-way' libertarian to my conclusion.

Robert Nozick (1981) presented a second anti-closure account of knowledge, *truth-tracking*. It features a principle according to which I know a proposition only if I would not believe it if it were false. Yet again, this is of no help for the third-way libertarian. If libertarians lacked moral responsibility, they might have still believed that they were responsible.[27] And if libertarianism had been false, they would have still believed it. Thus, libertarians do not know that they are responsible, and they do not know that libertarianism is true. It's best to keep closure.[28] (Remember that truth-tracking does not allow us to discard remote worlds. It is precisely why it implies that closure fails.)

On top of what I have just said, there is an independent reason for the libertarians to keep closure. In fact, the Consequence Argument,

---

[26] There might be trickier cases, brought up by Dretske's view that we don't know a plateful of cookies is a plateful of mind-independent objects.

[27] This counterfactual is hard to assess; and this difficulty tells against the claim of the third-way libertarians according to which they know they are responsible.

[28] See Sehon (2013, 375-379), who makes a similar argument about several theories of moral responsibility.

probably the main incompatibilist argument ever championed, uses a principle which resembles closure:

> β: If *S* has no control over *A* and that *A* necessarily implies *B*, then *S* has no control over *B*.

β is intuitive, but those who find the Consequence Argument appealing typically invoke the fact that other closure principles as evidence for β (see Fischer (1994)).[29] This makes sense if having control is understood as a kind of possibility; if, for instance, having control over a state of affairs just is for this state of affairs to be agentially possible.[30]

In the last section (§ 3), we identified a significant conclusion:

> It's possible that I know that I am responsible and that I don't know that determinism is true. (~B)

It does not follow that compatibilism is true, as the Blunt Argument would have it. Rather, it follows that I cannot know that libertarianism is true, independently of science. And we confirmed this conclusion in this section (§ 4) by showing that the libertarian could not reject closure.

5. Knowledge of Determinism (Again)

Having granted closure, we come back to the Subtle Argument and its devastating conclusion for the libertarians. Its weakest point is the claim that I cannot know that determinism is false independently of science.

[29] In a similar spirit, Dretske (2005) argues against closure under known entailment by arguing that other closure principles (memory closure, proof closure, perception closure, etc.) are false.

[30] See List (2014) and Kenny (1976) for a discussion.

This takes us back to the strategy that we rejected in § 2: accepting that we can come to know that determinism is false on the grounds that we know that we are responsible.

I do not have much more to say about the unlikelihood of this claim but let me bring up three dialectical points which count against this strategy. The first point is that we are in a context where compatibilists and incompatibilists seem in a deadlock.[31] On both sides, there are convincing arguments. It is true that incompatibilist arguments are persuasive, but not comparatively enough to warrant that we disregard the counter-intuitive consequence that we could come to know that determinism is false without the help of physics. Differently put, the libertarians would need to be in a much stronger position to accept the audacious claim that they know that determinism is false independently of science.

The second dialectical point is that it would be disingenuous on the libertarian's part to champion the conclusion that we know that determinism is false as a surprising though necessary anti-sceptical response. It is a surprising response, but not a necessary one: we can reject (B) instead; and we can do so without endorsing compatibilism. All we need is to accept that we cannot know libertarianism. It does not mean that libertarians should stop fine-tuning their arguments—they

[31] It is this deadlock that leads several philosophers to agnosticism (see Double 1996; Byrd 2010; Kearns 2015).

might come handy if one day we discover that determinism is false. Or, they might make it harder for others to embrace compatibilism.

The third dialectical point is that we can explain how it is that convincing libertarian arguments cannot give us knowledge of determinism. Convincing libertarian arguments are grist to the sceptic's mill. It is, therefore, no wonder that we cannot come to know that determinism is false on the basis that (1) we are sometimes responsible and (2) incompatibilism is true. For, absent our knowledge that determinism is false, (1) and (2) are evidence against one another. They are evidence against one another because if incompatibilism is true, it is harder to be responsible (some of our actions need to be undetermined); and if we are responsible, it is less likely that incompatibilism is true (since it is not any less likely that determinism is true). As such, incompatibilist arguments are, for libertarians, a double-edged sword: the better the arguments, the more fragile their knowledge that they are responsible.

6. An Objection from Epistemic Scepticism

Let me consider a final and important objection to what I have said so far. I have argued that we cannot know that determinism is false on non-physical grounds. But I have also made a clear parallel in the beginning between the sceptical argument regarding free will and regarding knowledge. Remember: in both cases, I don't know that I am not in the bad case (being systematically deceived by a Cartesian Demon, living in a

deterministic world); if I don't know this, I don't know something I took for granted (that I have hands, that I am ever morally responsible for my conduct); therefore, I don't know the thing that I took for granted.

This exact parallel was used by Peter van Inwagen to argue for the view that we can come to know determinism without the help of science. His argument is that if we accept that we cannot come to know that determinism is false on moral grounds, by parity of reasoning, we should accept that we cannot come to know that we are not being universally deceived by a Cartesian demon on empirical grounds. But this is unacceptable:

> You say that our having good reasons for believing in moral responsibility is not the sort of thing that could provide us with good reasons for believing that determinism is false. But you *also* say that our having good reasons for believing that our knowledge-claims are generally correct *is* the sort of thing that can provide us with good reasons for believing that there is no Universal Deceiver. But isn't it true that you are responding to philosophical problem that are essentially the same in arbitrarily different ways? (van Inwagen 1983, 213)

This objection is ingenious, for it is hard to resist the analogy (for a very similar objection, see Steward forthcoming). Coffman (2016, 89–90) offers three possible responses; let me mention the two more simple ones. The first response is that we know directly that we are not deceived by a Cartesian Demon, but not that determinism is false. This means that we do not come to know that we are not deceived on the

basis of a problematic Moorean inference. Therefore, we can cast doubt on the validity of the Moorean inference from responsibility to indeterminism without having to worry that this leads to epistemic scepticism. The second response is that we have abductive evidence for the claim that we are not being deceived (we can explain our world better with the non-sceptical hypothesis), but we have no such abductive evidence for indeterminism. There is no data that is left unexplained without the hypothesis of indeterminism. So, again, we do not come to know that we are not deceived from a problematic Moorean inference and can therefore cast doubt on the validity of the Moorean inference from responsibility to indeterminism.

I remain unconvinced by these suggestions, which are epistemologically controversial. (If they are right, though, my conclusion stands.) I prefer a different solution. Of course, one way to respond to van Inwagen's objection is to deny closure; and go back to our argument that libertarians do not cope very well with this the denial of closure. But this is the epistemic nuclear option; and I prefer not to go there. Instead, I want to insist that I *do* treat symmetrically the two philosophical problems under consideration up to a point (that of coming to know that I am not determined and that of coming to know that I am not systematically deceived).

Take the epistemic scepticism case. I know that I have hands; and if I know that if I have hands, then I know that I am not being universally deceived; therefore, I know that I am not being universally deceived

(perhaps by a Cartesian demon lurking at the end of the universe). There is no doubt that this conclusion comes as a surprise. How could I know this? We should not discount this surprise because we are used to taking the hard-line against scepticism. If we take our surprise seriously—the surprise of coming to know something about a Cartesian deceiver at the end of the universe from our sofa—we should say this: the fact that we infer something so abstrusely is *evidence* that we got something wrong in our deduction.

Compare: suppose that during dinner I get in an argument about whether Algerian cuisine offers the best African food. Imagine that I somehow end up inferring that Algeria is the largest country in Africa, *from purely culinary premises*. I should take this as *evidence* that I made a bad inference along the way or that I said something false. Perhaps I slipped from 'dessert' to 'desert' and inferred that since Algeria has the most impressive des(s)erts in Africa, it is the largest African country. We might want to admit a rule of thumb like the following: if we infer a proposition of kind *X* from a proposition of kind *Y*, then the more (significantly) different *X* and *Y*, the stronger our suspicion should be about the inference.

Suppose I face the sceptical hypothesis. I consider my options. First, I can abandon the claim that I know that I have hands—but this is a no

go.[32] Second, I can abandon the claim that *if I know that I have hands, I know that I am not deceived*. But the only way to do this is to abandon closure. And I gave reasons in § 4 for setting this option aside. I have one option left: accept that I know that I am not being deceived. And, this, it seems, I can only know (this is what Coffman would deny) in virtue of a Moorean inference. But it does not mean that I need to accept the Moorean inference 'as if it was nothing', that is as if it was not epistemically strange to come to know things about deceivers from my knowledge that I have hands. It *is* strange to come to know something about a universal deceiver in this way. But this remains my best option against scepticism. Again, the strongness of the inference gave me pause before accepting closure. But I see now that I need to admit the Moorean inference, despite its strength.

Now, suppose I face the sceptical hypothesis about free will. I consider my options. First, I can abandon the claim that I know that I am responsible—but this is a no go (according to our working assumption). Second, I can abandon the claim that *if I know that I am responsible, then I know that determinism is true*. This can be done, again, by abandoning closure, but we have set this aside. Now, here lies the difference between the two sceptical challenges: in the case of free will but not of knowledge, there is another way to abandon the second claim than to abandon closure. As we have argued, we can simply endorse that

[32] This seems true for multiple reasons: because it is obvious, because trying to doubt this proposition would not be a real doubt since it is a hinge proposition, because it self-defeating to doubt such a proposition.

incompatibilism is unknowable*. Now, *this* is a good option. It is a better option than the third, which is to accept that I can come to know that determinism is false without the help of scientific inquiry.[33]

I think that what makes van Inwagen's argument so appealing is the proneness that we all have to reject epistemic scepticism. We see it and our philosophical training convinces us that we must be able to refute it. Therefore, we conclude that Moorean inferences are perfectly unproblematic. We remain blind to the fact that usually, when we make one such surprising inference, we did something wrong. (Although, not always, as the case of epistemic scepticism confirms.)

7. Conclusion

I have presented an unusual sceptical challenge, according to which since we do not know that determinism is true, we do not know that we are responsible. Like many sceptical arguments, this one reveals the epistemological commitments of the non-sceptics. It shows that we can know that we are responsible despite not knowing that determinism is false. It would be too hasty, as I have argued *pace* the Blunt Argument, to conclude that compatibilism is true. Rather, we must accept that, given the present state of science, libertarianism is unknowable. It is still an open question whether, now that we are aware of this, we can maintain a

[33] Ralf Bader asks whether idealism might play the role that compatibilism assumes in the knowledge case? If my trilemma guides us away from libertarianism, should the sceptical trilemma guide us away from realism? I find compatibilism more palatable than idealism.

reasonable belief in libertarianism (perhaps for pragmatic reasons, as James (1912) suggested). It is also an open question whether we can reasonably believe in compatibilism or in free will scepticism. Some philosophers (e.g. Epicurus, *The Extant Remains*; Kant, *Groundworks* 4:448; Lockie 2018) have argued that these views are self-defeating to believe in. I do not think that these arguments are sound, but they should be taken seriously (see Chevarie-Cossette 2019, forthcoming).[34]

Might the incompatibilist retaliate by showing that a mirror image of the Subtle Argument[35] afflicts compatibilism? If this were the case, this would show that non-sceptical compatibilists and libertarians are in the same boat. I think however that most compatibilists would agree that there is no such argument, on the grounds that moral responsibility does not imply the truth of determinism. Admittedly, the view that it *does* commit one to determinism has had its defenders—possibly Hume (1739) and Hobart (1934). And compatibilists tend to overplay how destructive indeterminism would be for freedom when discussing libertarianism. Until physics tells us that determinism is *true*, this form of compatibilism is unknowable— just like libertarianism. For, otherwise, we could come to know *determinism* on non-scientific grounds. In any

[34] I am grateful to anonymous reviewers of the *American Philosophical Quarterly* for raising these issues.

[35] Kearns (2015, 241) explores a similar parallel, when he argues that if compatibilism is true, we cannot know that we have free will. I think the argument fails because we could discover that compatibilism is true by discovering that determinism is, assuming that we know we are responsible. Also, if there is a gap between philosophical or ethical considerations and scientific considerations, this is *strong evidence* for compatibilism.

case, we should be prudent before requiring the cooperation of the microphysical world for us to be responsible.

ii. FREE WILL SCEPTICISM: CONTEXTS AND STANDARDS

Abstract: Most people who encounter the problem of free will end up leaving it where they found it: in the philosophy classroom. Contextualists argue that this natural tendency is correct. We should doubt the existence of free will in the philosophy classroom, but not in normal contexts. This, they say, is due to a variation of the standards of application of the word 'free' and its cognates across contexts. So, in the contexts which matter such as an interpersonal conflict, a trial, or a legislative assembly meeting, free will scepticism is incorrect; thus, it is irrelevant to our moral practice. I argue that contextualism is of no help against free will scepticism. On the one hand, contextualism per se does not meet the sceptical challenge at all. On the other hand, the conciliatory touch that contextualism is thought to bring to anti-scepticism is mostly unhelpful. Yet contextualism offers an interesting lesson on how to understand the sceptical challenge. The sceptical challenge is not a request to set the bar higher for freedom; therefore, it cannot be answered by the concession that no one is ever completely free.

1. Introduction

Are we free in the sense required for moral responsibility? *Free will sceptics* believe that we are not, and in support of this they champion several persuasive arguments (see e.g. Strawson 1994; Pereboom 2014a). Like its epistemic counterpart, free will scepticism is not the majority view. Yet, again like its epistemic counterpart, it is difficult to refute. It is therefore tempting to try to *dissolve* the dispute between the sceptic and the non-sceptic as follows. The dispute, we could say, starts on the wrong foot: it assumes that 'free' or 'free will' has one correct application. In reality, whether someone is free (or

whether they know) changes according to the context of utterance; different contexts of utterance apply different standards.[1] The philosophical context applies sceptical or high standards; as a result, freedom or free will (the difference does not matter for our purpose[2]) does not exist in it. But other contexts apply non-sceptical or low standards; as a result, freedom or free will exists in them. Thus, both the free will sceptic and the free will believer are sometimes right and sometimes wrong.

This is, very roughly, the *contextualist* strategy. It might look like a conciliatory move, but in fact contextualists typically take themselves to refute or at least 'deal with' scepticism about knowledge or about free will. Here is why: if contextualists are right, then sceptical doubts, while relevant in the context where they are normally raised (in the philosophy classroom), are irrelevant in the contexts that count (in day-to-day life, in a court of law, in a legislative assembly). Confined in the same ivory tower, the two kinds of sceptics fail to pursue an interesting revisionary doctrine.

We can regiment this anti-sceptical approach as follows:

> (1) Free will sceptics use the term 'free' or its cognates with higher standards than compatibilists do.
>
> (2) Political theorists, jurists, and people in their day-to-day interactions ought only to be concerned with the compatibilist standards of the term 'free' or its cognates.

---

[1] I prefer to talk in terms of standards to avoid difficult questions such as 'what sort of variations in meaning imply a *new sense*?'

[2] It does not matter because I will not pay attention to the notion of a *will*; and it is clear that by 'freedom', I do not mean political freedom.

> *Therefore*, political theorists and jurists ought not be concerned with the sceptical standards of the term 'free' or its cognates.

So, the contextualist aims at showing that the disagreement between the sceptic and the non-sceptic is a matter of stringency of standards or of degrees and that this is to the non-sceptic's advantage. But, as indicated in the argument, the only non-sceptics who are obviously concerned here are the compatibilists, those who believe that freedom or free will is compatible with determinism.[3] This, at least, is the opposition I want to focus on.

My claim is that the contextualist strategy is unhelpful in dealing with free will scepticism. In a nutshell, contextualism falls prey to the following dilemma. On the one hand, contextualists can argue that contextualism *per se* refutes scepticism, and that the claim that free will exists in normal contexts is merely a consequence of contextualism, not of a substantial compatibilist theory of free will. But this fails blatantly. Without a substantial compatibilist theory of freedom or free will, one cannot show that compatibilist standards are the relevant ones to courts of justice, to day-to-day interactions, and so on. Contextualism does not secure (2) on its own. On the other hand, contextualists can argue their view has a more limited purpose in dealing with free will scepticism. They can come to term with the fact that their view is merely an add-on to compatibilism—one which does not refute scepticism. But they can instead insist that contextualism comforts some sceptical intuitions that compatibilism neglected. Yet this strategy causes more troubles

[3] It seems in fact that free will sceptics and non-sceptical incompatibilists share similar standards. For both require that determinism be false for free will to exist.

than it solves. So, even if contextualists were to secure (2), they should stick to compatibilism.

I do not claim that contextualism is false, but that that it is unhelpful to meet the sceptical challenge mounted by free will sceptics. Yet there is something that contextualists get right regarding free will scepticism. *If* the dispute between compatibilists and sceptics were one of degrees or standards of application, as contextualists argue, then the advantage would go to the former. That is, the free will sceptic should also reject (1). If the only difference between her view of free will and the compatibilist's is a matter of degree or of stringency of standards, then there seems to be little point in scepticism, other than a penchant for excessive rigour.

Here is the plan. §2 explains with more precision what contextualism consists in and the different shapes it can take. §3 argues that contextualism *qua* contextualism is an inadequate response to scepticism, in accordance with the first branch of the dilemma. §4 discusses difficulties for the second branch of the dilemma. §5 extracts an important lesson for the sceptic.

2. Contextualism: A Compendium

Contextualism is a large family of theories. A contextualist theory of *X* claims that the propositions expressed by utterances containing '*X*' may vary according to the context of utterance. In this section, I sketch what contextualism about free will is by a comparison with contextualism about knowledge. This is important because the two contextualisms share most of their alleged upsides and downsides.

Epistemic contextualism claims that the propositions expressed by utterances of the form '*S* knows that *P*' may vary depending on the context of utterance. More precisely, they hold that the strength of epistemic position[4] that is required of *S* for the utterance of '*S* knows that *P*' to be true will be higher in some contexts of utterance than others. Hence, a knowledge claim which is true for a bystander in a day-to-day context might be false for another bystander in an epistemology classroom, because the standards which apply in day-to-day context are much lower than in an epistemology classroom (more on this soon).[5]

According to this view, the verb 'know' and its cognates, and perhaps terms that refer to requirements for knowledge, such as 'warrant' or 'justification', are comparable to adjectives like 'big', or 'important'. For instance, an elephant is certainly a big thing if we discuss biology, but small if we do astronomy; similarly, we know that we have hands when we explore gardening techniques, but we do not when we examine sceptical scenarios. What determines the context can be either the domain of inquiry—epistemology, astronomy—or elements raised in the conversations. Here is an example of a contextualist theory of knowledge:

---

[4] This can be interpreted in terms of the level of evidential state that *S* must be in, or the degree of justification that *S* must have, or the kind of reasons that *S* must possess.

[5] Another (close, though not equivalent) way to put things is that 'know' and its cognates are equivocal and that while under some sense of them we don't know anything, there is certainly a sense of 'know' such that we know that we have hands. The context of utterance is what determines the sense of 'know'. This is a handy way of explaining contextualism, but it raises questions about how to individuate different senses—suppose that you and I use a word slightly differently; do we use it in the same sense? I wish to leave this question aside.

> *S* knows that *P* iff *S*'s evidence eliminates every possibility in which not-*P*—Psst!—except for those possibilities that we are properly ignoring. (Lewis 1996, 554)

What possibilities may be properly ignored will vary according to the context of utterance.

The stated purpose of contextualism is to contribute to dealing with the sceptical challenge. Imagine a sceptic and a non-sceptic, in different contexts, looking in on Tanya eating her cereal. The former utters 'Tanya does not know that she has hands' while the latter utters 'Tanya knows that she has hands.' According to the contextualist, they might both speak truly. The sceptic and the non-sceptic do not really disagree with one another. Rather, due to their different contexts, they express different relations by the verb 'know' and its cognates—or at least they apply different standards when using the word.

There are two ways of understanding the contextualist strategy. A more ambitious (and traditional) contextualism maintains that contextualism *per se* dissolves the opposition between the anti-sceptic (e.g. Williamson 2000) and the sceptic (e.g. Unger 1984) (or anyone who believes that sceptical standards of knowledge or justification are acceptable, viz. cartesian internalists). It confines the sceptic to the theoretical realm. Contextualism implies that, in some contexts, sceptics are wrong and that, in some other contexts, anti-sceptics are wrong. As such, contextualism about knowledge is an explanatory hypothesis (Feldman 2004a, 261): it explains away the disagreement between the sceptic and her opponent. Call this *heroic epistemic contextualism.*

A humbler contextualism maintains that contextualism is a mere supplement to an anti-sceptical account of knowledge. The contextualist initially endorses a substantial anti-sceptical view of knowledge (say a form of externalism). Then she concedes that the sceptic is right to deny the existence of knowledge in some philosophical contexts. She then reconciles the sceptic and the anti-sceptic by giving a contextualist twist to her substantial theory. This twist is the recognition that, despite the fact that the compatibilist has refuted the sceptic, the sceptic is right in theoretical contexts when she claims she is ignorant. As such, contextualism features a conciliatory move, one which perhaps makes anti-scepticism more plausible. Call this *piggybacking epistemic contextualism*.

Let us now turn to free will or freedom. Contextualism about free will or freedom claims that the propositions expressed by utterances of the form '*S* freely φ-es' may vary depending on the context of utterance. More concretely, they hold that the level of freedom[6] that is required of *S* for the utterance of '*S* freely φ-es' to be correct will be higher in some contexts of utterance than in others. A freedom claim which is true for a bystander in a day-to-day context might be false for another bystander in a metaphysics classroom.[7]

---

[6] This can be interpreted in terms of the subject's having alternatives or having good alternatives or leeway, in terms of the degree to which she is the source of her action, etc.

[7] Another (close, though not equivalent) way to put it is that 'free' and its cognates (or some of the terms which refer to the requirements of freedom like 'sourcehood' and 'can' (see Lewis 1976, 150–52)) are equivocal and that while in one sense we are always unfree (or incapable of acting differently), in another sense we are free to do many things. The context of utterance is what determines the sense of 'free'. See note 5 for difficulties.

Either 'free' and its cognates or something that is constitutive of freedom or free will (e.g. sourcehood or leeway) would then share the implicit relativity to a standard which we attributed earlier to 'big', 'important', and 'knowledge'. Again, what determines the context could be the domain of inquiry—psychology, morals—or the elements raised in the conversations. Here is an example of a contextualist theory of free will or freedom:

> S does x freely only if S's action is free from causal explainers beyond S's control—Psst!—apart from those causal explainers that we are properly ignoring. (Hawthorne 2001, 68)

Hawthorne's pet theory (he does not endorse it) is a 'sourcehood' theory. It is a sourcehood theory because it suggests that free actions are those which are not adequately explained by certain sources. For instance, my action is free if it is not adequately explained by the coercion of another or by the laws of nature and the distant past (bear in mind that in many cases we might ignore the laws of nature and the distant past, but not the coercion of others).

I will use Hawthorne's theory as a paradigmatic example of contextualist for most of this paper, but it should be clear that other theories are equally affected. For instance, we could have a theory which says

> *S* does *x* freely only if refraining from *x* is compatible with the events over which *S* has no control—Psst!—apart from those events that we are properly ignoring.

This theory is a 'leeway' theory because it suggests that free actions are those that the agent had the ability to refrain from performing or for which the

agent had alternatives. For instance, my action is not free if refraining from performing them would require me to run faster than the speed of light.

The stated purpose of contextualism is to contribute to dealing with the sceptical challenge. Imagine a sceptic and a non-sceptic, in different contexts, looking in, yet again, on Tanya eating her cereal. The former utters, 'Tanya freely decided to eat her cereal' while the latter utters, 'Tanya did not decide to eat her cereal freely.' According to the contextualist, they might both speak truly. The sceptic and the non-sceptic, at least when not in the same context, do not really disagree with one another. Rather, due to their different contexts, they apply different standards when using the word 'free'.

Again, there are two ways of understanding the contextualist strategy. A more ambitious contextualism (see e.g. Hawthorne 2001, 73; Rieber 2006, 223) maintains that contextualism *per se* dissolves the opposition between the non-sceptic (e.g. Fischer and Ravizza 1998) and the free will sceptic (e.g. Pereboom 2014a) (or anyone who thinks that the sceptic's standards of freedom or free will are acceptable, viz. some incompatibilists). It confines the free will sceptic to the theoretical realm. Contextualism *implies* that, in some contexts, sceptics are wrong and that, in some other contexts, anti-sceptics are wrong. As such, contextualism about free will is an explanatory hypothesis: it explains (away) the disagreement between the sceptic and her opponent. Call this *heroic free will contextualism*.

A humbler contextualism (see e.g. Stace 1953; Willaschek 2010, 568) maintains that contextualism is a mere supplement to an anti-sceptical account

of free will or its components. The contextualist initially endorses a substantial anti-sceptical view (say classical compatibilism). Then she concedes that the sceptic is right to deny the existence of free will in some philosophical contexts. She then reconciles the sceptic and the anti-sceptic by giving a contextualist twist to her substantial compatibilist theory. This twist is the recognition that, despite the fact that the compatibilist has refuted the free will sceptic, the free will sceptic is right in theoretical contexts when she claims she is unfree. As such, contextualism is a conciliatory move, one which perhaps makes compatibilism more plausible. Call this *piggybacking free will contextualism*.

3. First Branch: Heroic Contextualism

I now turn to the dilemma which I summarised in the introduction. As I have just explained, there are two ways of championing a contextualist strategy. The first way, heroic free will contextualism, tries to refute scepticism independently of a comprehensive compatibilist theory of free will. The second way, piggybacking free will contextualism, accepts that it depends on a substantial compatibilist theory and therefore does not by itself refute scepticism. I deal with heroic contextualism in this section.

My claim is quite simple. Contextualism *per se* does not deal with scepticism, whether about free will or about knowledge. It depends on a substantial non-sceptical theory of freedom or free will and of knowledge. In fact, it typically piggybacks on compatibilism (for free will) or on externalism

(for knowledge).[8] Therefore, contextualism *per se* does not warrant the second premise of the anti-sceptical argument we considered in the introduction:

> (2) Political theorists, jurists, and people in their day-to-day interactions ought only to be concerned with the term 'free' in the sense which is used by compatibilists.

In fact, (2) can only be justified by an account of the superiority of compatibilism over scepticism. This argument applies to all forms of contextualism.

Here is why. Contextualism *per se* only specifies two things. First, it tells us that the senses of 'free' and 'know' change according to the context of utterance. Second, it tells us what a context of utterance is. These elements alone cannot tell us that sceptics are ever right or ever wrong.

First, that there is a change in the meaning of 'free' or 'know' does not tell us in what respect there is a change. The reason for this can be summarised by the slogan that *existing variation does not imply significant variation*. It is given by critics of contextualism about knowledge:

> That the propositions expressed by utterances of knowledge sentences in ordinary contexts are true, for example, is not secured by EC [epistemic contextualism] on its own, nor is the truth of skeptical propositions expressed in contexts where skeptics assert

[8] The reason why contextualism typically piggybacks on externalism rather than on internalism seems to be the following. Contextualism appears as a desirable solution to the problem of scepticism mainly for those who believe that internalism has failed to respond to scepticism. It would be better if we could prove that we are not brains in vats in virtue of the apparent reasons that we have access to, but that's not possible. If it were possible, the contextualist solution would be unnecessary; scepticism would be false in all contexts. Contextualism serves then to 'soften' externalism, which often seems dogmatic.

> them; only substantive theory can secure either result. (Rysiew 2016)

In fact, conversational contextualism in itself does not deliver substantive conclusions regarding the truth of actual knowledge attributions. This is because the thesis that the standards of knowledge vary does not entail anything about the *way* in which and the *extent* to which they vary. This is the same reason why the fact that not all fossil fuels are equally damaging is irrelevant to whether we should transition into renewable sources of energy.

The same applies to free will scepticism. It might be that sometimes more possibilities or explainers are relevant (that the word 'free' has higher standards). But it does not follow that scepticism is *ever true*. And vice versa: it might be that it is only in the philosophy class that determinism is relevant; but it does not follow that in ordinary contexts, causal explainers which relate to events at the moment of my birth are irrelevant when explaining my action. Therefore, scepticism might be true *at all times*.

Second, a given way of specifying contexts of utterance does not imply anything substantive about scepticism. There are two main theories of contexts of utterance for the purpose of contextualism. The first is *conversational*: a context of utterance is constituted by conversational moves (Hawthorne 2001; Rieber 2006):

> If one does not ignore a causal influence then, a fortiori, one does not properly ignore it. As a result, the truth value of 'S did x freely' in the mouth of an utterer can change without a change in the causal factors relevant to S's doing x but *due instead to a change in the causal factors attended to by the utterer*. […] Freedom, then, is

> dependent upon the *context of attention*, in addition to being dependent on the causal influences upon action. (Hawthorne 2001, 68–69, emphasis is mine)

For instance, if I discuss with a friend whether his house is large and someone mentions Versailles, then the context of utterance might change. The same applies if we discuss whether I know something or am responsible for an action and someone brings up scepticism. The second theory of contexts of utterance is *domain specific* (Willaschek 2010): what constitutes a context of utterance are conventions determining what is relevant and what is not. For instance, if lawyers are debating the culpability of a person in court, the context of utterance *does not* shift by the mere mention of scepticism. But the context of utterance shifts if the same two lawyers then move to the pub to discuss philosophy.

What is important to see is that, whether the contexts of utterance are conversational or domain-specific, contextualism does not by itself imply that scepticism is ever right or ever wrong. For, contexts of utterance are partly determined by non-contextual facts. Every contextualist accepts the need for rules of proper ignoring. It cannot be that *any* conversational move or absence thereof sets the context of utterance. For instance, if I have a conversation about whether my sausage dog is a large dog, I cannot say truly that 'it is a large' in virtue of the fact that no one in the conversation thought about the existence of other breeds of dogs. And, in the midst of a conversation about whether my house is large, if I bring up the size of Jupiter, this should not matter at all. Similarly, it cannot be that *any* convention or lack thereof

determines the context of utterance. If laws make no mention of a difference between adults and children and the accused is seven years old, then it does not follow that we can truly say that the child is responsible for his crime. And if laws specify that being a woman is an aggravating condition for a criminal offence, this has no consequence on the truth-value of assertions like 'she deserves to be severely punished.'

How should we determine rules of proper ignoring? This is where the debate between the compatibilist and the free will sceptic resurfaces. To see this, consider Hawthorne's view on the matter (2001, 70–72).[9] He suggests three ways of determining rules of proper ignoring without endorsing any of them. Let me mention them briefly. My purpose is not to assess any of them, but to show that they do not necessarily favour the compatibilist.

Firstly, Hawthorne suggests, we can use (rule) consequentialism to determine which type of causal explainers we have to attend to. Attending to more causal explainers (or considering more actions as alternate possibilities) might lead to more or less happiness and this could be what guides the determination of rules. Secondly, we can take a descriptive strategy and try to identify the rules which are followed in ordinary practice. If in ordinary practice people ignore such and such causal explainers in such and in such contexts, then this is what we should do. Thirdly, we can pursue a 'Strawsonian transcendental strategy' that maintains that ignoring certain causal explainers is a precondition for humane and respectful interactions.

[9] Others have suggested different strategies (see e.g. Rieber 2006, 238), but they suffer from other problems.

When respect requires that one attributes free will, the causal explainers which might suggest that the agent lacks free will should be left aside.

Hawthorne thinks that each option gives the upper hand to non-sceptical compatibilism:

> The analysis provides us with an attenuated compatibilism but a compatibilism nonetheless. The compatibilism is attenuated because it does not allow one who attends to the determining causes of human behavior to truly say 'That behavior is free'. But it is a compatibilism because it allows that ordinary folk in a deterministic world speak the truth when they say things of the form 'S acted freely'. (Hawthorne 2001, 73)

In other words, contextualism should lend support to:

> (2) Political theorists, jurists, and people in their day-to-day interactions ought only to be concerned with the term 'free' in the sense which is used by compatibilists.

Hawthorne's conclusion is hasty, and the sceptic should not accede to it. She can firstly reject each of Hawthorne's suggested strategies and claim that what determines rules of proper ignoring is partly a good metaphysical account of free will. Second, she can argue that Hawthorne is wrong to think that any of his three strategies favours compatibilism. For all we know, maybe a rule-consequentialist strategy commands that a vast number of causal explainers be taken into account such that free will is never attributed. In other words, maybe rule-consequentialist rules of proper ignoring lead to 'sceptical' rules of proper ignoring. It is also possible that Strawson is wrong about the pre-conditions of respect—maybe true respect demands that we understand more extensively the particular circumstances that led a person to act as she did.

And perhaps the descriptive strategy counts against compatibilism: perhaps compatibilism (but not free will scepticism) suggests revisionist rules of proper ignoring.[10]

Let me summarise. For epistemic contextualism to deliver substantive conclusions regarding the truth-values of actual knowledge attributions, rules of proper ignoring have to be determined. In fact, the standards of knowledge might vary according to the context of utterance but this variation in no way entails that people are generally or even *ever* right in their knowledge ascriptions. The same goes for what the question of what a context of utterance consists in—no good answer to this question (the conversational or the conventional answer) directly favours scepticism or compatibilism. So, epistemic contextualism must rely on rules of proper ignoring to deal with scepticism. But even if the sceptic and the non-sceptic were to accept contextualism, they would disagree about these rules. Differently put, they would disagree about the standards required for knowledge even if they were to agree that these standards vary and if they were to agree that they vary according to conversational moves or according to conventions. The same applies to free will.

Feldman (2004b, 32) made a similar point when he held that because the quality of our evidence is not agreed to in the debate between the sceptic and the non-sceptic, simply talking about standards of evaluation of evidence does

[10] It is often argued that compatibilists are revisionist about what 'free' means. The experimental philosophy literature on this question has shown that this is a complicated question.

not refute scepticism. The sceptic believes that our evidence is *worthless*, that it is not real evidence. So, the question of whether such and such situation requires such and such degree of evidence is entirely beside the point. The same can be said of free will scepticism. Because the quality of our alternatives or of our sourcehood is not agreed to in the debate between the sceptic and the non-sceptic, simply talking about standards of evaluation of sourcehood or leeway does not refute scepticism. The sceptic believes that our freedom or free will is but an *illusion*, that we have no such thing. So, the question of whether such and such situation requires such and such degree of leeway or sourcehood is entirely beside the point. (Trying to refute scepticism with the contextualist solution looks like trying to buy a carpet with what the salesman believes to be Monopoly money by claiming that his carpet is less valuable than he thinks.)

4. Second Branch: Piggybacking Contextualism

Perhaps because of these difficulties, some free will contextualists do not take themselves to occupy a real middle ground between compatibilists and sceptics. For instance, Willaschek (2010, 568) declares that the purpose of his contextualism is to make compatibilism more plausible, not to refute incompatibilist (or sceptical) concerns. And Stace (1953) argues that we can dispel sceptical concerns about free will once we see that in day-to-day contexts, 'free will' is an obviously compatibilist notion. Piggybacking contextualism is not a middle ground between sceptics and compatibilists; it is a kind of compatibilism.

Again, I am unsure whether contextualism is *true*. But my inquiry concerns whether contextualism is helpful in dealing with scepticism. We saw that contextualism does not refute scepticism. But it could be helpful in dealing with scepticism by either making compatibilism more plausible or by making a conciliatory move towards the sceptic. In a word, the contextualist is trying to sugar-coat compatibilism for the sceptic. But this comes at a cost.

One of these costs, well studied for the case of knowledge (see e.g. Williamson 2001, 26), is that contextualism makes it hard to deal with statements made in one context about the other. Suppose that half an hour ago, Hans had to fix his bicycle chain on his way to the Philosophy Faculty. He arrived in his epistemology class with his hands oil-slicked. By then he was very conscious of his hands; too much so in fact. For the contextualist, Hans knows he has hands. A few minutes into the lesson, he is presented with a sceptical argument about the external world. Now at the end of the class, the professor presents contextualism as a solution to scepticism and Hans is immediately convinced. He reflects on his past beliefs: 'I knew then I had hands then, but I don't know now that I had hands then.' This assertion is Moorean-paradoxical: 'I knew then I had hands then' implies that *I had hand then* and so Hans' thought implies 'I had hands then, but I don't know I had hands then'—which is an instantiation of '*p* but I don't know *p*'.

Well, perhaps that's too quick; in the classroom, Hans must not believe that he ever knows anything about his hands. In the classroom, Hans will have to believe that he never knew and outside of the classroom, he will have to believe that he always knew. But then, Hans must never believe that

contextualism is true. For 'contextualism is true, and I never know that I have hands' (believed in the classroom) is just as infelicitous as 'contextualism is true, and I always know that I have hands' (believed outside the classroom). And if Hans believes in epistemic contextualism, he should be able to infer the Moorean-paradoxical claim we just presented.

A similar kind of example shows that contextualism about free will has problems with cross-contextual claims. On her way to the Philosophy Faculty, Frida hits a pedestrian with her bike and feels badly about it—she did this freely. Once in class, Frida learns about free will scepticism and contextualism. She reflects on her past beliefs: 'I was free then, but I am not free now.' And she could even say, 'I was free then, I am free now, but nothing happened to my freedom.' This only seems felicitous if Frida has lost her freedom, e.g. if she has been submitted to the malevolent manipulation of her professor. But she has not lost her freedom.

Again, perhaps that's too quick; in the classroom, Frida should not believe that she is ever free. Perhaps she needs to think, in the classroom, that she is *never* free; and outside the classroom that she is always (or almost always) free. But, then, Frida must never believe that contextualism is true. For 'contextualism about freedom is true, and I am never free' is just as infelicitous as 'contextualism about freedom is true, and I am always free'. And if Frida believes in free will contextualism, she should be able to infer the infelicitous claim we just presented. It is a high price to pay for sugar coating

The response to this problem might be to shift from conversational contextualism to domain-specific contextualism (see §3). After all, the problem that we just noticed results from conversational shifts. If 'free' has different standards of application according to the domain, this problem does not appear. At any time, we can see that we are free according to low standards, though not free according to high standards. The problematic assertion, 'contextualism is true, but I am never free' becomes 'contextualism is true and I am never high-standard free', which is acceptable.

But domain-specific contextualism faces another problem. For the contextualist must either mischaracterise the nature of theorising or abandon contextualism. When we theorise, we wish to do this *about concepts which are relevant to other contexts*. Philosophers of physics wish to discuss concepts that physicists use or should use, and ethicists wish to discuss concepts that people use or should use in their moral lives. Similarly, epistemologists wish to discuss concepts that people use or should use in contexts of discovery or belief formation, and free will theorists wish to discuss concepts that people who deliberate or hold others responsible use or should use. These four philosophers might each think that ordinary concepts should be *revised*, but they are not trying to create a separate conceptual world for their own intellectual pleasure. To avoid mischaracterising the activity of theorising, domain-specific contextualism must not concede anything at all to the free will sceptic and simply tell her that she uses 'free' differently from all other practice

albeit wrongly.[11] But then this is not a form of contextualism any longer: our domain-specific contextualists believe, like any other invariantist compatibilist, that free will sceptics misunderstand and misuse the terms 'freedom' and 'free will' and that compatibilists do not. Not only is contextualism not a refutation of free will scepticism, but it does not feature a significant conciliatory move.

5. A Lesson about Degrees and Kinds

I have argued that contextualism did not provide by itself a refutation of scepticism, at least *qua* contextualism. Rather, what does the heavy lifting in the alleged contextualist refutation of scepticism is an underlying compatibilist theory in the case of free will and an externalist theory in the case of knowledge. Contextualism does not help us to avoid substantial debates. Thus, heroic contextualism fails. And piggybacking contextualism brings more troubles than it solves. So, contextualism is unhelpful in dealing with the sceptical challenge.

Yet, this discussion should be instructive to the sceptic. Contextualists rightly point out that the standards of application of certain concepts vary according either to conventions or to conversational moves. Take the concept of 'flat' and of 'friend'. What counts as flat is not the same thing when we buy a table and when we do geometry; who counts as a friend is not the same when we write our will and when we decide who to invite to a housewarming party. Once we notice this, we should also notice that the way in which the

[11] Stace (1953) is explicit: sceptics have a *wrong definition* of free will as opposed to determinism.

contexts change is usually such that the concept remains broadly applicable. In our day-to-day life, we use the low-standard concept of 'flat' because the high-standard concept is useless; and similarly for 'friend'. It would be silly to insist that 'I won't buy the table because it is not (geometrically) flat'; similarly, it would be silly (or antisocial) to use the standards of will-writing when we are deciding whom to invite to a housewarming party. (And vice versa: the normal concept of 'flat' is useless in geometry; and the broad concept of 'friend' is irrelevant to will-writing.)

Now, *if* the contrast between sceptics and non-sceptics about free will and knowledge is like the contrast between the silly (or pseudo-rigorous) and the non-silly person in the examples above, then this gives the advantage to the non-sceptic in both debates. There seems to be little theoretical motivation to have concepts which are never applicable simply because of their high standards. If scepticism is simply a theory that tells us that the standards of freedom or of knowledge are higher than we think, then scepticism at best serves to prevent us from thinking that we are *absolutely free*; but it is unclear that this is much more useful than for me to have someone reminding me that *'these people are not your will-level friends'* while I am writing down the guest list, although it might be a good thing when I consider what to spend on the evening.

In a word, the contextualist lesson is that the sceptic should not frame her challenge as one about the *degree* or *relevant standards* of leeway or sourcehood necessary for freedom. If she does, contextualism will point to a good compromise. You say that we do not have enough eggs for dinner and I say

that we do. Our disagreement vanishes if you concede that we have enough for a *small* dinner and that I agree that we do not have enough for a *large* one. Similarly, you say that we don't have enough evidence for knowledge and I say that we do. We resolve our disagreement if you concede that we are in a good enough epistemic situation to have 'lower' knowledge and I agree that we are not in a good enough epistemic to secure 'higher' knowledge. And again: you say that we do not have enough control to be free and I say that we do. We resolve our disagreement if you concede that we have enough control to be 'free enough' and if I agree that we are not 'perfectly free'.

This means that the truly interesting form of scepticism is *not* the high-standards one. In epistemology, the interesting question of scepticism is not the question of whether we should have high or low standards. And knowledge scepticism is independent from the claim that we never know anything *for certain* or at 100%—contrary to what undergraduate essays on scepticism often claim. The sceptic argues that we lack knowledge, full stop; she does not merely think that we are never *certain* regarding matters of the external world. She believes that we are no more justified in claiming that we have hands than in claiming that we have wings. For, in each case, we can be (or are) wrong even with the same evidence. Differently put, our evidence is worthless. We are not in the right *type* of situation to know. The sceptic who argues that you don't know whether you will die in the next month *is* relevantly different from the life-insurance company which applies self-servingly high standards. Whereas the life-insurance company might overestimate the impact on your life-expectancy of your diabetes, it does not

think that insuring the diabetic is as risky as insuring the professional rock-climber. So, the view that we are not 100% certain that we have hands is not a good response to scepticism—it neither implies that scepticism is right, nor that it is wrong.

The same goes for scepticism about free will. The interesting form of scepticism about free will is neither refuted nor vindicated by the claim that people are not 100% free. Rather, it claims that we are no more the source of our action than the drug addict, or no more able to do otherwise than the paralytic. That very strong claim is what the interesting form of free will scepticism maintains. (The sceptic can still recognise an important difference between the drug addict, the paralytic and the others; but it will be a different kind of difference.)

The contextualist lesson can be expressed in a different way, using the main argument we considered in the introduction. Remember its first premise:

> (1) Free will sceptics use the term 'free' or its cognates with higher standards than compatibilists do.

Now, my point in this section is that, if we grant this premise, even though it does not follow that free will scepticism is false or innocuous, it is harder to resist the second premise:

> (2) Political theorists, jurists, and people in their day-to-day interactions ought only to be concerned with the compatibilist standards of the term 'free' or its cognates.

In sum, if we accept that the difference between the sceptic's idea of freedom and the compatibilist's is a matter of degree or of the stringency of standards,

then we lack a theoretical motivation for taking the sceptical option seriously. This simply comes from a comparison with other concepts: big, warm, friend, and so many more. To be taken seriously, the sceptic should not let the contextualist interpret her as holding the superficial view that freedom requires high standards of control or, say, leeway or sourcehood. Premise (1) is an anti-sceptical Trojan Horse and premise (2) is a thoroughly compatibilist premise which has little to do with contextualism. The sceptic should reject both.

The point is not that the sceptic has some difficult concessions to make or even that she needs to accept a surprising claim. Rejecting (1) and (2) is not particularly costly for the sceptic. The point is rather that we can better understand the sceptical challenge by way of examining the inadequacy of the contextualist response.

6. **Conclusion**

I have argued that contextualism does not refute scepticism, either about knowledge or about free will. This is true of any form of contextualism, whether about freedom or knowledge. For, while the non-contextualist elements of these theories might refute free will scepticism, it is not the proposition that standards vary that provides reason to abandon scepticism. It is the proposition that these standards can be met (and are typically met) regardless of whether determinism is true. And this is a compatibilist claim which can only be argued for with the help of a substantive compatibilist arguments. Contextualism is not disproved by these remarks, but the truth of contextualism is a different matter.

However, contextualism is not irrelevant to sceptics in that it teaches them a lesson. For *if* the sceptical challenge were about degrees or standards, then the contextualist theory would perhaps be a good response. Contextualism can solve superficial disagreements. People who agree on what free will is and requires qualitatively and on what knowledge is and requires qualitatively can nevertheless disagree about matters of degree. It is there, and only there, that the contextualist theory can deal with the sceptic. This warns us against conceiving the debate between sceptics and non-sceptics as a superficial one.

III. SELF-DEFEATING BELIEFS AND MISLEADING REASONS

Published in the International Journal of Philosophical Studies (Chevarie-Cossette 2019a)

Abstract: We have no reason to believe that reasons do not exist. Contra Bart Streumer (2017), this has nothing to do with our incapacity to believe this error theory. Rather, it is because if we know that if a proposition is true, we have no reason to believe it, then we have no reason to believe this proposition. From a different angle: if we know that we have at best misleading reasons to believe a proposition, then we have no reason to believe it. This has two consequences. Firstly, coming close to believing the error theory is idle or pointless. Secondly, philosophers who argue that believing sweeping theories like determinism or physicalism is self-defeating because they are either false or believed for no reason pursue a worthwhile argumentative strategy.

1. Introduction

Do we have reasons to believe propositions and to perform actions? Some error theories, of the *sweeping* kind, say 'no' (see e.g. Streumer 2017). More *modest* error theories accept that there are such reasons but deny their 'objectivity' or 'authority'. These modest error theories usually claim that reasons depend, for instance, on our desires (see e.g. Mackie 1977, 27–28), or on what we take to be important practical problems (see e.g. Husi 2013, 432).

Whatever the merit of modest error theories, it is tempting to conclude that we have no reason to believe the error theory of the sweeping kind (henceforth 'the error theory'). For if the error theory is true, there are no

normative properties, including reasons to believe the error theory—call this 'The Basic Argument'. Not so quick, says Bart Streumer:

> this argument does not show that there is no reason to believe the error theory. All it shows is that
>
> (1) *If the error theory is true*, there is no reason to believe this theory.
>
> And the belief that (1) is true can only make us believe there is no reason to believe the error theory if we already believe the error theory […]. (Streumer 2017, 155)

This response is partly right. The mere fact that <if the error theory is true, we have no reason to believe it> does not imply that we have no reason to believe the error theory. Nor does it explain why we have no reason to believe it. In this sense, *The Basic Argument* is lacking.

But this response is partly wrong. For not only is it true that <if the error theory is true, we have no reason to believe it>, but we *know* that this is the case. Yet, this epistemic fact implies that we have no reason to believe the error theory. It also explains why we have no reason to believe the error theory. So, the spirit of *The Basic Argument* is correct.

My central claim is thus that we have no reason to believe the error theory because we know that we have no reason to believe the error theory if it is true. This explanatory claim has two consequences. The first concerns the error theory. Streumer has argued in favour of *coming close to believing the error theory*. Like me, he thinks that we have no reason to believe the error theory, but unlike me, he thinks that this is because of a *psychological fact*, namely that

we cannot believe it (Streumer 2017, 171)[1]. But, he also argues, if we have no reason to believe the error theory because we cannot, we should still try to come close to believing it. Several have already criticised Streumer's approach by arguing that we *can* believe the error theory.[2] Yet this sort of response is unnecessary. For, if I am right, we have no reason to believe the error theory *whether this is psychologically possible or not*. Thus, it seems idle or pointless to come close to believing the error theory.

There is a second and philosophically more important consequence of my central claim. My central claim is correct because *in general* we have no (epistemic[3]) reason to believe propositions that we know we have no reason to believe if they are true. For instance, since I know that <if I am a plant, then I have no reason to believe that I am a plant>, I have no reason to believe that I am a plant. Now, this general fact is crucial for many philosophical debates. Several philosophers since Epicurus (*On Nature*, XXV, 34)[4] have criticised sweeping doctrines like physicalism, scepticism, and determinism on the basis that we can only have reasons to believe them if they are false. My paper vindicates this argumentative strategy. To believe propositions which we know to be either false or unsupported by reasons is *self-defeating*.

---

[1] Streumer argues forcefully that we have no reason to believe what we cannot believe just like we have no reason to do what we cannot do.

[2] See, for instance, Bergamaschi Ganapini (2016), Lillehammer and Möller (2015), and Olson (2016). For responses, see Forcehimes and Talisse (2016) and Streumer (2016).

[3] There might be practical or all-things-considered reasons to maintain these beliefs (see e.g. James 1912; Rinard 2017).

[4] See, for instance, Bailey (1926); Popper (1972, 223–24); Long (1992); Dworkin (2011, 225); Slagle (2016); Lockie (2018).

Here is the game plan. Section 2 defends my central claim by salvaging *The Basic Argument*. Section 3 defends the most contentious premise of my reconstruction. Section 4 responds to three objections and makes a modest concession. Section 5 discusses the two consequences I have briefly outlined. §6 shows that these results sit well with a plausible account of self-defeat.

2. Salvaging the Basic Argument

Consider again *The Basic Argument*:

> If the error theory is true, there are no normative properties.
>
> The property of being a reason for belief is a normative property.
>
> *Therefore,* there is no reason to believe the error theory.

Streumer (2017, 155) is right to protest against this argument, for it is invalid. What deductively follows from the argument is:

> (1) If the error theory is true, there is no reason to believe this theory.

Streumer accepts the two premises of *The Basic Argument* and so he also accepts (1).[5] I will follow him in accepting them.

But might *The Basic Argument* just be *enthymematic*? Streumer asks whether the missing premise could be:

> (2) If a theory is false, there is no reason to believe this theory.

[5] See Streumer (2017, ch. 8), where he explicitly endorses both premises.

Streumer rightly doubts (2). Aristotelian physics is false, but until Galileo, scholars had good reasons to accept it. This seems true of many claims of various kinds—not just theories. For instance, during the Olympics of Lake Placid in 1980, viewers had good reasons to believe that the Olympic Russian hockey team would wipe the floor with the American team. But the Russian team lost. Therefore, we might sometimes have good reasons to believe a falsehood.[6]

I think that *The Basic Argument* is enthymematic, though (2) is *not* the missing premise. When we come to know that (1) is true, we realise that the error theory is either false or unsupported by reasons. And it follows from this that we have no reason to believe the error theory. For this argument to work, though, we need to slightly weaken the conclusion of our argument by limiting the scope of the conclusion to those who have followed the arguments of this paper ('we'). This is mainly because not everyone knows that <either the error theory is false or there is no reason to believe it>. This amendment is unproblematic for our purpose.

The revised argument has two steps, the first of which Streumer accepts:

First step

If the error theory is true, there are no normative properties.

The property of being a reason for belief is a normative property.

---

[6] If (2) is true, *The Basic Argument* is salvaged. This might tempt those who believe that *being justified* is factive (e.g. Williamson 2000).

(1) If the error theory is true, there is no reason to believe this theory.

Now we come to know (1).[7] This takes us to the second step:

Second step

We know that <If the error theory is true, there is no reason to believe this theory.>

(3) If we know that < If the error theory is true, there is no reason to believe this theory.>, then we have no reason to believe the error theory.

(C) We have no reason to believe the error theory.

We will refer to this argument as *The Two-Step Argument*. Once we accept this argument, we can explain why we have no reason to believe the error theory. We have no reason to believe the error theory because we know that it is either false or unsupported by reasons.

3. Reasons and Misleading Reasons

(3), the most crucial premise of *The Two-Step Argument*, is true because it instantiates the following:

No Reason: If we know that <If there are reasons to believe *p*, *p* is false.>, then we have no reason to believe *p*.

No Reason is the cornerstone of this paper. (You might be worried that we virtually never have *no* reason to believe a proposition. This could be an

[7] Streumer (2017, 105) can admit this because he accepts the existence of knowledge. All we need here is something like **Closure:** If *S* knows that *p*, *S* knows that <*p* entails *q*> and *S* competently deduces *q* from *p*; *then, S* knows that *q*.

acceptable hyperbole. After all, we do say things like 'you have no reason to worry' all the time, meaning perhaps 'your reasons to worry are negligible'. Anyhow, I will show in Section 4 that we can qualify the principle at no cost.)

No Reason already looks good as it stands, and you might find it more appealing than any argument that could be given in its support. You know that the proposition *p* is either false or unsupported by reasons. If *p* is false and you believe it, you will believe falsely; if *p* is unsupported by reasons and you believe it, you will believe it for no reason. Why bother believing?[8] Here is an analogy. On a visit to an English pub, you know the beer you are considering is either tepid or weak. If the beer is tepid, it will be unrefreshing; if the beer is weak, it will be tasteless. Why bother drinking?

Responding to this is not easy. At times, it looks like Streumer (see 2017, 171) has to free-ride on the possibility that the error theory be false for the belief in it to be reasonable; and to free-ride on the possibility that the belief in the error theory be unreasonable for the belief to be true. When forming her belief (or when coming close to do so), the error theorist hopes to be a reasonable false-believer; when she considers her belief content, she hopes to be an unreasonable true-believer. But this looks like a bad trick. If I know that *A* or *B* is true, that I know that *A* implies something unacceptable (e.g. falsehood, tepidness) and that *B* implies something unacceptable (e.g. unreasonableness, weakness), then I know that *A* or *B* implies something unacceptable (e.g. falsehood or unreasonableness, tepidness or weakness).

[8] You *might* form a true belief for no reason but you have no reason to believe this because you have no reason to believe that *p* is true.

Well, beliefs are not beers: while 'weak or tepid' is an undesirable quality for a beer, 'false or unsupported by reason' might not be an undesirable quality for a belief. In fact, it is an undesirable quality, and I can explain why by showing how *having no reason to believe* and *believing falsely* can be captured by a common property. This will explain why *No Reason* is true.

The train of thought starts from the idea that reasons to believe false propositions are all *misleading*. More precisely:

> **Misleadingness:** If *p* is false, and *S* has a reason *r* to believe *p*, *r* is a misleading reason for *S* to believe *p*.

Suppose that Alexis is nauseous because he is ill. Given his symptoms, his being nauseous is also a reason for thinking that he had food poisoning, albeit a misleading one. But Natalia, who has the same symptoms but whose doctor told her she has a stomach flu, has no reason to believe that she had food poisoning, and thus no misleading reason.

The train of thought continues as follows: if a subject knows that any reason to believe a proposition is at best misleading, she has no reason to believe that proposition. More precisely:

> **Reasonableness:** If we know that <for all *x*, if *x* is a reason to believe *p*, *x* is misleading>, then we have no reason to believe *p*.[9]

[9] *Reasonableness* does not mention the possibility of the inexistence of reasons, because doing so would be useless. <For all *x*, if *x* is a reason to believe, *x* is misleading> is logically equivalent to <For all *x*, if *x* exists and is a reason to believe, *x* is misleading>. One way of knowing this proposition is to know that there are no reasons. Another is to know that either there are no reasons, or they are misleading.

Why should we accept *Reasonableness*? Consider the following analogy: I tell you that you have to perform a certain manual labour; say, unscrewing a slotted screw. We both know that this labour requires a certain tool, i.e. a slotted screwdriver. But I tell you that you will either have access to an inadequate tool for the purpose, e.g. a crosshead screwdriver, or you will have no tool at all. Then, it would be unreasonable for us to expect you to succeed. It is not likely that the screw will budge. Similarly, suppose that I tell you that you have to perform a cognitive labour; say, forming a true belief about *p*. We both know that this cognitive labour requires a certain tool, e.g. a reason. But now, I tell you that you will either have access to a malfunctioning tool, e.g. a misleading reason, or no tool at all, i.e. no reason. Then it would be unreasonable for us to expect you to succeed. You are not likely to believe truly. If you form a true belief about *p*, it would merely be by lucky guess. It would in fact be just as lucky as if, dismissing your useless crosshead screwdriver, you unscrewed the *luckily loose* slotted screw with your fingernail.[10]

Misleading reasons are malfunctioning tools in the sense that they do not fulfil their function, namely guiding us to truths.[11] In fact, reasons to believe a proposition are reasons to believe *that this proposition is true*. But if I know that a reason is misleading, I know that it does not guide me to a truth. So, if I know that my reasons for *p* are at best misleading and yet believe *p*, I do not do any

[10] The error theorist might want to reply that her belief in the error theory is not like a lucky guess. The difficulty is to avoid reintroducing normativity in her explanation of this.

[11] This is not to oppose the view that knowledge is the aim of belief, as Hyman (2015, chs. 7, 8) demonstrates.

better than the person who believes a proposition by flipping a coin. In neither case am I *guided* to a truth. Identifying that a reason is at best misleading is a bit like realising that an informant is either working for the enemy (the informant is misleading) or uninformative (the informant is not really an informant). Listening to him is to accept being misguided, which is of course unreasonable.

In fact, *Reasonableness* captures what reasonable people do when they realise that their reasons are misleading. Take the following example. Charlotte's father asks her why she looks stressed. She responds that she misunderstands the basic rules of French grammar and that she will therefore fail her exam. Her father then shows Charlotte that in fact she understands the *basic* rules; what Charlotte misunderstood were *advanced rules* of French grammar, which she could ignore for the exam (she was unaware of this fact). Now Charlotte's father says: 'see, Charlotte, you had no reason to worry'. Charlotte is subtle: 'well, I don't have any reason to worry *any more*.' This seems right.[12] And Charlotte is right to relax now. *Reasonableness* also captures what would be wrong if Charlotte said: 'Alright dad, my reason to believe I would fail was misleading (if it exists). But I still believe I will fail the exam.' This would be a case study in *unreasonableness*: Charlotte would knowingly believe something for no reason.

[12] On the background of what Charlotte knew (or rationally believed), the fact that she misunderstands the advanced rules made it more probable that she would fail her exam.

To better see the interdependence between *No Reason* and *Reasonableness*, consider the following argument:

> 1. We know that <If there are reasons to believe *p*, *p* is false.> (By hypothesis.)
>
> 2. If there are reasons to believe *p*, *p* is false. (By factivity of knowledge & 1.)
>
> 3. If there are reasons to believe *p*, they are misleading. (By *Misleadingness* & 2.)

Since we competently made this inference (still assuming our hypothesis), we get:

> 4. We know that <If there are reasons to believe *p*, they are misleading.>
>
> 5. We have no reason to believe *p*. (By *Reasonableness* & 4.)
>
> **No Reason**: If we know that <If there are reasons to believe *p*, *p* is false.>, then we have no reason to believe *p*. (By discharging 1.)

You might think that *Reasonableness* and *No Reason* are equally controversial and thus that introducing *Reasonableness* fails to provide support for *No Reason*. But this would be to misunderstand its role: to *explain* why *No Reason* is true. It 'cooks' our raw intuition by explaining why we lack reasons to believe certain propositions in terms of misleading reasons.

Let us take stock. My central claim is that we have no reason to believe the error theory because we know that we have no reason to believe the error theory if it is true. To support this claim, I vindicated the spirit of *The Basic*

Argument in the form of *The Two-Step Argument*. To do so, I used a principle I called *Reasonableness*. I now turn to three objections to *Reasonableness* (Section 4) and to two consequences of my central claim (Section 5).

4. Three Objections

I want to consider three objections to *Reasonableness* (which could also be cast as objections to *No Reason*). First, some might be tempted to identify a counterexample to *Reasonableness* which exploits the fact that sometimes we know without knowing that we know.[13] Thus, Sarah might know deep down that her reasons to believe she has been unfairly treated are misleading. But, until she becomes aware of this, she might still have reasons to believe that she has been unfairly treated.

I believe this first objection is a diversion: so far, our discussion has only concerned *transparent* knowledge, i.e. knowledge that we know we have. The proposition that we focus on is the sweeping error theory, and we not only know that it is at best supported by misleading reasons: *we know that we know* that it is at best supported by misleading reasons. Restricting *Reasonableness* to cases of transparent knowledge would therefore have no impact on the principle's usefulness in reconstructing *The Two-Step Argument*.

Second, some might worry that *Reasonableness* warrants epistemic dogmatism. *Epistemic dogmatism*[14] claims that if a subject knows a proposition, then she is justified in disregarding any future evidence against it, for she

[13] On cases of non-transparent knowledge, see Williamson (2000, ch. 4).

[14] Kripke presented this thesis in 1972 as a paradox in a lecturer at Cambridge University to the Moral Sciences Club.

knows that such evidence is misleading.[15] This looks like *Reasonableness* put in terms of evidence. Since epistemic dogmatism is highly controversial and raises worries that I clearly have not tackled[16], this could be a problem. Yet *Reasonableness* is much weaker than epistemic dogmatism, for it does not warrant epistemic subjects *disregarding* misleading reasons. It only indicates that so long as the subject knows that any reason to believe *p* is misleading, this subject has no reason to believe *p*. This leaves open the possibility that the initial piece of knowledge (the knowledge that any reason to believe *p* is misleading) be defeated.

The third, and more serious, objection is as follows: *Reasonableness* overstates the consequence of coming to know that our reasons are at best misleading. Consider the following example to see the motivation for this worry. It is not raining (in Hackney), but Mike sees in the Daily Mail the headline 'HACKNEY FLOODED AGAIN'. Mike acquires a misleading reason to believe that it is raining. He then looks up the BBC website, which reports sunny weather. Mike comes to know that it is not raining. The BBC is after all very reliable and website predictions are normally better since they are updated constantly (much more than the paper version of the Daily Mail, which Mike affectionately calls the 'Daily Fail'). Mike thus comes to know that his initial reason was misleading. Now *Reasonableness* says that Mike has *no reason* to believe that it is raining. Yet some might be reluctant to go this far. They might prefer to say that Mike's reasons to believe it is raining have been

[15] This exact principle is defended in Lasonen-Aarnio (2014b).

[16] Opponents of epistemic dogmatism include Wright (2007) and Baumann (2012).

weakened or *outweighed*, but not *erased*. According to this line of thought, Mike has an *insufficient reason* to believe that it is raining rather than *no reason* to believe.

This qualification might be important, especially if we want to say that Mike's reason could become sufficient again. Suppose that Mike later learns that the BBC website team has just been caught taking hallucinogenic drugs. Then, Mike might once again have sufficient reasons to believe that it is raining (the Daily Mail headline). In technical terms, Mike's *defeater* (the BBC announcement) could be *defeated* and thus his initial reason would be *undefeated*. Or suppose that Mike's neighbour waters his plants so much that Mike acquires the impression that it is raining. Then, you might want to say that Mike's overall set of reasons will now be sufficient for believing that it is raining, thus outweighing the BBC's claim. (Whether you think that Mike's reason *is* sufficient depends on whether you think knowledge can ever be defeated. I remain neutral on this issue.[17])

In summary, the third objection goes as follows. There is a difference between having *no* reason and having *insufficient* reasons to believe; further, a reason which is known to be (at best) misleading is merely *insufficient,* but not *inexistent*.

If this third objection is correct, then we should qualify *Reasonableness*:

> **Reasonableness*:** If we know that <for all x, if x is a reason to believe p, x is misleading>, then we have *insufficient* reason to believe p.

[17] See for instance Lasonen-Aarnio (2014a).

The disadvantage of this amendment is that it forces us to adjust the conclusion of the *Two-Step Argument* to read as follows: we have *insufficient reason* to believe the error theory.

However, this disadvantage is inconsequential because Streumer and I are in the same boat. If I should talk in terms of sufficient reasons, so should he. If *Reasonableness* must be weakened, then so should his principle according to which we have no reason to believe what we know we psychologically cannot believe. This simply owes to the fact that both claims are subject to the 'defeat' objection. If my reason to believe *p* is merely defeated (made insufficient) by my knowledge that *p* is either false or unsupported by reasons, my reason to believe *p* should be merely defeated (made insufficient) by my knowledge that I cannot believe *p*. What matters is that we speak on the same level.

Still, I believe it is informative to press the objection. My original conclusion was supposed to be an indictment of the error theory. How much less of a bad thing is the new conclusion (namely that we have insufficient reasons to believe the error theory)? To answer this question, we need to see what insufficient reasons are *insufficient for*.

Believing for no reason or believing for insufficient reason can neither be *rational* nor *reasonable*, at least when we believe *p* despite knowing that our reasons for *p* are at best misleading. This is a point that we discussed in Section 3: the only reasonable or rational thing to do when we realise that our

reasons for believing are at best misleading is to refrain from believing.[18] Of course, believing for insufficient reasons in this sense might be *more rational* or *more reasonable* than believing for no reason. But to believe for insufficient reason in this sense is to remain under the threshold of reasonableness or rationality. And it is this threshold that matters. Note that there might be other epistemic statuses which are incompatible with believing for insufficient reasons. However, I will leave this stone unturned.[19]

The gap between believing for no reason and believing for insufficient reasons narrows further once we focus on reasons which, like the reasons in favour of the error theory, *could never be sufficient*. It is fair to be careful with the concept of 'sufficient reasons' because of the phenomenon of defeat. But could our reasons be defeated here? So long as we know that our reasons to believe the error theory are at best misleading, then we know they will necessarily remain insufficient. However, many powerful arguments the error theorist could muster, there is no reason why we should change our mind: we still *know* that we will at best be given more misleading reasons. As we have seen in Section 3, this follows almost directly from what a misleading reason is and from what the sweeping error theory is. The error theorist can put up a good intellectual fight, but it is doomed to failure.

In summary, the first two objections fail straightforwardly. The third objection, if it is correct, asks Streumer and me to revise our conclusion, but in

---

[18] This is true regardless of whether rationality is a formal requirement or a substantial requirement.

[19] It's not implausible to think that I cannot *know* what I believe for insufficient reasons, setting aside self-evident and perceptual beliefs.

a way that favours neither of us. For simplicity's sake, I will leave aside the suggested revision for the rest of the paper.

5. Two Consequences

By defending *The Two-Step Argument*, I have vindicated my *central claim*: we have no reason to believe the error theory because we know that if the error theory is true, reasons do not exist. Let me now turn to the two consequences of this claim.

The first is that it undermines Streumer's project to convince us to come close to believing the error theory. Consider the following:

> I [Streumer] argued in chapter X [i.e. in the passage quoted in Section 1] that there is no reason to believe the error theory, but not because the property of being a reason does not exist if the error theory is true. [*contra The Basic Argument*] As I have said, that only shows that
>
> (1) *If the error theory is true*, there is no reason to believe this theory,
>
> and the belief that (1) is true can only make us believe that there is no reason to believe the error theory if we already believe this theory, which I have argued we cannot do. Instead, I argued that there is no reason for us to believe the error theory *because we cannot believe this theory*. But that does not make the error theory polemically toothless. For I showed […] that we can come close to believing this theory […]. (Streumer 2017, 171, emphasis mine.)

Coming close to believing a theory could mean several things: having a low credence in it, a belief in different parts of it at different times, or a belief that there are sound arguments which together seemingly imply its truth (Streumer 2017, 152–154).

If *The Two-Step Argument* is sound, Streumer's strategy is threatened because it now seems idle or pointless to come close to believing the error theory. In fact, it is because *the property of being a reason does not exist if the error theory is true* that there is no reason to believe it—not because of our incapacity. Or, more precisely, it is because we know this. But if this is true, it is hard to see the point of coming close to believing the error theory.

Might Streumer show that it is *also* because we cannot believe the error theory that we have no such reason? I somehow doubt this. Take a very complex physical theory, so complex that you cannot believe it. Now, imagine that you also know that believing it would be irrational. You would need, for instance, to completely disregard the scientific consensus against it. Why do you lack reasons to believe the theory? To me the answer seems obvious: 'because believing it would be irrational'; but certainly *not*, 'because you cannot believe it'. Again, take an immoral or irrational action which you cannot perform. Why do you lack reasons to perform this action? Because it is immoral or irrational, not because you cannot perform it.

Suppose, for the sake of argument, that I am wrong: it is *also* because we cannot believe the error theory that we have no reason to believe it. Even then, Streumer's project is threatened. What he needs is for us to have no reason to believe the error theory *solely* or *mainly* because we cannot believe it. Otherwise, it seems idle or pointless to try to come close to believing the error theory. Consider a proposition which you know you would have no reason to believe; or an action which you know you would be morally wrong to perform. Since you lack reasons to believe the first or to perform the second (reasons

that have nothing to do with your abilities), you also lack reasons, at least *prima facie*, for *coming close to* believing the first or for *coming close to* performing the second. Whether you can or cannot believe the theory or perform the action is irrelevant.

Take a case where you have no reason to believe a proposition or perform an action *merely because* you are incapable of believing this proposition or performing this action. Say you are pathologically suspicious. You might then, because of this, have no reason to let your lovely neighbour in or to believe what she says—you simply cannot. But you could have reasons to come close to letting her in (talking to her through the window) or to come close to believing what she says (raising your credence in what she says as a result of hearing her testimony). By contrast, it seems idle or pointless to come close to believing the error theory. For we know that, contrary to the neighbour case, what explains why I have no reason to believe the error theory is not solely or mainly our lack of ability.

Now, to the second consequence of my central claim. To understand it, we need to see that my central claim is correct because *in general* we have no reason to believe propositions which we know we have no reason to believe if they are true. For instance, since I know that <if I am a plant, I have no reason to believe it>, I have no reason to believe that I am a plant. This general fact is philosophically important: it vindicates an argumentative strategy which is used in several domains of philosophy.

Consider Karl Popper's argument:

> For according to [physicalism][20], any theories […] are held because of a certain physical structure of the holder […]. Accordingly we are deceiving ourselves […] whenever we believe that there are such things as arguments or reasons which make us accept [physicalism]. Or in other words, physical[ism] […] is a theory which, if it is true, is not arguable, since it must explain all our reactions, including what appears to us as beliefs based on arguments, as due to purely physical conditions […]. (Popper 1972, 223–24)

Popper concludes that physicalism (or believing it) is 'self-defeating'. This strategy is, like my strategy against the error theory, legitimate. (I also believe Popper's strategy is successful, but this is a different story.)

The same applies to Ronald Dworkin's strategy to discredit belief in pessimistic determinism:

> If pessimistic determinism [(determinism is true and so we lack free will)] is true, no one could responsibly think that he had made a wise decision in believing it. He had no choice but to believe it. (Dworkin 2011, 225)

I disagree with Dworkin: I think that to believe responsibly does not require choosing. But if this *were* correct, then Dworkin's line of reasoning *would* be a challenge for the pessimistic determinist. Dworkin's premise might be false, but his strategy is legitimate.[21]

Streumer has explicitly considered the opposite (2017, 192–193): he suggests (without endorsing this strategy) that the pessimistic determinist

---

[20] Popper (e.g. 1983, 103) uses the term 'physical determinism' to refer to physicalism.

[21] A striking example of this strategy can also be found in Malcolm (1968, 70).

admit Dworkin's principle and use it *in favour of pessimistic determinism*. In fact, Streumer suggests that the pessimistic determinist take the following three steps: first, show that it is impossible to believe pessimistic determinism[22]; second, argue that it is because of this (and presumably because of this alone) that we have no reason to believe it; third, convince us to come close to believing pessimistic determinism. Again, if I am right, this is *not* a good strategy for the pessimistic determinist for the exact same reason that it is not a good strategy for the error theorist. The second step is unwarranted.

6. Self-Defeat

Before I conclude, I want to show that what we have said so far sits well with a simple account of self-defeat. I do not mean to say that this account gives us an additional reason for rejecting the error theory. But it gives us a more general tool to avoid similarly problematic theories.

First, 'self-defeat' is a property of propositional attitudes like beliefs and inferences. It is not a property of propositions and arguments. If you think otherwise, you probably have in mind propositions like <No universal proposition is true.> or arguments like <No interesting argument is sound. This argument is interesting. Therefore, this argument is unsound.>. I think that we do better to follow the philosophical tradition in calling these examples 'self-refuting'.[23] A proposition or argument is self-refuting just when

[22] Dworkin agrees (2011, 204).

[23] This is the account presented in Mackie (1964); White (1989, 75); Page (1992, 423). Churchland and Popper also made use of it (Churchland 1981, Popper 1983).

it is contradictory or unsound because it is self-referential. The error theory, physicalism, and pessimistic determinism are *not* self-refuting.[24]

There are two ways for a belief to be self-defeating.[25] First, imagine that Tommy believes that he has no beliefs. The fact that he believes this proposition implies (and even *ensures*) that it is false. Tommy's belief is 'self'-defeating in the sense that the *believing* (the attitude) defeats *what is believed* (the content).

Second, imagine that Tommy acquires the belief that he will have acquired no adequate (perhaps justified, reasonable, or rational) belief by the end of the day. This might in principle be true; still, his belief is self-defeating. This is because the presumed truth of Tommy's belief implies that he does not believe it adequately. Tommy's belief is 'self'-defeating in the sense that the truth of *what is believed* (the content) precludes the *believing* (the attitude) from being adequate.

Thus, self-defeat is a property which, when it applies to beliefs, plays on an act/object ambiguity. Nothing here really defeats *itself*—it is either the *attitude of believing* which implies that the *belief content* is false or, alternatively, it is the alleged truth of the *belief content* which implies that the *attitude of believing* is inadequate.

The following definition captures our remarks:

_S_'s believing that _p_ is self-defeating =$_{df}$

[24] Some authors including Dworkin (2011) have a different opinion. See Streumer (2017, 124–28) for an illuminating discussion of this issue.

[25] Authors normally only focus on the second way (see e.g. Slagle 2016, 41–43).

> either
>
> **(Act Defeats Content)** The fact that S believes that p ensures that p is false;
>
> or
>
> **(Content Defeats Act)** The presumed truth of p ensures that S believes that p inadequately.

In my definition, I use 'ensures that' instead of 'implies' to avoid the strange consequence that believing any necessary falsehood is self-defeating. There is much more to be said, but I leave it for another day.

This account allows us to maintain that it is self-defeating for us to believe the error theory, just like it is self-defeating for me to believe that I am a plant. *Content Defeats Act* is clearly instantiated in each case.[26] Similarly, if either Popper or Dworkin was right, it would follow that believing physicalism or pessimistic determinism is self-defeating respectively.

7. Conclusion

Bart Streumer accepted that:

> (1) *If the error theory is true*, there is no reason to believe this theory.

I have argued that once we come to know (1), we have no (sufficient) reason to believe the error theory and thus that it would be irrational or unreasonable to believe it. *Contra* Streumer, this has nothing to do with our incapacity to

[26] This claim is true given that believing a metaethical theory for no reason is inadequate.

believe the error theory. Similarly for 'coming close to' believing the error theory: we, at least *prima facie*, lack reasons to do this as well. This conclusion generalises to any proposition that *we know* is as follows: if it is true, we have no reason to believe it. Believing these propositions is self-defeating. This has significant consequences for debates over physicalism, determinism, scepticism, and other sweeping doctrines. Believing each of them could be self-defeating, just as believing the error theory is.

IV. IS FREE WILL SCEPTICISM SELF-DEFEATING?

Published, in a different version, in the European Journal of Analytic Philosophy. (Chevarie-Cossette 2019b)

Abstract: Free will sceptics argue that we lack free will, that is the command or control necessary for moral responsibility. The Epicureans have alleged that it is somehow self-defeating. To interpret this charge charitably, we must first realise that it is propositional attitudes like beliefs and not propositions themselves which are self-defeating. So, *believing* in free will scepticism might be self-defeating. The charge becomes more plausible because, as Epicurus insightfully recognised, there is a strong connection between conduct and belief—and so between the content of free will scepticism and the belief in it. Second, we must realise that an attitude can be self-defeating relative to certain grounds. This means that it might be self-defeating to be a free will sceptic on certain grounds, such as the putative fact that we lack leeway or sourcehood. This charge is much more interesting, and, although it ultimately fails as it stands, it delivers important lessons to the sceptic. The most important lesson of the Epicurean is that the free will sceptic must either accept the existence of leeway or reject the principle that "'ought" implies "can"'.

1. Introduction

Free will scepticism is the doctrine that we do not have free will, i.e. the kind of command or control of our own conduct that we would need in order to be morally responsible for it.[1] The concept of free will allows for different more precise *conceptions*. It is sometimes as the ability to

[1] See e.g. Strawson (1994), Waller (2011), and Pereboom (2001; 2014a).

choose amongst real alternatives or, as philosophers say, leeway;[2] and it is sometimes understood as being the source of one's conduct.[3] Thus, sceptical arguments typically conclude that free will does not exist on the grounds that we lack leeway or that we are not the source of our actions.

Denying the existence of free will broadly defined might seem unthinkable to us. Radical as it stands, free will scepticism *must be wrong*. But if it is so radical, could it be turned against itself like some other sweeping philosophical doctrines? Several philosophers in history took up this challenge.[4] Their arguments do not share a common philosophical lexicon. Nor do they always target free will scepticism explicitly—they often target determinism. But if they are sound, it is self-defeating to believe in free will scepticism, whether it is true or false. Since Epicurus first marshalled these arguments, let us call them 'Epicurean arguments.'

This paper is slightly unusual in that it does not belong to a contained contemporary debate where it could make a very specific contribution. Few philosophers after the 70s took Epicurean arguments seriously. Those who did proposed complex reconstructions which relied on a vast array of controversial views about epistemology and free will.[5] Going back to the simple argument of Epicurus, distilling its general idea, and using it to regiment refined but simple Epicurean arguments is

[2] See e.g. Vihvelin (2013) and van Inwagen (2017).

[3] See e.g. Frankfurt (1971) and Fischer (1994).

[4] See, for instance, Hinman (1979), Dworkin (2011), Slagle (2016), and Lockie (2018).

[5] See e.g. Lockie (2018) and Slagle (2016).

my first goal. My second goal is to assess how the sceptic might respond to each type of argument and to unearth what she might already learn. I ask the reader to judge this critical engagement as it is: a precocious step in a nascent literature.

Here is the game plan to meet my two goals. I start with Epicurus' own argument (§2) and discuss what it means to be self-defeating. I then move to reconstructing the argument in the most charitable way (§3). This gives rise to arguments related to sourcehood (§4) and leeway (§5), which I examine in turn.

2. Epicurus' Argument

About the view that all comes to pass by necessity, Epicurus argued:

> This sort of account (λόγος') is self-refuting (τρέπω περικάτω), and can never prove that everything is of the kind called 'necessitated'; but he [the sceptic or the determinist] debates this very question on the assumption that his opponent is himself responsible (δι ἑαυτοῦ) for talking nonsense. (Epicurus, *On Nature*. XXV, 34; translation Annas, Taylor, and Sedley 1983, 19–23)

To extract the best possible argument from this passage, we need to answer two questions. First, who does the Epicurean criticise: the determinist or the free will sceptic? Second, what does the Epicurean criticise this target for: making a discourse, maintaining a belief, or posing an action? For simplicity, call these two things respectively the *target* and the *object* of the Epicurean arguments.

2.1 The Target and Object of Epicurus' Argument

The official target of Epicurus is the Stoic who endorses the doctrine of necessity. This doctrine approximates *physical determinism*, the thesis that the conjunction of all the physical states at a given time and the laws of nature determines the states at another given time. However, Epicurus' argument is mostly worrying to determinists who also believe that determinism undermines free will, namely hard determinists—as opposed to determinists who do not, namely soft determinists.

Why? Epicurus' argument, as we will soon discover, turns on the claim that rationality or reason requires free will (whether sourcehood or leeway). The soft determinist may simply respond to Epicurus that the apparent, though illusory, difficulty of reconciling determinism with rationality is entirely inherited from the apparent, though illusory, opposition between determinism and free will. But this response is unavailable to the hard determinist. The hard determinist must be an *incompatibilist* between determinism and free will but a *compatibilist* between determinism and rationality. This is the real challenge. So, there is no doubt that, if Epicurean arguments work, they work against hard determinists. The most charitable reconstruction of Epicurean arguments thus reads them as primarily targeting hard determinists.[6] To simplify, I will consider that the target of Epicurus extends to all free will sceptics. For one thing, sceptics typically admit the possibility of determinism; for another, Epicurus' argument hinges on the alleged

[6] Cf. Slagle (2016, 17, 28, 201–2) and Lockie (2018).

consequences of determinism for free will (or responsibility), not on determinism itself.

Moving to the object of the argument. In the passage, Epicurus focusses on the sceptic's *logos* (λόγος). But *logos* is famously ambiguous and suggests at least two possible translations. It could mean 'account', i.e. the sceptic's doctrine; or it could mean 'reasoning', i.e. the sceptic's inference or belief from reasons. What is wrong with this account or with this reasoning? They are described as *trepetai egkalein* ('τρέπω περικάτω'), a technical term which literally translates as *defeating upside down*. Translators talk about 'self-refutation' or 'self-defeat', but what does that mean?

2.2 Self-Defeat and Self-Refutation

To understand Epicurus' argument, we must distinguish self-refutation from self-defeat and their respective objects.[7] A proposition (or an account) is self-refuting, in one common sense of the term, just when it is contradictory because it applies or refers to itself; for instance, 'No universal proposition is true'. Similarly, an argument is self-refuting just when it is unsound because it applies or refers to itself; for instance, 'single-premise arguments are invalid; therefore, single-premise arguments are invalid'.[8]

Free will scepticism is *not* self-refuting in this sense, since it is neither contradictory nor self-referential. The arguments supporting free will

[7] In this, I follow Slagle's distinction (2016, 41–43).

[8] This conforms to Mackie's account (1964); see also Page (1992, 423).

scepticism are not self-refuting either: while they might be unsound, this is not because they apply or refer to themselves. This suggests that when Epicurus attacks the sceptic's *logos*, he does not mean her 'account'.

However, there might be *epistemic* problems with *believing* or *reasoning* (in the sense of reasoning to a conclusion) that free will scepticism is true. After all, this is what Epicurus' argument suggests: that the content of a doctrine interacts with the epistemic attitudes of its proponent in an unfortunate way.

This is where the notion of self-defeat is relevant; for its object is precisely attitudes like beliefs or inferences rather than abstracta like propositions or accounts. There are two kinds of case of self-defeating propositional attitudes like belief. First, imagine Tommy, an undergraduate philosophy student who, fortunately for our purpose and unfortunately for his peers, likes to embrace the most radical theses he encounters. Today he believes that he has no beliefs. The fact that he believes this proposition ensures[9] that it is false. Tommy's belief is 'self'-defeating in the sense that the *believing* (the attitude) defeats *what is believed* (the content)—see **Act defeats content** below. Second, imagine that Tommy acquires the belief that he would not acquire a single justified belief that day. Even assuming that this is true, Tommy's belief is self-defeating. The presumed truth of Tommy's belief ensures that he believes it inadequately, e.g. without justification, irrationally,

[9] One might read 'ensures that' as 'is a sufficient reason for', 'explains why', 'grounds', or 'makes'.

unreasonably, etc. Tommy's belief is 'self'-defeating in the sense that truth of *what is believed* (the content) precludes the *believing* (the attitude) from being adequate—see **Content defeats act** below.

Thus, self-defeat is a property which, when it applies to beliefs, plays on an act/object ambiguity. It is either the *attitude of believing* which ensures that the *belief content* is false or, alternatively, it is the alleged truth of the *belief content* which ensures that the *attitude of believing* is inadequate. In light of this, it is no surprise that the object of Epicurus' argument was uneasy to identify.

A last remark on self-defeating beliefs: believing a proposition could be self-defeating because of its grounds. Last week, Tommy came to think that he was hopeless at remembering events in the distant past. That's fine. But Tommy formed his belief on the grounds that, when he was eight, he forgot his best friend's name four times. Now, *this* is problematic. Tommy's belief could well be true; and this time, it could be adequate if it was differently grounded. But the presumed truth of the belief content ensures that the believing is inadequate, *grounded as it is*. Thus, whether a belief is self-defeating is relative to the grounds on which it is believed.[10]

[10] As a reviewer of the *European Journal of Analytic Philosophy* pointed out, this remark might stretch the notion of self-defeat since the grounds of a belief is something external to it and so cannot contribute to *self*-defeat. Yet what a belief contains is contentious: a belief in the same proposition but which has been rebased on other grounds is not obviously the same belief. When I say that a belief is self-defeating, I use this broader view of what a belief contains.

The following definition captures our remarks. (The same applies for inferences.)

> **Self-defeating beliefs =$_{df}$** Suppose that a subject *S* believes that *p* on the grounds that *q*. *Then, S*'s believing that *p* is self-defeating *if and only* if *either*:
>
> **(Act defeats content)** The fact that *S* believes that *p* on the grounds that *q* ensures that *p* is false; or,
>
> **(Content defeats act)** The presumed truth of *p* ensures that *S* believes that *p* on the grounds that *q* inadequately.[11]

What is an adequate belief? For the purpose of this discussion, I will assume that it is a belief which is rational or reasonable. For our purpose, I find it useful to understand rationality in believing as believing according to some rules of rationality and reasonableness in believing as believing on good grounds or for good reasons.

I also think that self-defeating beliefs are beliefs that we should dispose of, at least upon recognising that they are such. This is because they are beliefs which generally fail to be reasonable. Typically, a self-defeating belief is unreasonable in that it implies that it is unsupported by reasons: so either it is false or unsupported by reasons. Upon discovering this fact, we lose any reason to maintain this belief (see Chevarie-Cossette 2019a). We should note however that some philosophers still insist that they can reasonably question the existence of reasons (Olson 2014, chap. 9; cf. Cuneo 2007, chap. 4). The question turns essentially on whether reasons are inherently *normative* and on

[11] I have presented this account elsewhere (see Chevarie-Cossette 2019a).

whether we can, in distinguishing reasonable from unreasonable beliefs, use a non-normative notion like evidence. We must leave this interesting question for another day.

3. Reconstructing the Epicurean Argument

We now have some tools to partly explain one[12] interpretation of Epicurus' claim. Free will scepticism is not self-refuting, for there is nothing contradictory or self-referential about the *thesis* that free will does not exist. However, free will scepticism is a thesis that concerns conduct, but which extends to attitudes. The suggestion is that there is an unfortunate interaction between the content of the thesis of free will scepticism and the attitude of believing or inferring. This is crucial: a key to all Epicurean arguments is a connection between conduct and beliefs.

So, free will scepticism might be self-defeating to *believe in*. But if this is true, this cannot be in virtue of the fact that to believe it makes it false (as in Act Defeats Content). What remains is the possibility that the presumed truth of free will scepticism (perhaps combined with something else) ensures that one does not rationally or reasonably believe it (Content Defeats Act). Why would this be true? The most straightforward explanation is that rationality or reasonableness requires free will.[13]

---

[12] There is also a plausible moral interpretation of Epicurus' argument, but I leave it aside for our purpose.

[13] By 'free will', it is important to insist that I mean nothing more than the command or control necessary for moral responsibility. Otherwise, this would

This explanation sits well with another remark of Epicurus on the topic:

> The man who says that all things come to pass by necessity cannot criticise (ἐγκαλέω) one who denies that all things come to pass by necessity: for he admits that this too happens of necessity. (Epicurus, *Extant Remains*, Frag. XL; translation Bailey 1926, 112)

According to this point, the sceptic cannot legitimately criticise her opponent. This is because in trying to persuade her opponent, the sceptic is appealing to her opponent's rationality; but, again, rationality implies free will.[14] So, in order to convince someone of a doctrine, the free wil sceptic must presume its falsehood.

We need not focus on this dialectical problem. *If* it is self-defeating to rationally persuade someone to be a sceptic, it has everything to do with the *rational* character of persuasion and nothing to do with its *interpersonal* character (cf. Castagnoli 2007, 16). So we should ask again: is it true that rationality or reason implies free will?

It seems, after all, that I can rationally believe that 2+2=4 without having free will, i.e. the command or control necessary for moral responsibility. Here are two *pro tanto* reasons to strengthen this impression. First, there seems to be no necessary opposition between rational belief and constrained belief. Quite the opposite, constraint

---

be an anachronistic interpretation of Epicurus. See Bobzien (2000) and Frede (2011).

[14] Again, there is also a moral interpretation of this argument according to which it is unfair to criticise someone for not being a free will sceptic.

seems like a part of rationality, as Nozick (1981, 4–8) and James noted (1912, 168–69): to give in to *forceful* or *knockdown* arguments, or to be *forced to* a conclusion, is not irrational at all. Second, there seems to be no necessary opposition between rational action and unfree action (in the sense of an *inner freedom*). A heroin addict in rehabilitation acts rationally in taking the safe dose given by a doctor, regardless of whether he has the command or control to refuse to take it. If there is no opposition between rationality on the one hand and coercion or absence of inner freedom on the other hand, it is unclear that there is an opposition between rationality and absence of free will.

We can however refine our explanation of why it is self-defeating to believe in free will scepticism. Epicurus targeted the free will sceptic's *logos* or *reasoning*. But we reason *from premises* and we believe *on certain grounds*. Now, the grounds or premises for which one is a free will sceptic includes or implies the proposition that humans lack leeway or sourcehood. The Epicurean can take advantage of these points. She does not need to argue that the existence of rationality or reason implies the existence of free will, understood generally. She can instead argue more narrowly that either leeway or sourcehood implies rationally. Thus, believing in free will scepticism on the grounds that we lack leeway or sourcehood is self-defeating.

4. Sourcehood

Epicurus himself was a sort of sourcehood theorist. In fact, he claimed that the sceptic debates the question of responsibility 'on the assumption

that his opponent is himself responsible (δι ἑαυτοῦ) for talking nonsense.' Yet the Greek for 'responsible' is δι ἑαυτοῦ (*di eautou*), which means 'because of himself'—as opposed to because of something external (see Bobzien 2000, 291–292). This is sourcehood. So, according to Epicurus the sceptic cannot argue that we are never the *source of our conduct* because this presumes that she is not the *source of her reasoning or belief*. This is supposed to be self-defeating.

What is it to be the source of one's conduct? The contemporary literature gives us two main answers. We are the source of our conduct when our actions and omissions stem from reason-responsive mechanisms (see e.g. Fischer 1994; Hurley 2003; Sartorio 2016). Alternatively, we are the source of our conduct when our actions and omissions are the product of some of *our* desires or *our* values—something to which we identify or which belongs to our 'real selves' (see e.g. Frankfurt 1971; Shoemaker 2015). So, the first view emphasises reason while the second emphasises identification or ownership.[15]

There is a sceptical concern corresponding roughly to each answer. The first is that reasons are irrelevant to the explanation of our conduct: our conduct, including our reasons for action, can be entirely accounted for in terms of physical events. Call this *the exclusion concern*. The second concern is that our conduct does not ultimately belong to our real selves—or if it does, this is partly insignificant—because our conduct is

[15] These answers are not exclusive: for instance, Fischer endorses an endorsement condition on responsibility.

ultimately determined by events outside of our control. The same goes for reasons for which we act. Call this the *ownership concern*. The two concerns are importantly linked: the exclusion concern makes the radical claim that we never act *for* reasons; the ownership concern leads to the less radical view that we never act for reasons that are ultimately *ours*. In a word, the first concern casts doubt on reason for conduct; the second on ownership of these reasons. Each sceptical concern gives rise an Epicurean argument.

4.1 Exclusion

The exclusion concern about reasons stems from so-called *exclusion arguments* about mental states (see e.g. Kim 2007). Roughly, mental states seem to be unidentical to physical states and yet realised in them. But then the cause of our conduct could be entirely accounted for in terms of physical states. This leaves no causal role to be played by our mental states. The conclusion is not that mental states do not exist, but rather that we have no reason to think that they have a causal effect.

The same argument applies more specifically to *our* reasons. For the purpose of this discussion, I understand reasons as facts which explain or favour other facts (or which explain both normative and non-normative facts). And I understand *our reasons* as pieces of knowledge or justified beliefs in these reasons. When *S* acts for reason *R*, *S* acts because *S* knows *R* or because *S* is justified in believing *R*. I make these plausible[16] assumptions for simplicity's sake. On this picture, if mental

[16] See Hyman (2015) and Alvarez (2017).

states have no causal role, nobody acts for a reason. It vindicates free will scepticism.

The worry is that if our reasons have no causal role on our conduct, the same applies for our beliefs, including our belief in free will scepticism. This Epicurean response was most famously made by Karl Popper (and has not been much discussed in the recent literature):

> [P]hysical determinism is a theory which, if it is true, is not arguable, since it must explain all our reactions, including what appears to us as beliefs based on arguments, as due to *purely* physical conditions […]. But this means that if we believe that we have accepted a theory like determinism because we were swayed by the logical force of certain arguments, then we are deceiving ourselves, according to physical determinism […]. (Popper 1972, 223–24, emphasis is mine; see also Lockie 2018, chap. 10)

Thus, it is self-defeating to believe in free will scepticism because if it were true, it would follow that we do not believe it *for* a reason, since we believe only because of natural events. (The use of 'purely' is supposed to mark an exclusion.)

Popper was right to maintain this connection between actions and beliefs. If we never act for reasons, we never believe for reasons. If my *giving* to charity being caused by a natural event is incompatible with (or excluded by) my *giving* to charity for a moral reason, then my *believing* that I should give to charity being caused by a natural event is incompatible with (or excluded by) my *believing* it for a moral reason. For, in general, what counts as a reason to act can also count as a reason to believe. The

fact that wealth is unequally distributed in our society is *both* a reason to *give* to charity and to *believe* that we should do so. And once we know that wealth is unequally distributed, then it becomes our reason to act and believe. What counts as a cause for a belief can also count as a cause for an action. The Wall Street Crash of 1929 caused many people to believe that they had lost their fortune; and it caused Roosevelt to propose his New Deal. And what counts as acting or believing *for* a reason must include some sort of causal connection between an agent's holding to that reason and her action or belief.

This clear but admittedly controversial picture leaves no room for an asymmetry that the free will sceptic could exploit in responding to the Epicurean. This suggests that the Epicurean has a sound argument:

The Reason Argument

> If everyone acts because of natural events rather than for reasons, everyone believes because of natural events rather than for reasons.
>
> If everyone believes because of natural events rather than for reasons, no one believes rationally.
>
> *Therefore*, if everyone acts because of natural events rather than for reasons, no one believes rationally that free will scepticism is true.

I have just discussed the first premise, which is the typically Epicurean premise that connects conduct and beliefs. And the second rests on a natural connection between rationality and reasons and on the fact that by 'rather', I signal an incompatibility. As I indicated, it follows that

believing in free will scepticism on a certain ground is self-defeating, the ground being that we act because of natural events rather than for reasons. This is all I need to show for my purpose:

> **First lesson:** The free will sceptic cannot and should not rely on the doubtful claim that we act because of natural events rather than for causes.

So free will sceptics should try, as we all should, to find a way out of the exclusion concern.

Since the problem is general and applies to all mental states, the free will sceptic will not be able to solve it by drawing distinctions between different kinds of explanations or reasons. True, she can insist—as did Ayer (1963, 266–67) and Wolf (1993, 72)—that something can be both caused and justified. If I ask you 'why do you believe that Italian is the most beautiful language?', you may *both* respond that it is because you spent some time in Italy (which merely explains your belief) or that it is more colourful than any other language (which justifies your belief). But this does not help to deal with the exclusion concern. This is because action or belief *for* a reason (*because* I am justified) must make room for a causal role to reasons; and this causal role is precisely what the exclusion concern targets.

4.2 Ownership

We now move to the second sceptical concern and its corresponding Epicurean argument. The concern is that even if we act for reasons (*contra* the exclusion concern), these reasons—and so our actions—are

not truly *ours* since they can all be ultimately explained by facts or events *foreign* to us. Alternatively, even if some reasons (or desires and values) were ours, this would not mean much since our having these reasons would not ultimately be up to us. (see Strawson 1994; Pereboom 2014a).

The Epicurean can respond to this challenge, again by connecting actions and beliefs. If our actions and our reasons for action are foreign to us, so are our beliefs. But then our beliefs do not stem from *our* reasons and threaten to be irrational.[17] The argument can be put in terms of sourcehood:

The Sourcehood Argument

> If no one is the adequate source of their conduct, no one is the adequate source of their beliefs.
>
> If no one is the adequate source of their beliefs, no one believes rationally.
>
> *Therefore*, if no one is the adequate source of their conduct, no one believes rationally that free will scepticism is true.

The first premise is the typically Epicurean connection between conduct and beliefs. I shall argue that although the Epicurean is right to connect beliefs and actions in this way, she is wrong to think that moral free will (or responsibility) and rationality each requires sourcehood in the same sense. In a word, the *Sourcehood Argument* equivocates.

[17] See Kant (Groundworks 4:448) for the claim that foreign influence is incompatible with the work of reason.

To be the adequate source of her conduct in the sense relevant to responsibility and free will, a subject must first be the *source* of her conduct. Her conduct must *stem from her*. Thus, it cannot be the direct result of an external factor such as someone pushing her. Nor can it be the direct result of an internal factor which is not the agent's, such as an uncontrollable impulse or a disease. For a similar reason, someone who acts under duress is not the adequate source of her action: this action does not really *stem from her* in the relevant sense. It somehow is the action *of the coercer* (Nozick 1997, 38).

But this is not enough. A child soldier who participates willingly in an act of war is still not the *adequate* source of his conduct. He did not *make himself*; and the commitments that he manifests in his violent behaviour are not fundamentally *his*. Some sceptics suggest that this is because the child is not *the source of the source* of his conduct (see Strawson 1994; Pereboom 2014a). To use Hurley's phrase (2003), the free will sceptic (and some incompatibilists) requires that to be responsible for Φ, someone be responsible for the causes of Φ. Now, to be responsible for Φ or to be the source of Φ, a subject does not need to be responsible for or be the source of everything that leads to Φ. The subject must be responsible for or the source of what *determines* that she will Φ (see Istvan 2011). This still leads to a regress, unless the causal chain can stop in the agent's free act of will—which the sceptic denies. The subject's conduct is ultimately determined by something that she is not the source

of—and so for which she is not responsible. The subject will never be the *ultimate source* of her conduct—she will never be truly responsible.[18]

Free will sceptics can say the very same thing about beliefs: to be the adequate source of them, we need to be the adequate source of their cause. But it leads again to a regression: for the causal chain continues up until a point where we no longer are the adequate source of the thing we are considering. Since this reasoning is just as plausible—no more, no less—for beliefs and for conduct, the first premise looks true: if no one is the adequate source of their conduct, no one is the adequate source of their beliefs (the first premise).

Now, the second premise also seems true. Believing for good reasons or on adequate evidence is generally insufficient to believe rationally. For this, a subject must usually believe for good reasons *that are her own*—on *her* evidence. If Detective Chief Inspector Japp does not believe that there is an earring in the room, then it would be irrational at best for him to believe that the killer is a woman on the grounds that there is an earring in the room. Japp clearly needs to *acquire* the evidence or the reason—that he makes it *his*—and this in turn implies that he forms the relevant belief or that he comes to know the relevant fact.

For this, one might think, Japp needs again to believe on *his* evidence or for *his reasons*. And this might look like it causes a regress which shows that rationality, just like responsibility, is impossible. Yet this is mistaken. The fact that, to believe rationally, Japp needs to believe

[18] For a similar reasoning, see Galen Strawson (1994) and Istvan (2011).

for his reasons does not imply that the reasons for which he believes must have been acquired because of a further reason that was his own. Why? There are two classic answers to this. The foundationalist (most notably Aristotle) claims that there might be self-evident reasons or special pieces of evidence that Japp acquires simply by paying attention to them. The existence of these things clearly has nothing to do with free will or determinism. The coherentist (such as Blanshard) will claim that a series of belief can justify each other in a circle. Again, whether someone holds coherent beliefs has nothing to do with free will and determinism. The foundationalist can claim that while a responsible action requires a further responsible action *ad infinitum*, it is clear that a justified belief does not require a further justified belief *ad infinitum*. And a coherentist can claim that while a series of responsible actions does not secure the existence of responsible, a series of justified beliefs can.

This means that, whether coherentist or foundationalist in nature, the structure of justification differs from the structure of responsibility. What is required for justification and for free will are fundamentally different; and only the former can plausibly give rise to an infinite regression. We saw that there is a sense in which the concept of 'adequate sourcehood' is fitting in both the case of responsibility and of justification since each implies a kind of ownership. But since the structure of justification is so different from the structure of responsibility, this strongly suggests that 'sourcehood' does not refer to

the same property in each case or that the standards of adequacy differ fundamentally from premise 1 to premise 2.

The sceptic can therefore insist that the *Sourcehood Argument* equivocates: rationality requires sourcehood or adequacy in a different sense than moral responsibility (according to the sceptic). It is not, then, that 'adequate sourcehood' always means something different when it applies to actions and when it applies to beliefs. The Epicurean is right to maintain a symmetry between these things. Rather, 'adequate sourcehood' is different whether we are talking about what is necessary *for moral responsibility* and whether we are talking about what is necessary *for rationality*.

Despite having refuted the *Sourcehood Argument*, the free will sceptic has a concession to make. She must concede that when some compatibilist argues that we have free will because, roughly, we act for our own reasons, he is partly right. We *do* act for our own reasons, in the same sense that we believe for our own reasons. But this sourcehood, the sceptic must then argue, is different from the sourcehood necessary for moral responsibility. In a word:

> **Second lesson:** The free will sceptic must recognise the existence of some property of sourcehood, one that is necessary for people to believe and act for their own reasons and rationally.

The *Sourcehood Argument* fails but it urges the sceptic to explain why responsibility and justification use different notions or standards of sourcehood.

5. Leeway

As I indicated previously, some philosophers doubt the existence of free will on the grounds that we lack leeway. It suffices for our purpose to roughly indicate why. Typically, the leeway sceptic combines the Consequence Argument (see e.g. van Inwagen 2017) and the Mind Argument (see e.g. Pereboom 2014a). The Consequence Argument claims roughly that if determinism is true, then our actions are the consequence of the remote past and of the laws of nature. Thus, to act differently, we would have needed to change the laws of nature or the distant past, which is impossible. The Mind Argument says, crudely, that indeterminism is determinism plus chance, which is irrelevant to our abilities. Therefore, we lack the ability to do otherwise.

This suggests an Epicurean argument:

> If pessimistic determinism [free will scepticism] were true, no one could responsibly think that he had made a *wise* [or rational] *decision in believing* it. He had *no choice* but to believe it. (Dworkin 2011, 225, my emphasis)

The notion of choice is obviously tied to that of alternatives: Dworkin is not merely talking about the mental event of choosing, but the selection of one amongst several courses of action. Dworkin tells us, without much argument, that wisdom requires leeway.

A stronger philosophical treatment is given by Robert Lockie:

> [I]f determinism is true, then no-one can do otherwise and therefore *no-one may reason otherwise*. Assuming that the ability to reason otherwise is necessary for someone to be held epistemically irresponsible, no-one may then be held responsible for their intellectually wrong actions or unjustified, *irrational cognition*. But if no-one is responsible for their unjustified cognition, then no-one is epistemically justified either—in the intended, internalist sense. […] So one cannot be epistemically justified in claiming that determinism is true. (Lockie 2018, 183; my emphasis, premise numbering omitted)[19]

I shall argue that there is a simplified and less ambitious version of Lockie's argument. This argument does not suppose, controversially, that determinism is incompatible with leeway; it does not dip into the question of epistemic responsibilities; and it does not tackle the difficult topic of 'holding' responsible. My argument simply captures Lockie's insight that rationality ('internalist justification') requires leeway because leeway is required for obligations (responsibilities).

That leeway implies rationality cannot be established directly: it is false that if one believes a proposition rationally, one could have believed otherwise. A subject could believe a proposition rationally in the absence of any kind of leeway. Thankfully, most of us would be incapable of disbelieving that 2+2=4 and yet nothing impedes (directly) the rationality of our belief that 2+2=4. We need an indirect argument.

---

[19] In this chapter, Lockie discusses two related arguments, but they are of a different kind. The first (179) concludes that determinists cannot accept deontic ethics; the second (about the lazy argument, 178) fundamentally makes the claim that in deliberating we must presume that we have options.

5.1 The Leeway Argument

The Epicurean can instead pursue an indirect strategy and claim that if one *irrationally* believes a proposition, one could have believed otherwise. This principle is more plausible. For, suppose that a subject believes that 2+2=5 and could not believe otherwise. Then, the Epicurean may still declare that this subject's belief is *not* rational: it is perhaps *arational*, but not *irrational*.

Why would the Epicurean say this? Irrationality, she can say, implies a failure to satisfy obligations to believe and if leeway does not exist, then these obligations do not exist either. Why does the inexistence of leeway imply the inexistence of obligations? The Epicurean says: first, 'ought' implies 'can'—henceforth OIC. By *modus tollens*, 'no can Φ' implies 'no ought to Φ.' This principle, together with the thesis that we cannot do or believe otherwise, directly implies that we have no obligations to do or to believe otherwise. (I leave aside the abstruse view that we only have the obligation to perform the acts that we in fact perform and to believe the propositions that we in fact believe.)

Once we have admitted that the inexistence of leeway implies the inexistence of irrationality, we should admit that it implies the inexistence of *rationality* in general. Rationality and irrationality go hand in hand. This explains why it is quite natural to think that animals and toddlers are *a*rational rather than *ir*rational. This does not have to mean that for one to rationally believe that 2+2=4 one is able to believe otherwise *in this situation*; it simply means that the existence of rationality

in general implies the existence of irrationality in general.[20] We can make sense of the philosophical view that rationality *and* irrationality do not exist: rationality would be a property without extension (perhaps because they are no norms of rationality) and irrationality would not exist because irrationality is *absence of rationality where rationality could apply*. Everything would be *a*rational because arationality means the absence of rationality where rationality could not apply (e.g. the table is arational). But to make sense of the view that rationality exists, but not irrationality, we would need to imagine that if rationality can apply, then it exists. This only works in an ideal world, a place we do not live in.

It is now helpful to summarise the argument:

The Leeway Argument

If no one has leeway, there are no obligations.

If there are no obligations, no one believes irrationally.

If no one believes irrationally, no one believes rationally.

Therefore, if no one has leeway, no one believes rationally that free will scepticism is true.

Here again we have a typically Epicurean connection between conduct and beliefs in the two first premises (since 'no one has leeway' is at least partly about conduct).

If the Leeway Argument is sound, it follows that it is self-defeating to believe that free will scepticism is true on the grounds that we have no

[20] This argument might recall the theistic argument that the existence of evil implies the existence of good.

leeway. The presumed truth of free will scepticism ensures that either it is believed on false grounds (since we *have* leeway) or that it is not believed rationally (since we have no leeway and thus cannot believe rationally).

It is worth insisting at this point that, although the sceptic could admit the existence of leeway and find other grounds for her doctrine, this is a major consequence. As we just saw, some of the main arguments for free will scepticism feature leeway. And, although many sceptics[21] and non-sceptics[22] are now sourcehood theorists, they often concede in passing that, in a sense, we lack leeway. If the Leeway Argument is sound, this is problematic; belief leeway needs to be maintained for rationality.

This means that the sceptic should try to find a way to refute the Leeway Argument. I suggest that she abandons OIC, although this might be mistaken. Let me simply point out that OIC is popular amongst free will sceptics (see e.g. Pereboom 2014c, 222; cf. Waller 2011). Some sceptics have appealed to it in defence of another principle that is sometimes used by free will sceptics, the principle that we are only responsible if we could have done otherwise.[23] And if free will sceptics were to concede that we *can* do otherwise, it might be harder to argue that we are not ultimately the source of our actions.

---

[21] See Pereboom (1995, 27; 2014a, 138; 2014c, 221).

[22] This might include semi-compatibilists like Fischer (1994).

[23] On this, see e.g. Widerker (1991) and Copp (2008).

Instead of making this costly concession, the sceptic might want to follow a third way and declare that the Leeway Argument equivocates on the sense of 'obligation'. There are two ways to argue for this. The first is to oppose obligations to believe and obligation to act and insist that the second but not the first requires leeway. The second is to oppose epistemic obligations and moral obligations and maintain that the second but not the first requires leeway. I consider each in the following subsections.

5.2 Obligations to Believe, Obligations to Act

We have some power over our conduct and so while it is controversial that an obligation to act implies 'can', it remains plausible. By contrast, obligations to believe, one might suggest, cannot imply 'can' because we have obligations to believe, but we do not control directly our beliefs.[24] So, although we might have thoughts with the Epicurean that all obligations were similar in their requirements, this is incorrect: obligations to believe and to act are made of a different fabric.

While tempting, this counterargument is unavailable to our sceptic (and so there is no need to consider whether we directly control our beliefs). This is simply because the sceptic who accepts that we lack leeway, following the same reasoning, would have to conclude that the moral 'ought' does not imply 'can' either, since she denies the existence of leeway in conduct. But then the strategy is no longer a third-way

[24] This is precisely what Alston (1989) advocates.

escape focussing on the asymmetry between two kinds of obligations to preserve OIC. It throws OIC under the bus.

5.3 Epistemic Obligations, Moral Obligations

The contrast between epistemic and moral obligations might look more promising. As Clifford insisted, there are moral obligations to act, but also to believe. Similarly, there are epistemic obligations to believe, but there might also be epistemic obligations to act. So the epistemic/moral distinction cuts across the to believe/to act distinction. If I must choose between happily believing against my evidence and depressingly believing on my evidence, it is possible that the two kinds of normativity pull me in different direction. We can see this if we consider a case where a doctor could gather more evidence for the efficacy of a new vaccine or spend this time to administer it to patients. The general idea is that there are two kinds of normativity which aim at two kinds of value and which are subject to different requirements. Perhaps the anti-Epicurean can use this picture.

Her strategy would be to argue that moral, but not epistemic, obligations require alternatives. For this, the anti-Epicurean will try to show that epistemic normativity is like other 'minor' kinds of normativity. In fact, in general, certain kinds of obligations, like obligations of etiquette and professional obligations, do not imply 'can.' The Dean of the Faculty might have an obligation to attend a meeting even if he cannot be present—he might have been in a car accident. He thus fails to satisfy an obligation, although he is excused. To be sure, his

work-description does not specify that his obligations are suspended if he is unable to meet them, at least, not in all cases where he lacks the relevant abilities. Similarly, etiquette certainly recommends that a guest show appreciation for a meal, but this might be impossible if a food allergy forces her to rush to the bathroom. An obligation was infringed, despite the excuse. And the epistemic 'ought' might be very similar to etiquette and professional obligations as regards 'can': 'our friend in his tinfoil hat can't make himself stop overtly believing contradictions.' (Carr 2015, 752) But, surely, our friend in his tinfoil hat is not believing rationally and therefore he infringes an obligation of rationality.

But then, can't we say the same thing of moral obligations? There is an analogous character to our friend in his tinfoil hat—the inexorably evil man with his dictator cap. Surely, he infringes his moral obligations—and so the moral 'ought' does not imply the moral 'can'. Remember however that the anti-Epicurean strategy here is to keep "'moral ought" implies "can'" and reject "'epistemic ought" implies can'. So she must insist that somehow the dictator does not have the *true* moral obligation to refrain from 'dictating.'

She can make the following case. True obligations imply can (perhaps because their infringement is blameworthy), and other obligations—such as professional obligations, chess obligations, and epistemic obligations—are just rules that it would be ridiculous to insist 'imply can'. What distinguishes our friend in his tinfoil hat from our foe in his dictator cap is that the first fails to follow epistemic *rules* whereas

the second is not under moral *obligations*.[25] Satisfying epistemic rules or rules of etiquette then has no intrinsic significance: it does not matter, in and of itself, whether one has rational beliefs or not. The significance of following those rules is quite unlike, for instance, the significance of finding meaning in one's life or of respecting others. Hence, it is no surprise if general principles like OIC do not apply to epistemic obligations (or rules).

I am disinclined to admit this response because I think that at least some epistemic statuses, as opposed to statuses of etiquette, bear intrinsic or final significance or value. To be fair, some false beliefs seem too negligible to matter for their own: whether I truly believe that there is an odd number of grains of sand on this beach is negligible. And yet, it seems that *in general* having true beliefs—but not following rules of etiquette—*is* intrinsically or finally significant or valuable. We think it is intrinsically or finally valuable to be connected with the world—we wanted to be guided by the truth. And this connection, just is an epistemic connection (see Hyman 2015, 209).

Now, to be sure, the proposition that *knowledge* or *rationality* is finally or intrinsically valuable does not entail that *following epistemic* obligations is. But this seems enough to close the gap between ethics and epistemology and to widen the gap between epistemology and etiquette. (After all, perhaps *following moral rules* has no intrinsic value. Perhaps it is only resulting states of affairs that have intrinsic value.)

[25] See Côté-Bouchard (2017).

We should add that the anti-Epicurean is in no position to insist too much on individual cases of valueless knowledge or of true belief. For a similar problem will appear in the case of morality. Imagine that we have a pair of cases of identical situations of horrible actions except that in the first, but not in the second, the villain could not have done otherwise, and so infringes no moral obligation. Just like it seems indifferent whether we have a true belief or an item of knowledge about the number of grains of sands, it seems indifferent whether we are in the first or in the second case—whether the villain is infringing an obligation. We should not infer from these sorts of cases that there is no real epistemic or moral normativity.

It seems, then, that sceptics have a solid case for the symmetry of moral and epistemic obligations regarding leeway. This takes us back to the two other options, which are costly to the sceptic. Thus, we can say:

> **Third Lesson:** The free will sceptic must either: (1) accept that we may have leeway; or (2) reject OIC.

None of these avenues is blocked to the sceptic, but, as we have seen, each is at least *prima facie* very costly.

6. Conclusion

Epicurus' argument was *prima facie* tempting. There seemed to be something suspicious about claiming, on the one hand, that we have no free will but, on the other hand, engaging into fierce deliberation about what would be best and rational for us to achieve and believe.

Trying to give the best reading of Epicurus' argument helped us to identify some of the arguments underlying this initial sentiment. There was *something* to it: not just a confused feeling, but sourcehood-related and leeway-related arguments concluding that to believe in free will scepticism is self-defeating. Yet none of the considered arguments was fully successful. That's a major point in favour of the free will sceptic.

Still, our discussion left the sceptic with three lessons:

> **First lesson:** The free will sceptic cannot and should not rely on the doubtful claim that we act because of natural events rather than for causes.
>
> **Second lesson:** The free will sceptic must recognise the existence of some property of sourcehood, one that is necessary for people to believe and act for their own reasons and rationally.
>
> **Third Lesson:** The free will sceptic must either: (1) accept that we may have leeway; or (2) reject OIC.

These lessons should be added to the sceptic's breviary. What is their significance? The third lesson makes it much harder to be a free will sceptic of the leeway kind; and the first lesson deprives the sceptic of an important argument (from the exclusion concern). The obvious way forward for free will scepticism seems therefore to revolve around the ownership concern, according to which we cannot be responsible for our conduct because even if we act for our reasons, desires, or values, they are not as much ours (or significantly ours) as we thought; in fact, we own them because of factors lying outside of our control. This might

give confidence to the free will sceptics, who have mostly turned to this sort of view in the past decades (Strawson 1994; Pereboom 2001; Waller 2011; Pereboom 2014a). Yet sourcehood sceptics were mistaken in thinking that since leeway was not the crux of responsibility, they could deny the existence of leeway on top of sourcehood at no additional cost.[26]

Although, I did not have the chance to tackle this question seriously (see §2.2), I believe it would be a mistake for the sceptic to accept that her doctrine is self-defeating and insist that at least she has a true belief. After all, what made her view attractive was that she gave us powerful *reasons* to believe that some of our institutions are *unjustified* and some of our attitudes are *irrational*. To admit that being a sceptic is itself neither rational nor reasonable would undermine this. Free will scepticism is a philosophical doctrine supported by arguments, not a creed supported by hopes.

[26] See Pereboom (1995, 27; 2014a, 138; 2014c, 221).

V. FREE WILL SCEPTICISM AND THE MORAL LIFE

Abstract: Can the free will sceptic engage adequately in the moral life? This paper explores three related claims for thinking that she cannot. The first is that free will scepticism is incompatible with coherent and fair moral criticism. The second is that free will scepticism is incompatible with the existence of moral wrongfulness. The third is that free will sceptics cannot explain the moral significance of excuses. If these claims were correct, it would be self-defeating to pursue the practical programme of free will scepticism since this programme is a fundamentally moral one. But they are incorrect; or at least can be resisted by free will sceptics. Yet, each claim strongly suggests that, in order to engage in the moral life adequately, the free will sceptic needs to recognise the existence of moral responsibility in some sense.

1. Introduction

Free will scepticism claims that we lack the command or control that agents require in order to be morally responsible for their conduct. Since the days of Ancient Stoicism and of Epicureanism, this doctrine—and similar ones like determinism—has been discarded on the grounds that it is incompatible with the moral life. Amongst other things, the sceptic stands accused of being incapable of identifying wrongful actions in the world.

If this is true, then free will scepticism seems to lose much of its current traction. What makes it so appealing is precisely that it offers an

attractive ethics.[1] This scepticism is not a mere philosophical 'ism'. It is a doctrine which invites us to change our interpersonal relationships and political institutions. We should abandon blame and punishment and embrace understanding and reform, says the sceptic. We should refrain from blaming, resenting, and praising others, she adds. For it is wrong to punish or to blame those who lack free will.

In fact, free will sceptics claim not only that they can preserve the ethical life, but that they can improve it. If free will scepticism was incompatible with the moral life, this would be a disaster for the sceptic's sceptical programme. It could no longer be seriously considered an improvement on the actual system of moral responsibility. (As we shall see, it would also perhaps mean that it is self-defeating to pursue the practical sceptical programme.)

There is a sense in which free will sceptics could keep doing ethics any way. They could engage in moral *theorising*. Just like atheists can do theology, free will sceptics can do ethics. Theologians can discuss whether God would be omniscient *if* he existed; ethicists can discuss whether killing animals would be wrong *if* responsibility existed. This is important, because free will scepticism itself rests on moral principles, principles which must remain available to the sceptic. For instance, it rests on the principle that someone cannot be responsible for unavoidable events. No argument that I will consider in this paper

[1] See Spinoza's *Ethics*, Pereboom (2001; 2014a), Waller (2011), and Caruso (2018).

threatens *this*. However, I take it that free will sceptics want more than the possibility of conducting theoretical ethics—they want to be able to engage in the moral life.

In this paper, I explore three more precise claims which support the view that free will sceptics cannot engage in the moral life. The first, championed by Epicurus and Zeno, is that free will scepticism is incompatible with moral criticism. The second, maintained by Peter van Inwagen, is that free will scepticism is incompatible with the existence of moral wrong. The third, presented by Ronald Dworkin, is that free will sceptics cannot explain the moral significance of excuses. I critically assess each claim respectively in §2, §3, and §4.

I conclude that the three arguments fail. But they are all instructive, in a similar way: they force the sceptic to acknowledge the existence of some sense of moral responsibility. In particular, I think they put pressure on the sceptic to admit the existence of responsibilities, conceived as obligations or duties, and of responsibility, conceived as the capacity to be the bearer of obligations or duties. This, I think, is the only way for free will sceptics to fully partake in the moral life.

2. First Claim: Criticism Implies Responsibility

This section examines a first anti-sceptical claim: it is problematic (either unfair or incoherent) for the sceptic to *criticise* someone for rejecting free will scepticism. If this is true, it follows that the sceptical programme is hardly feasible. This programme partly aims at convincing people to

recognise the inexistence of free will and to abandon blame and punishment. For this purpose, criticism seems indispensable, since the sceptic's project is not merely to feed people 'anti-blame pills', but to eliminate allegedly irrational practices by means of philosophical arguments.

The history of this first anti-sceptical claim starts in antiquity:

> The man who says that all things come to pass by necessity [the free will sceptic][2] cannot criticise (prosecute, ἐγκαλέω) one who denies that all things come to pass by necessity: for he admits that this too happens of necessity. (Epicurus, *Extant Remains*, Frag. XL; translation Bailey 1926, 112)

When Epicurus claims that the free will sceptic 'cannot criticise' her opponent, he must mean that it would be unfair or incoherent for the free will sceptic to do so, not that it would be impossible. It is unfair for the free will sceptic to criticise others because if free will scepticism is true, then no one deserves to be criticised or blamed. Alternatively, it is incoherent for the free will sceptic to criticise others because free will sceptics themselves claim that criticism of others is an illegitimate practice. This means that to criticise others *for failing to accept free will scepticism* is self-defeating.[3] If the criticism is correct—namely if free will

---

[2] Elsewhere (Conditional acceptance, Chevarie-Cossette), I argue that this argument is stronger when it targets the sceptic rather than the determinist.

[3] Epicurus himself uses a term similar to 'self-defeating' in another passage:

> This sort of account (λόγος') is self-refuting (τρέπω περικάτω), and can never prove that everything is of the kind called 'necessitated'; but he [the free will sceptic] debates this very question on the assumption that his opponent is himself responsible (δι ἑαυτοῦ) for talking nonsense. (Epicurus, *On*

scepticism is true—then it is unfair or incoherent, because the person criticised lacks free will and thus responsibility for their conduct.[4]

The Stoic Zeno of Citium made a similar point:

> They say that he [Zeno] was once scourging a slave whom he had detected in theft; and when he said to him, 'It was fated that I should steal' he rejoined, 'Yes, and that you should be beaten.' (D.L. VII, 23, translation Yonge 1853)

Zeno does not think that his slave's rejoinder is convincing. The fact that all events are fated is not a reason to exculpate him. For, if it were, it would also be a reason to exculpate Zeno for punishing his slave. Thus, either the slave is wrong, or Zeno is excused in advance for the punishment he inflicts on his slave; hence, the punishment.

Zeno and Epicurus diverge in their approach to moral responsibility. To admit the existence of moral responsibility, Epicurus rejects the claim that all things come to pass by necessity while Zeno, like other Stoics, accepts it but disagrees that this implies that humans are unfree or non-responsible. (To use modern terms, Epicurus is an indeterminist while Zeno is a compatibilist.) Regardless of this divergence, Zeno and Epicurus agree that the free will sceptic engages in an activity that she

---

Nature. XXV, 34; translation Annas, Taylor, and Sedley 1983, 19–23)

If we understand 'responsibility' as 'moral responsibility', then we get a similar point. (I have defended a different reading elsewhere.)

[4] Elsewhere (Chevarie-Cossette 2019a), I argue that a propositional attitude (such as belief) is self-defeating when the presumed truth of the content of the attitude ensures that the attitude is problematic.

does not have the right to. Moral criticism cannot be part of the free will sceptic's moral repertoire.

In *Justice for Hedgehogs*, Ronald Dworkin makes a similar claim:

> Some lawyers and criminologists insist that we must abandon traditional criminal law, with its apparatus of guilt and punishment, and substitute only therapeutic treatment because people are never responsible for what they do. They contradict themselves. If no one ever has judgemental responsibility, then officials who treat accused criminals as responsible for their actions are not responsible for their own actions, and it is therefore wrong to accuse them of acting unfairly. Of course, it would then also be wrong of me to accuse the criminologists of acting wrongly in accusing the officials of acting wrongly, because the criminologists are not responsible either. […] This recursive nonsense shows, even if nothing else did, that we cannot believe the proposition on which it hinges, which is that we all lack judgmental responsibility for anything. (Dworkin 2011, 224–25, emphasis mine)

Like Epicurus, Dworkin accuses free will sceptics of being unfair and incoherent. For he assumes that the putative truth of free will scepticism makes criticism unfair or wrong. Dworkin also assumes that criticism is prohibited by free will sceptics—hence the accusation of contradiction. In fact, Dworkin believes that free will sceptics deny what he calls 'judgemental responsibility', which he defines in terms of liability to

criticism or praise (2011, 103). This suggests that he equates blame and criticism, for it is *liability to blame* that free will sceptics reject.[5]

It seems however that the accusation of incoherence and the accusation of unfairness are misplaced precisely because we can distinguish blame from criticism. The sceptic is coherent because he only asks that we do not *blame* one another; not that we do not *criticise* one another. And the sceptic is fair because he *criticises* those whom he considers lack free will, but he does not *blame* them. In virtue of this distinction, the sceptic makes the following assessment of Epicurus' and Zeno's men. Epicurus' man, who claims that all things come to pass by necessity, can (fairly and coherently) insist that what his interlocutor maintains is *both* incorrect *and* that it comes to pass by necessity. Similarly, Zeno's slave can (fairly and coherently) claim that his theft was wrong and thus criticisable; that Zeno's punishment would be wrong and thus criticisable; *but also* that neither the theft nor the punishment is the correct object of *blame* or *punishment*.

In other terms, the sceptic believes that Epicurus, Zeno, and Dworkin fail to distinguish *evaluative* concepts (linked to the activity of valuing) such as criticism and judgement and *hypological* concepts (linked to the activity of holding accountable or liable[6]) such as blame and

[5] Van Inwagen makes a similar equation when he declares that 'one cannot simultaneously excuse and censure' (1983, 208).

[6] The term was coined by Zimmerman (2002) from ὑπόλογος (hypologos), which means to hold accountable or liable.

responsibility. Once we realise that criticism and blame are conceptually distinct, claims the sceptic, we can retain the first and reject the second.

This response leaves two problems unanswered. The first is suggested by Zeno's attitude. By ignoring his slave's rejoinder, he seems to assume that if his slave is right, then he, Zeno, has no obligation to respond to criticism. Differently put, if free will scepticism is true, no one has to listen to criticism. If this is right, this is a problem for sceptics: they do not want merely to be able to criticise others; they want others to respond appropriately and they hope that others feel obliged to respond. Consider blame. If I blame someone and I am right to do so, then I have a *claim* against them: minimally, a right to demand that they acknowledge fault and make amends; and perhaps an expectation that they feel regret or shame. This sort of claim, we might think, is precisely what distinguishes blame from mere criticism. But when the sceptic criticises others, she acts as if she had this sort of claim over those that she criticises. She expects her non-sceptical audience to take ownership and make amends; and perhaps also to accept bad consequences and feel regret. This looks like blame in all but the name. (Zeno wouldn't fall for this trick!)

I think that there is something to this response; but fully assessing it would require that we evaluate different accounts of blame, something for which we lack the space.[7] However, I think that the sceptic can respond roughly as follows. To fully participate in the moral life, the

[7] See Wallace (1994) and Scanlon (2008), for instance.

sceptic needs to be able to morally criticise others. On top of this, she must be able to generate the responsibility to respond (i.e. the *duty* or *obligation* to respond) as a result of criticism. The criticised person *ought to* assess whether the criticism is right; and if it is, try to make amends. But the emotional and retributive elements of blame—the feeling of shame or regret; and the punishment or opprobrium—can be set aside. Obviously, for this response to work, the sceptic must admit the existence of responsibilities in the sense of duties or obligations—and this is something that not all sceptics are ready to do, as we shall see in §5.

To dispel any possible confusion, we should distinguish explicitly different senses of the word 'responsibility', two of which we have encountered:

> **1. A liability relation:** the relation one has with one's conduct which makes one a fitting or deserving[8] subject of blame or punishment (if it is bad) or of praise or reward (if it is good). (E.g. 'He is responsible for stealing from the poor.')
>
> **2. An obligation:** the modal fact that one is obliged to do or refrain from doing something. (E.g. 'The Little Prince was responsible for his rose.')
>
> **3. A capacity:** the capacity that one requires in order to participate in the moral life—perhaps including in order to be the bearer of moral obligations. (E.g. 'Do not open this package unless accompanied with a responsible adult.')

[8] See Haas (2013) for the difference.

> **4. A virtue:** the disposition to uphold some of one's moral obligations. (E.g. 'Give her the job: she's a responsible worker.')[9]

In what follows, I will ignore virtue-responsibility. The free will sceptic claims that we should reject moral responsibility understood in the *first* sense, i.e as liability for praise or blame. Some free will sceptics add that obligations also do not exist—as such, they reject moral responsibility in the *second* and perhaps *third* sense. We will meet these sceptics again in §5.

So far so good. But there is a second problem with our response to Epicurus' claim. Perhaps it does not matter whether criticism and blame are distinct attitudes or activities. For, criticism might *also* be unavailable to the sceptic. This is the problem to which I now turn.

3. Second Claim: Wrongfulness Implies Responsibility

It seems that we can draw a conceptual distinction between moral criticism and blame; and it also seems that we can criticise without blaming in virtue of the fact that we can omit the emotional and retributive aspects of blame. But there remains a problem concerning criticism. The problem is not so much about the interpersonal aspect of criticism as Epicurus and Zeno seemed to suggest, but about its

[9] This typology of responsibility is presented in (Chevarie-Cossette 2017). It partly follows H.L.A Hart's (1968, 211). Hart distinguishes four senses of responsibility: role-responsibility (duties), causal-responsibility, liability-responsibility, and capacity-responsibility. I prefer to amend this nomenclature in a few ways since we are only interested in senses of *moral* responsibility. First, we add *virtue-responsibility*. Second, we take role responsibilities to encompass all moral obligations—since, in a sense, *qua* moral agents, it is our role to respect all our moral obligations. Third, we ignore 'causal responsibility' since it is not moral.

cognitive aspect: moral judgement or evaluation. In particular, free will sceptics cannot coherently judge that non-sceptics are guilty of moral wrongdoing, or so goes the accusation. This is simply because someone's actions or omissions are wrong only if she is responsible for them. If this is right, then it is an additional difficulty for the free will sceptic on her path to participating in the moral life. For, she is not only unable to criticise others, but she is unable to judge that they have acted wrongly.

The view that an action is wrong only if its subject is responsible is presented by Peter van Inwagen[10]:

> I have listened to philosophers who deny the existence of moral responsibility. I cannot take them seriously. I know a philosopher who has written a paper in which he denies the reality of moral responsibility. And yet this same philosopher, when certain of his books were stolen, said, "That was a *shoddy* thing to do!" But no one can consistently say that a certain act was a shoddy thing to do *and* say that its agent was not morally responsible when he performed it. (van Inwagen 1983, 207)

Van Inwagen makes similar remarks about other moral concepts, such as 'despicable'. An action cannot be despicable unless someone is responsible for it. More generally,

> If *S*'s φing (where φ is an act or omission) is morally wrong, then *S* is morally responsible for φing.
>
> Free will scepticism implies that, for any φ, no one is morally responsible for φing.

[10] See also Campbell (2013) and Wolf (1981, 386).

> *Therefore*, the sceptic cannot coherently judge that anything is wrong.

From this, we can conclude that believing that an act is wrong on the grounds that free will scepticism is true is self-defeating. For instance, it is self-defeating to believe that the current system of private property is wrong on the grounds that no one is responsible for anything (including the fact that they are wealthy). The truth of the claim ensures that it is believed on false grounds, since it implies that free will scepticism is false.

Van Inwagen's view that *if S's φing (where φ is an act or omission) is morally wrong, then S is morally responsible for φing* is clearly controversial. For instance, act consequentialists deny this view. Act consequentialism claims that φ is morally wrong just if it is produces (or has a tendency to produce) certain bad consequences. Clearly, whether my conduct has bad consequences can come apart from whether I am responsible for it. By calling a friend for his birthday, I can cause a car crash—the ringtone might distract him at the fatal moment. But I am not responsible for his death. So, for act consequentialists, responsibility and wrongfulness come apart.

Of course, act consequentialists must be able to maintain that if you push me out of the window and that I fall on a bystander, *I* did nothing wrong, despite my body's relevance in the causal chain which leads to the production of pain. At first, the act consequentialists might suggest that the difference is that in the case where I am pushed, I am not *acting*.

This will not do, as the act consequentialist does not want to put on par the case where I am pushed and the case where I omit something important (and where I am, again, not *acting*). As a second pass, the act consequentialist will have to give a certain story about how my mental life (my mental states, my reasons, my desires) is explanatorily relevant in cases like an omission or a normal action which leads to the production of pain, but not in the case where I am pushed out of the window. And, they should continue, the relevance of this mental life does not commit them to admit the existence of moral responsibility.

Ronald Dworkin, a friend of the claim that wrongfulness implies responsibility, is not impressed by this act consequentialist response:

> [Free will sceptics assume that] […] [if] no one was ever blameworthy, it would not follow that nothing was ever the wrong thing to do. But that depends on why we think certain acts are wrong. On any plausible theory, concepts of responsibility figure just as pervasively in identifying wrongful acts as in adjudicating blame for those acts. Yes, some philosophers, including some utilitarians, believe that an act is wrong if it has bad consequences, whatever the mental state of the agent. Someone who gives to charity acts wrongly if, unknown and unknowable to him, he would have improved overall happiness by spending on his own pleasure instead. This is not plausible. (Dworkin 2011, 142)

It is obviously not my purpose in this paper to determine whether utilitarianism (or other forms of consequentialism) is correct. As I said, I think it is true that for act consequentialists, wrongdoing does not imply

responsibility.[11] We can, at least for this discussion, grant Dworkin that act consequentialism is false.[12]

However, it is not obvious that rejecting act consequentialism can help Dworkin and van Inwagen to show that wrongfulness implies responsibility. Here is John Stuart Mill:

> The reality of moral distinctions, and the freedom of our volitions, are questions independent of one another. My position is, that a human being who loves, disinterestedly and consistently, his fellow creatures and whatever tends to their good […] is naturally, necessarily, and reasonably an object to be loved, admired, sympathized with, and in all ways cherished and encouraged by mankind; while a person who has none of these qualities […] is a natural and legitimate object of their fixed aversion, and of conduct conformable thereto: *and this whether the will be free or not, and even independently of any theory of the difference between right and wrong*; whether right means productive of happiness, and wrong productive of misery, or right and wrong are intrinsic qualities of the actions themselves, provided only we recognise that there is a difference, and that the difference is highly important. What I maintain is, that this is a sufficient distinction between moral good and evil […] that if there be

[11] Slote (1990, 369–72) argues very convincingly for this view. However, in its crudest version, utilitarianism claims one of two things. Either there is no such thing as moral responsibility, in which case the question of whether free will sceptics can engage in utilitarian ethics is trivially 'yes'. Or there is such a thing as moral responsibility, for one is responsible if and only if it would bring good consequences to blame or praise him (e.g. Smart 1961). On this second reading, free will scepticism is implausible.

[12] However, Dworkin's example is not a good counterexample to *rule* utilitarianism because the action-type 'giving to charity' still has good consequences in terms of utility in general. But this is unimportant: other counterexamples could be found.

> another distinction, we can dispense with it. (Mill, [1865] 1979, 456–57, emphasis mine)

Mill states eloquently the view that ethics without free will is possible *regardless of whether consequentialism is true*. Badness—and perhaps also wrongness—does not imply free will.[13]

This view is common. Free will sceptics have responded on a number of occasions to van Inwagen's claim by drawing the evaluative–hypological distinction I made in §2 (it's one thing to criticise; quite another to blame).[14] To make this point clearer, it is useful to briefly recall the opposition between volitionists and attributionists. Volitionists[15] believe that we are only directly responsible for our conduct if it lies within our control—and typically, for our choices and what results from them. Attributionists[16] believe that we are responsible for all our actions and omissions which reflect our evaluative attitudes, whether or not these attitudes are the result of our choices and whether or not they are under our control.

The difference between attributionists and volitionists is clearest when considering cases of implicit biases. Implicit biases are evaluative attitudes which impact our conduct unbeknownst to us (unbeknownst to us at the moment of their influence and perhaps against our best

---

[13] Waller (2011, 189) quotes Mill as talking about responsibility.

[14] See Slote (1990), Otsuka (1998), Zimmerman (2002), Waller (2004, 2011), Pereboom (2014a), and Caruso (2018).

[15] See e.g. Levy (2005), Fischer (2006).

[16] See e.g. Scanlon (1998), Smith (2005).

intention).[17] Some of these biases are mistaken; others are not only mistaken but morally problematic. For instance, suppose that I implicitly judge that people of colour are lazy. I am not aware that I have implicitly judged this; if I had, I would be ashamed of this attitude and I would try to eliminate it. Add, if you want, that I live in the 1960s and that I could not have encountered the concept of implicit bias in my studies. Now, I happen to reject a candidate for a job precisely because of this bias. Attributionists often argue that they are in a better position to morally assess this sort of situation than volitionists. They claim that I am blameworthy for my conduct during the interview process, since it poorly reflects on my evaluative attitudes. Volitionists respond that this is not clear—it depends on how much control I exercised. In particular, we need to make sure before blaming me that my racist evaluative attitudes are the result of a wrong choice on my part. Attributionists think that we will leave too many wrongdoers off the hook; volitionists are prone to thinking that attributionists are being unfair.[18]

Crucially, the case of implicit biases clearly points to a common assumption of volitionists and attributionists, namely that, whether or not I am responsible for my racist conduct, it is *morally wrong*. It is morally wrong to discriminate on the basis of prejudice against people of

[17] We can be aware that we have such attitudes without making them explicit; for they remain under our radar at the precise moment when they are causally efficacious.

[18] It is unfair because to hold someone responsible for something outside of their control is analogous to holding someone responsible for their height. See Levy (2005, 9–11).

colour, *whether or not this prejudice is within my control and whether or not it results from my choices*. It is morally wrong for multiple reasons (reasons which do not imply that consequentialism is true): it results from a wrong evaluation; it shows insufficiently good will towards certain people; it manifests bad dispositions and perhaps even vices. In fact, the distinction between volitionists and attributionists makes no sense without the unequivocal distinction between morally wrong actions and blameworthy actions. For it is about whether blame is appropriate for conduct which is *morally wrong but uncontrolled* that the two sides disagree. I think that this is right.

There is an amended version of van Inwagen's claim which is compatible with what I have just said. He could perhaps concede that in specific cases, an agent can act wrongly without being responsible for her act. But he could add that if an agent acts wrongly, she is a responsible *agent*, that is responsible in the sense in which a human adult, but not a child or an animal, is responsible. (I think that we can draw the distinction by claiming that the first, but not the second, is the potential bearer of obligations.) The advantage of this revised claim is that it allows us to reject a seemingly absurd view about the range of moral wrongness. Most non-consequentialists do not believe that e.g. hippos stampeding humans or toddlers being ungrateful is *morally wrong*. We *could* say that the wrongness of an action is only a matter of intention and that the hippos are guilty of moral wrongs since they have violent intentions when they charge others; and we *could* say that a toddler who refuses to

say 'thank you' to his grand-mother is guilty of a moral wrong if it is because he lacks good intention. But this seems problematic. The revised claim can dispense us of having to say this while acknowledging that *that φ is wrong implies that someone is responsible for φ.*

This, then, suggests the following anti-sceptical line:

> If *S*'s φing (where φ is an act or omission) is morally wrong, then *S* is a *morally responsible agent*.
>
> Free will scepticism implies that no one is a morally responsible agent.
>
> *Therefore*, the sceptic cannot coherently judge that anything is wrong.

Yet, as I have suggested at the end of §2, sceptics do not need to deny the existence of moral responsibility in this sense—and so they can reject the second premise. They need to remain careful not to accept just any notion of 'morally responsible agent', though. For instance, if by 'morally responsible agent', we mean the capacity for being liable for praise or blame then the sceptic should reject it.[19] We already see that having distinguished different senses of 'moral responsibility' is fruitful for the free will sceptic; this distinction continues to be useful in the next section.

---

[19] The general sceptical strategy should be the following: (1) identify what distinguishes the conduct of animals and kids from the conduct of competent humans; (2) if this alleged difference is incompatible with the sceptical doctrine (for instance, if the alleged difference is the ability to do otherwise), deny the moral difference between animal conduct and human conduct; (3) if this alleged difference is compatible with the sceptical doctrine, maintain that it is this difference which should define what a 'morally responsible agent' is and accept the existence of morally responsible agents in this sense.

4. Third Claim: Without Responsibility, Excuses and Exemptions are Insignificant

This slowly takes us to our third and final anti-sceptical claim. For, so far, we have argued that the sceptic had access to certain conceptual distinctions and that she could, in virtue of this, distinguish right and wrong actions in the world. But Dworkin has something left to say:

> [Free will sceptics] insist that [their doctrine] would not rule out any number of other ethical or moral judgements. It would still be possible, they say, to declare that a criminal has done something morally wrong even though he is not blameworthy for having done it. Or that he has a bad character. It would still be possible to think, they say, that someone has acted prudently or imprudently or that some states of affairs are better than others. I believe all this is wrong. Morality is an integrated web of standards. It is not a collection of detachable modules, each of which can be eliminated, leaving the rest even more or less intact. Judgmental responsibility is the weft of all moral fabric. (Dworkin 2011, 224)

The claim that (liability) responsibility is the weft of all moral fabric is hard to interpret. Responsibility is supposed to be a central notion—perhaps one in terms of which all the other moral notions must understood.

This is not yet to say that sceptics cannot engage in the ethical life. For, sceptics do not reject the existence of the *concept* of liability to praise and blame; they simply think that it is a concept which lacks an extension in this world. Thus, the sceptic could, at least in principle, define a wrong

action as one which would be blameworthy if free will existed. The sceptic would then give, in ethics, the conceptual lion's share to the concept of liability responsibility without committing to its existence. Remember the parallel we made about the atheist and the sceptic: as the first can do theology, so the second can do theoretical ethics.

But Dworkin has yet more to say. There are certain essential moral distinctions which the free will sceptic might be able to draw, but whose moral significance the free will sceptic cannot explain. So the claim is no longer that the free will sceptic cannot engage in the moral life because she cannot coherently judge that some actions are wrong. Rather, the claim is that the free will sceptic cannot engage in the moral life (at least, adequately) because she cannot make sense of the ethical significance of some essential moral distinctions. The sceptic has no reason to acknowledge the significance of distinctions which *ought to be significant*. What are these distinctions?

Here is Dworkin again:

> [O]ur concepts of responsibility, right and wrong, character and prudence are interwoven in such a way that we could not sustain any of the practices built around those concepts if we abandoned the distinction between those acts for which people are and are not judgmentally responsible. We would lose an essential pillar of our ideas about wrongful acts, for instance, if we were not able to distinguish them from accidents or tics or what madmen do. (Dworkin 2013, 142)

He mentions a new distinction here: the distinction between accidental or compulsive conduct and other conduct. This might remind the reader that Dworkin brought up a similar sort of distinction earlier:

> Someone who gives to charity acts wrongly if, unknown and unknowable to him, he would have improved overall happiness by spending on his own pleasure instead. This is not plausible. (Dworkin 2011, 142)

Here, Dworkin distinguishes conduct which results from ignorance and conduct which does not.

Again, Dworkin's point is *not* best interpreted as the claim that free will sceptics cannot draw some critical moral distinctions. In particular, the point should not be that free will sceptics cannot distinguish conduct which results from ignorance and conduct which does not; accidental or compulsive and non-accidental or intentional conduct (alluded to by the example of tics); and between the conduct of mentally disabled individuals ('madmen') and children and the conduct of others. For free will sceptics firstly do not question the existence of epistemological achievements, such as knowledge and justification; and of epistemological failures, such as irrationality and ignorance. Secondly, they do not question the existence of a distinction between accidental (or compulsive) and non-accidental (or non-compulsive) conduct because intentions play a different role in each (perhaps a different causal role).

Thirdly, they accept that mentally disabled individuals and toddlers have more limited intellectual capacities than others.[20]

Dworkin's point is better interpreted as the view that free will sceptics cannot explain the ethical significance of the three aforementioned distinctions and similar ones. In a word, free will sceptics cannot explain the ethical significance of *excuses* such as ignorance and compulsion or of *exemptions* such as mental disability or childhood; that is, they cannot explain why excuses and exemptions are relevant to the evaluation and criticism of conduct.

Consider again the parallel between atheists and sceptics. Atheists *can* draw a distinction between sins (in its religious sense) and good deeds. But they have no reason for thinking that this distinction is morally significant. Dworkin's idea, I think, is that free will sceptics are in the same situation: they can draw a distinction between conduct which is excusable or exempt from criticism and conduct which is not. But they have no reason for thinking that this distinction is morally significant because in a sense *everyone is excused or exempted* for them.

I think that Dworkin is right to think that excuses and exemptions are important to evaluation and criticism of conduct (even though he does not specifically speak in these terms); and I think that in order to engage adequately in the moral life, we need to be able to acknowledge or perhaps even to explain the significance of specific excuses and

[20] This is true even if they are the kind of free will sceptics who believe that, because of determinism, no one has the opportunity to act differently than they do.

exemptions. But I disagree that free will sceptics cannot acknowledge or explain this significance.

To explain the significance of excuses, free will sceptics can first maintain that they track good intentions or good will; or absence of bad intentions or of bad will. One crucial difference between the ignorant wrongdoing and the non-ignorant wrongdoing is roughly that in the second, but not in the first, the agent manifests bad intentions or bad will. If I had known that calling my friend would provoke a car crash, I would not have called him—I did not intend that he be hurt. The difference between compulsive 'wrongdoing' and non-ignorant wrongdoing is again roughly that in the second, but not in the first, the agent manifests bad intentions or bad will. If I break the ceremonial minute of silence because I have hay fever or if I perjure myself in court because of a threat, I do not manifest the same bad intentions or bad will as if I break the silence for the pleasure of offending old people or if I perjure myself because I refuse to answer for my conduct.

In several cases, the significance of exemptions can be explained in a similar way, i.e. by the difference of intentions or of quality of will. The person affected by severe dementia often does not mean the violent words that she utters—perhaps she does not understand what she is talking about at all. And the toddler is also unaware of the fact that he will make his grandmother sad by not greeting her. This explains why Cicero believes that ignorance is the right plea for several categories of exempted people:

> Why do you bear with the delirium of a sick man, or the ravings of a madman, or the impudent blows of a child? Because, of course, they evidently do not know what they are doing: a man be not responsible for his actions, what does it matter by what malady he became so: the plea of ignorance holds equally good in every case.' (*De Ira*, 3, XXVI, (trad. Stewart 1900))

If this is right, then the moral significance of the distinction between the conduct of the exempted and of the non-exempted can be explained like the moral significance of the distinction between excused conduct and non-excused conduct. And if this is right, free will sceptics can acknowledge and explain the ethical significance of the distinctions that Dworkin accused them of sacrificing. Liability is not then the 'weft of all moral fabric'.

5. A Different Weft: Responsibilities

This response is too quick (but, as I shall argue, it can be improved). For it should be clear that the significance of exemptions cannot always be given in terms of intentions or quality of will. Severely mentally disabled people and young children still manifest terrible intentions or show ill will from time to time. Some people affected by Alzheimer's say things which are particularly vicious—this often manifests a lack of control, not the ignorance that these words are mean. This contradicts Cicero: the ignorance plea might not be available in these cases.[21]

[21] The case of Tourette is especially clear here, although it does not contradict Cicero's quoted passage directly. A Tourette knows that the words she utters

There is a similar problem about excuses. Not all forms of ignorance and compulsion are born equal.[22] Some of them constitute good excuses, some of them do not. More importantly, conduct which results from bad excuses and from good excuses can involve the exact same kind of intentions or will.

Consider two pairs of cases. The first pair is about ignorance. We have, on the one hand, a rich man from the 19th century who heats his house with a coal furnace, completely unaware of the carbon emissions that will ensue. We have, on the other hand, a contemporary rich man whose main source of information is Fox News and who decides to heat his house using gas because he sincerely believes that global warming is a hoax. It seems quite clear that there is a normative difference between the conduct of the first and the second man, even if there is no significant difference of intention or quality of will. Turning to compelled behaviour. If, on the one hand, I am chosen by friends as the provider of dessert for tonight's party but I give the delicious *kouign amanns* I bought to a hungry burglar who demands it, it seems that I have a good excuse: I was forced or compelled to do so. If, on the other hand, I am entrusted by the State with nuclear plans and sworn to protect them and I nonetheless give them to the same threatening burglar, I lack a good excuse: considering the circumstances, it would be a stretch to say

---

are blasphemous and that blasphemy in public space can make people feel uncomfortable. But she can't help it.

[22] I follow Aristotle (*EE*) and Hyman (2015, chap. 4) in taking compulsion and ignorance as the main excuses.

that I was forced or compelled. But I may have the very same good intentions and good will in each case: I mean no harm to anyone; I simply want to avoid being harmed.[23]

The examples about ignorance and about coercion make it clear: intentions or quality of will cannot always explain the moral significance of excuses. And we have seen that intentions cannot always explain the moral significance of exemptions either. Does this mean that free will sceptics cannot after all acknowledge or explain the significance of exemptions or excuses—or that they cannot always do this or do this adequately?

No, because free will sceptics can admit the existence of moral responsibilities in the sense of duties or obligations. And this is sufficient to resolve their difficulties. If duties or obligations exist, then they can be used to explain the difference between cases where ignorance excuses and cases where ignorance does not. The ignorance of the 19th century man is excusable because he has no obligation to know that carbon emissions are damaging for the environment. The ignorance of the modern Fox News enthusiast is not excusable. The same can be said about cases of coercion. I have no duty to protect the desserts against the hungry burglar; but I do have the duty to resist to protect the nuclear plans I was entrusted with. More generally, the view is that free will

[23] The example of coercion might—contrary to the ignorance case—misleading suggest that what makes the difference here is merely the *stakes*. The difficulty here is that obligations are stakes sensitive. But we can imagine a case where my promise (or, say, my enrolment as a police officer) increases significantly how much coercion I should be ready to face.

sceptics can explain the ethical significance of excuses correctly by combining observations about intentions or quality of will and observations about responsibilities.

The same explanation is available for cases of exemption. Toddlers and the severely handicapped in general lack the capacity for moral responsibilities as obligations or duties. This is often useful to explain the significance of the occasional ill-intentioned conduct of exempted agents: their conduct is not as great a cause for concern as the conduct of ill-intentioned adults because they, unlike ill-intentioned adults, do not infringe obligations.

What this means is that the free will sceptic can acknowledge and explain the ethical significance of excuses and of exemptions. (Whether this explanation is sufficient or whether it can be completed is a story for another occasion.) But this comes at a price: the free will sceptic must recognise the existence of responsibilities conceived as obligations or duties. I say that this comes at a price because in fact some free will sceptics deny the existence of obligations.[24] Why do these sceptics believe that obligations require free will? Roughly, they endorse the principle that "'ought" implies "can'" and argue that we lack the ability to choose amongst true alternatives. But, they continue, it is implausible

[24] Alternatively, they define obligations in an axiological sense which does not allow us from distinguishing the pairs of cases that we scrutinised. Consider the view that for *S* to have the obligation to φ just is for φing to be an action that it would be good that *S* performs. This notion is of no help to distinguish the 19th century man from the contemporary Fox News enthusiast.

that the only obligations we have are the ones we fulfil. Therefore, obligations require free will.[25]

As I just argued, Dworkin's claim is effective against these sceptics.[26] In his words, the 'module' of responsibilities cannot be detached from morality. If one of these three suggestions (about coercion, about ignorance, or about disability) is right, then Dworkin's argument works against those free will sceptics who deny the existence of liability-responsibility *and* obligation-responsibility. If there is a way around for this sort of sceptic, she will need to show it.[27] In the meantime, I maintain that responsibility, in the sense of a duty or of an obligation but not in the sense of a liability to blame, is the weft of all moral fabric.[28]

---

[25] Haji (2012) has argued extensively for the view that *determinism* is incompatible with obligations. Pereboom (2014c) claims that obligations claims can be true if we understand them as claims about what would be good to happen.

[26] Waller (2011, chap. 10) argues that moral responsibility scepticism is compatible with ethics partly by arguing against "'ought" implies "can'". But he does not directly argue, as I do here, that if sceptics deny the existence of obligations, they cannot fully participate in the moral life.

[27] There might be further problems for the denier of obligations. One of them is that absent obligations, it is difficult to make sense of virtues and rights. It is difficult to make sense of virtues because virtues are dispositions to satisfy obligations or duties for the right reasons. And it is difficult to make sense of rights because for *S* to have a right just is for others to have an obligation towards *S*; see Nelkin (2014).

[28] We should note that since "'ought" implies "can'" and certain forms of scepticism imply that we have no obligations; and since we cannot participate in the moral life without recognising the existence of obligations, then these sceptics must deny that "'ought" implies "can'".

6. Conclusion

Many philosophers consider that the type of objection that Epicurus raised against free will sceptics is fallacious if not silly. Others think that these objections are resoundingly successful. They are all wrong. The objections fail, but for deep and instructive philosophical reasons.

First, Epicurus' claim compels the sceptic to distinguish criticism and blame precisely and explain how the first is possible without the second. This claim suggests that there is more to criticism than an objective evaluation—in its most important sense, legitimate criticism imposes responsibilities (obligations or duties) on its targets. The existence of such responsibilities should be acknowledged by free will sceptics who want to engage in moral criticism.

Second, van Inwagen's claim forces the sceptic to give a clear answer to the question of how an act can be wrong if no one is responsible for it. It can also be turned into a more plausible claim—a claim according to which an act is wrong only if its subject has, unlike animals and toddlers, 'capacity responsibility' (the capacity to bear obligations). If this claim is true, then this forces the sceptic to admit the existence of such a capacity.

Third, Dworkin's claim about the ethical significance of excuses and exemptions shows that free will sceptics cannot engage in the ethical life adequately unless they admit the existence of moral responsibilities. Facts about quality of will or intention do not suffice to explain the significance of excuses and exemptions.

Can free will sceptics participate in the ethical life? It seems so. But it also seems that it is a mistake to consider this answer an obvious one—that is, an answer which anyone who understands the difference between the hypological and the axiological should immediately admit. To draw conceptual distinctions is insufficient to rebut Epicurus', van Inwagen's, and Dworkin's views. In their most refined versions, their challenges allow for this sort of moral distinction; but they also draw important moral connections: between criticism and blame; between wrongness and responsibility; or between excuses and obligations. They remind us that the existence of a distinction does not imply the inexistence of a connection.

VI. SCEPTICAL DELIBERATIONS

Conditionally accepted, in a different version, in the *Pacific Philosophical Quarterly*

Abstract: This paper has three aims. First, it provides a unified account of practical deliberation. Deliberation is *decisional* because its aim is to form decisions; it is *rational* because it is subject to norms of rationality; and it is *provisional* because it is always run on a number of provisions. Second, this paper argues that the unified account helps to correct the set-up of the deliberation question. The deliberation question asks whether leeway sceptics or determinists can deliberate rationally. Deliberation incompatibilists say 'no' on the grounds that we cannot deliberate rationally about a determined outcome; deliberation compatibilists say 'yes' on the grounds that deliberation merely requires a belief in its efficacy and that deterministic deliberation is efficacious. But each side is mistaken in asking for a *belief* (as opposed to an assumption), since deliberation is provisional. And each side is mistaken in targeting the determinist instead of the leeway sceptic, since one can rationally believe in a deterministic open future. Third, this paper argues for a (corrected) deliberation incompatibilism. Roughly, since deliberation is decisional and that leeway sceptics cannot decide rationally to Φ unless they already know they will Φ, they cannot typically deliberate rationally about what to do.

1. Introduction

Suppose that Cate deliberates about whether to award the *Palme d'Or* to *The Wild Pear Tree* or to *Shoplifters*. The first contains exceptional shots, but the acting of the second is superior. She finds it hard to choose. Now imagine Cate comes to believe that one of the films is ineligible (perhaps not knowing which) or that she is not on the jury. In a sense, she can no longer deliberate rationally.

Suppose instead that Cate comes to believe that her choice is determined by the past and the laws of nature. Or suppose that she becomes convinced that she cannot choose amongst real alternatives, or, as philosophers, say, that she lacks leeway. Can Cate still deliberate rationally? More generally, can determinists or leeway sceptics deliberate rationally about what to do? Call this the *deliberation question*. The deliberation question is important because there are strong reasons for being a determinist or a leeway sceptic, as the free will debate has long demonstrated.[1]

The deliberation question raises controversy. On one side of the aisle, deliberation incompatibilists say 'no' (Alexander of Aphrodisias, *De Fato*; MacIntyre 1957; Taylor 1966, Castañeda 1975; van Inwagen 1983; Searle 2007; Cohen 2018). From cases like that of Cate, they infer that we need to believe in an open future to conduct rational deliberation. On the other side of the aisle, deliberation compatibilists say 'yes' (Waller 1985; Kapitan 1986; Pettit 1989; Clarke 1992; Nelkin 2004, 2011; Pereboom 2008a, 2014a). They admit that Cate cannot deliberate rationally in the first cases, but that believing in determinism or in leeway scepticism is not an impediment to rational deliberation.

[1] 'Leeway scepticism' is an unusual term, but it potentially includes many people. It first includes free will sceptics who think either that free will *is* leeway (hard determinists like Holbach and Spinoza) or that it *implies* leeway (Pereboom 2014a, 2014c). It also includes free will believers who are happy to concede that we lack leeway because free will is completely unrelated to leeway (Dennett 1984, Fischer 1994 and semi-compatibilists in general). Levy (2011) gets close to leeway scepticism when he argues that we lack the control required for responsibility and that abilities are relevant to this control.

In order to make progress on the deliberation question and similar ones, we need to be clear on what practical deliberation consists in. This is why my first goal in this paper is constructive: in §2, I offer a unified account of practical deliberation. It is *rational* in that it is subject to rational requirements. It is *decisional* in that it aims at issuing decisions. It is finally *provisional* in that it is always run on a number of provisions. These three properties interact with one another.

My second goal is corrective: in §3, I use my account to readjust the set-up of the deliberation question. Both deliberation compatibilists and incompatibilists are partly mistaken in the general way that they specify requirements on deliberation. Since deliberation is provisional, each side is mistaken in asking for a *belief* rather than for an assumption. And since deliberation is rational (and that compatibilism is rational), each side is mistaken in targeting the determinist rather than the leeway sceptic.

My third goal is polemical: in §4, I side with the deliberation incompatibilist. Aided by my account in §2 and the correction in §3, I argue that leeway sceptics cannot deliberate rationally. In a nutshell, it is impossible to decide rationally to do the impossible or the improbably possible. But this suggests that the leeway sceptic can never decide rationally to Φ, unless she already believes she will Φ. The sceptic can no longer deliberate but merely introspect or theorise. In §5, I respond to four objections.

Before I start, I wish to insist that the three main parts of my paper are tightly connected, but that each remains to some extent independent.

Deliberation compatibilists can adopt my account of deliberation and adopt my way of asking the deliberation question even if, in the end, my polemical argument falls short of convincing them. To them, I ask that they at least consider the deliberation question afresh.

2. Construction: A Unified Account of Practical Deliberation

Practical deliberation is the activity of deciding what to do on the basis of reasons. I explore in turn three properties of deliberation: it is decisional, rational, and provisional.

2.1 A Decisional Activity

Practical deliberation is *decisional*: its aim is to make a decision (see Taylor 1966, 168). Other philosophers agree in different words. They claim that deliberation is a form of reasoning or active thought whose aim is to form 'an intention to do something' (Clarke 1992, 107), or to make 'a commitment to a course of action' (Watson 2003, 175-176). Practical deliberation is thus a conative activity whose aim is a decision rather than a cognitive activity whose aim is a belief.[2] That is not to deny that practical deliberation includes the identification or at least the weighing of reasons.[3] But once we have figured out the reasons supporting each course of action (having perhaps failed to identify the best one), we still need to make a decision. This is especially clear

[2] It is contentious whether the *conclusion* of practical reasoning is action (Aristotle, *On The Soul*, 433a13–20), intention (Broome 2013), or belief (Raz 1999). This question is orthogonal to my discussion.

[3] See Clarke (1992, 107–8), Watson (2003, 176), Pereboom (2008a, 291). Coffman and Warfield (2005, 28) exclude the weighing of reasons from deliberation to isolate its conative part from its cognitive part.

in cases of irresolvable dilemmas, where we have no commensurable reason to favour one option over another (see Sartre 1946). Theoretical deliberation might be over, but practical deliberation will not end before a decision is made.

The deliberation question concerns only practical deliberation. But it is useful to note that there are other kinds of deliberation—or deliberation about other questions—which are available to the sceptic. There are fundamentally three questions concerning one's future conduct one can deliberate about:

1. Shall I Φ or Ψ?
2. Should I Φ or Ψ?
3. Will I Φ or Ψ?

We can distinguish different *kinds of deliberative activities* which correspond to some of these questions. The first question concerns primarily *practical* deliberation—its answer is a decision. The second and the third concern primarily *theoretical* deliberation—their answer is a belief. But these activities are linked: we typically ask the second to answer the first. Yet the fact that practical deliberation involves theoretical reasoning does not make practical reason a kind of theoretical reasoning (see Watson 2003, 176).

2.2 A Rational Activity

Like theoretical deliberation, practical deliberation is a *rational* activity: it is subject to norms of rationality or constraints. The exact specification of these constraints is controversial (as we shall see in §3), but individual cases are usually clear. We could not, for instance, deliberate rationally about whether to spend the weekend on the Moon or on Mars. And Cate could not deliberate

about whether to award a prize if she firmly believed that she was not a member of the jury.

It is interesting to ask what kind of irrationality is at stake in these clear cases. Instrumental rationality must be partly concerned: it is *pointless* to deliberate about whether to go to the Moon or to Mars. But it seems that there is something more going on—at least we should keep the possibility open at this stage. Even if deliberation can have a positive effect on my life or on my conduct, this can simply be in virtue of the fact that being irrational is sometimes beneficial. Deliberating about whether to accept a job that I was not offered might put me in a good mood or even make me get the job, in the Kavka-like case (1983, 33–34) where an eccentric Dean tells me that if I deliberate about whether to accept a tenure-track job I was not offered, he will offer it to me. And yet some irrationality remains. It seems then that, just like there seems to be such a thing as effective but irrational beliefs, there is such a thing as effective but irrational deliberation.

One way to capture the irrationality I have in mind in is to emphasise the difference between the irrationality *to* deliberate and the irrationality *in* deliberating. The first might simply be instrumental: it is rational *to* deliberate when it is a good means-to-end activity. The second form of irrationality, irrationality *in deliberating*, is different. Deliberating well does not mean that it is good to deliberate and vice versa. (Compare it being rational to play Monopoly and playing Monopoly rationally.)

What is irrational deliberation in this second sense? It is sometimes understood with the help of the distinction between right and wrong kinds of reasons, or of reasons and motives.[4] Here, I think that we have two choices. We may first give it an epistemic reading. Suppose that prior to deliberating about whether to go to Mars or to the Moon to spend the weekend, I believe it is impossible for me to travel in a spaceship in the foreseeable future. To deliberate, it seems likely that I will form some beliefs, such as the belief that I can in fact get on a spaceship by Friday. Now, this can lead me to hold contradictory beliefs (see van Inwagen 1983, 155–160). Alternatively, I might abandon, but for no reason, my belief that I cannot travel in a spaceship in the foreseeable future. In each case, I am clearly epistemically irrational.

Second, we can give irrational deliberation a constitutive reading. Without forming any new beliefs, I might make such an unreasonable *assumption* that I am guilty of constitutive irrationality. Here, I somehow fail to follow the constitutive rules of deliberation. I might for instance have shifted the nature of the activity—making deliberation a case of daydreaming or of theoretical deliberation. When Cate deliberates with the belief that she is not on the jury, she might shift to a different activity, i.e. she deliberates, like many of us in our living rooms, about who *should* receive the prize.

There are others questions we can leave open. The first is whether very irrational deliberation falls short of being deliberation. If rational deliberation

[4] See e.g. Kelly (2003), Hieronymi (2005), Schroeder (2007), and Nolfi (2018).

requires *X*, it might be because deliberation, not rationality, requires *X*.[5] The second is whether the irrationality I am concerned with is the irrationality *in engaging* in deliberation or *in conducting* deliberation. Like irrationality for maintaining or for forming a belief, these two kinds of irrationality seem similarly problematic.

Interestingly enough, decision-making also has some rational constraints. I cannot decide to go to the Moon rationally any more than I can deliberate about whether to spend the weekend on the Moon or on Mars. Now, since deliberation is a decisional activity, it is very plausible to hold that constraints on decision-making are somewhat exported to deliberation. If I am considering two options in deliberating and that it is clear to me that I could not rationally decide to do either, then my deliberation is also irrational. The decisional nature of deliberation interacts with its rational nature.

2.3 A Provisional Activity

I claimed that practical deliberation aims at making a decision, yet it does not always aim at making a decision *simpliciter*. When, after much deliberation, I decide to Φ, this is typically—if not always—*provided that* things go well. I decide to take Oxford Street rather than Divinity Street to go to work. My decision depends on the provision that Oxford Street is accessible. No one should call my decision 'dishonest' if I end up changing plans because I see that Oxford Street is blocked.

---

[5] See Coffman and Warfield who formulate their view without the term 'rationality' but end up discussing rational deliberation (2005, 43–44).

Most of our decisions are qualified with a series of implicit provisions, something barely acknowledged in the contemporary literature.[6] Stoics like Cleanthes and Chrysippus were aware of this possibility, but it left them unsatisfied. Thus, they located human agency in a realm that humans presumably had full control over; walking, they say, really happens in the soul.[7] It is only about this sort of mental action (*attempts*) that we might take decision to be unqualified by provisions. But even this is to assume, problematically, that we fully control these mental actions. When I decide to *attempt to write* a paper, I can fail if I fall asleep just before, just as I can fail to *write* a paper if electricity is cut off. We do better in maintaining that decisions and are always provisional, even when the object of the decision is an attempt. In a slogan, our decisions are always more or less 'decisions *insha'Allah*' (if God wills).

The provisional character of decisions is exported to deliberation, making it a provisional activity. We run our deliberation on a number of provisions and when it succeeds, it produces a provisional decision. Good deliberation is not impeded by the fact that our decision is provisional. The provisional nature of deliberation interacts with its decisional nature

However, there are constraints on what counts as a mere provision for a decision. I can decide to take Oxford Street (provided implicitly that it is accessible). But I cannot decide to buy a yacht (provided implicitly that I win the lottery). In the second case, the provision changes the nature of the

[6] Castañeda (1975, 134–135) might have been aware of this fact since he places constraints on deliberation which are related to what we can *assume*.

[7] See *Moral Letters to Lucilius* (113, 23).

decision: I no longer decided to buy a yacht; I simply decided how to spend money I don't have. The decisional nature of deliberation interacts with its provisional nature.

What explains the difference between the case of the yacht and of the street? It is the fact that in the first case, but not in the second, it is irrational to assume that the provision is a fact. It is not just '*insha'Allah'* that I will buy a yacht; it is if I am extremely lucky. I can assume for the purpose of practical deliberation that Oxford Street is accessible, but not that I will win the lottery. It follows that either I am irrational in deciding to buy a yacht (*simpliciter*) or that I am making a mere conditional decision, one which should be described as *daydreaming*. Daydreaming or conditional decision is not the aim of practical deliberation. And so deliberating about whether to buy a yacht is not real practical deliberation. The rational nature of deliberation interacts with its provisional nature.

To summarise, practical deliberation is an activity which shares three main properties. First, it is decisional: practical deliberation aims at making a decision rather than forming a belief. Second, it is provisional: we deliberate *insha'Allah*. Third, deliberation is rational: it is an activity which is subject to normative constraints of rationality. We have also seen that the three properties interact in several ways. Whether a provision changes the nature of a decision depends on what it is rational to assume. And whether it is rational to deliberate in a certain situation depends on whether it is rational to make certain decisions or to make certain provisions. These interactions unify the account.

3. Correction: The Deliberation Question

With this unified account of practical deliberation, we can revisit the deliberation question. In this section, I present a quick case for its two answers. I then offer some corrections to the set-up of the question based on the rational and provisional nature of deliberation.

3.1 The Standard Case for Deliberation Incompatibilism

Take again Cate, who deliberates about who to award the *Palme d'Or* to. If she believes she is not on the jury, it is clear that she cannot rationally deliberate about whom to award the *Palme d'or* to, although she can deliberate about who *should* receive the award or whom she *would* give it to in different circumstances. After all, these are questions that we deliberate about from the comfort of our homes. The same goes in all kinds of situations: if Cate believes that *The Wild Pear Tree* is not eligible for the prize; if she believes that she is incapable of writing down her vote; if she believes that even were she to decide wholeheartedly to give the *Palme d'or* to *The Wild Pear Tree*, she would compulsively give it to *Shoplifters.*

Why can't Cate rationally deliberate about which movie to vote for in each of these cases? An immediately attractive explanation is the following principle:

> **Specific Disbelief:**[8] If *S* deliberates rationally about whether to Φ or Ψ, then *S* does not believe that Φ is unavailable for her to

[8] 'Disbelief' here (as in the OED) means absence of belief, not belief of a negation.

> perform and *S* does not believe that Ψ is unavailable for her to perform.[9]

Note that in *Specific Disbelief*, as in any other principle that I will consider, Φ and Ψ are incompatible actions or omissions. (Why philosophers always focus on action is a mystery to me; for the most usual case is when we deliberate about whether to Φ or to refrain from Φing.) An action or an omission counts as 'available for *S* to perform' when *S* has both the ability and the opportunity to perform it. Consider again the four cases featuring Cate. If Cate believes she is not a member of the jury or that *The Wild Pear Tree* is ineligible, then Cate believes (at least she should) that she lacks the *opportunity* to cast her vote for it. If Cate believes that she will compulsively vote for *Shoplifter* or if she believes that she is illiterate (assuming the vote must be written down), she believes that she somehow lacks the *ability* to cast her vote for it. Specific Disbelief gives us the desired verdict that, in all the cases, Cate does not deliberate rationally.

Now, *Specific Disbelief* fits well with deliberation compatibilism. For it is inconsequential for leeway sceptics who don't know which path is unavailable when they initiate deliberation.[10] It will only be revealed to them once they have decided. The deliberation incompatibilist needs other cases to support her favourite principle, the Belief in Ability Thesis (BAT):[11]

---

[9] See Pettit (1989) and Nielsen (2011, 410).

[10] *Specific Disbelief* would pose problem to leeway sceptics if it is possible to rationally about what we know we will do. See Cowan (1969, 59) and Clarke (1992, 108).

[11] See e.g. Taylor (1966, 170–84), van Inwagen (1983, 160), Coffman and Warfield (2005), and Henden (2010).

> **BAT:** If *S* deliberates rationally about whether to Φ or Ψ, then *S* believes that both Φ and Ψ are available for her to perform.

Note that BAT is a *doxastic requirement on deliberation*. If BAT were true, the leeway sceptics could not rationally deliberate. In deliberating, they would be irrational either by forming contradictory beliefs, by abandoning some beliefs for no reason or by making a deliberative mistake.

We have a strong reason to favour BAT over *Specific Disbelief*. The first but not the second can explain the irrationality of the agent in some cases, which Taylor (1966, 177–78) and van Inwagen (1983, 153) devised. Take the following:

> **Cinema:** Sandrine wants to go to the cinema. She deliberates about whether to see *Isle of Dogs* or *The Square*. She then learns that one of the two is no longer showing, but she does not know which one. She keeps deliberating as if she had not learnt anything.

Yet Sandrine cannot deliberate, or at least she cannot deliberate *rationally*, about which movie to watch. This is an intuition that almost all participants to the debate have registered about structurally identical cases.[12]

It is worth recognising, in accordance with our account of deliberation, that Sandrine can deliberate about other questions. She can deliberate about what she shall or should *try* first. Secondly, she can deliberate about what she

[12] The classic case features an agent who deliberates about which of two doors to leave the room through. Since he knows that one door is locked not knowing which, this is either impossible or irrational. Taylor (1966, 177–78), van Inwagen (1983, 154; 2017, 184), Waller (1985, 49), Kapitan (1986, 247), Clarke (1992, 104, and note 8), Nelkin (2004, 219), and Pereboom (2008a, 297) all agree about this judgement. One exception is Nielsen (2011, 410), who thinks that deliberation is always theoretical.

thinks she will *end up trying first*. She might be painfully aware of how much she tends to follow the film critiques out of snobbery and thus suspect that she will pick *The Square*. Thirdly, Sandrine can deliberate about what she *would* or *should* pick, *were both options available*. This is similar to daydreaming about what we would do if things were different. Fourthly, Sandrine can deliberate about what movie is showing and so what is her true available option. So BAT, but not *Specific Disbelief*, gives us the right verdict about a clear case. Sandrine does not believe that her options are available and thus fails to deliberate rationally. Thus, the case for BAT (which implies deliberation incompatibilism) is that it is an explanatory principle for scenarios like Cinema.

3.2 The Standard Case for Deliberation Compatibilism

Deliberation compatibilists grant that Sandrine is irrational in deliberating. But they think that they can dispose of BAT when explaining this irrationality. Deliberation compatibilists now recognise that the following kind of principle will not work (see Nelkin 2011, 130; Pereboom and McKenna 2016, 297–298). It is tempting to read BAT as a principle about *epistemic availability* or openness.[13] For *S* to deliberate about whether to Φ or Ψ, *S* must believe that each is available to her in the sense that they are compatible with what she knows or believes. So, since Φ and Ψ are presumably exclusive, sceptical deliberators must not have yet predicted or be certain about whether they will Φ or to Ψ. Since sceptics—like any of us—do not know or are certain about what they will do prior to deliberating, this requirement sits well with

[13] See e.g. Dennett (2015, 118–26), Kapitan (1986, 1996), Nelkin (2004), and Pereboom (2008a).

deliberation compatibilism. (In reality, it is very similar to *Specific Disbelief*.) The epistemic requirement might be sound (cf. Cowan (1969) and Clarke (1992)) Sound or not, the epistemic openness requirement does not address the standard case for deliberation incompatibilism. In fact, it leaves undiagnosed Sandrine's irrationality in Cinema. Going to see *The Square* and going to see *Isle of Dogs* is each compatible with what she believes or knows. The case for BAT remains.

The real challenge for deliberation incompatibilism comes from a different requirement, that of deliberative efficacy. Sandrine's irrationality can be explained without invoking her lack of leeway. She knows that, in an important sense,[14] her deliberation is futile and yet she deliberates. There lies the irrationality; not in her belief that she lacks alternatives.

According to Kapitan (1996), Nelkin (2004, 2011), and Pereboom (2008a, 2014a), we can only rationally deliberate if we believe that our deliberation is causally efficacious. But, as they each acknowledge, we need to refine this requirement. Imagine that you are asked to deliberate about whether to buy a jet or not (I presume you are not rich enough). If you proceed, you will win a thousand dollars. Since it would be profitable, your deliberation would be causally efficacious. And yet your deliberation would be irrational in the sense

[14] It might still be important for accidental reasons that they deliberate, for instance to cultivate good character.

that we fixed in §2. Hence, to deliberate rationally, you must believe that your deliberation is causally efficacious *in the right way*.[15]

After reviewing several formulations which I cannot present here for reasons of space,[16] Pereboom endorses the following requirement:

> **Deliberative Efficacy:** In order to rationally deliberate about whether to do A_1 or A_2, where A_1 and A_2 are distinct actions, an agent must believe that if as a result of her deliberating about whether to do A_1 or A_2 she were to judge that it would be best to do A_1, then under normal conditions, she would also, on the basis of this deliberation, do A_1; and similarly for A_2. (Pereboom 2008a, 299)

This explains why Sandrine cannot deliberate rationally about whether to go see *Isle of Dogs* or *The Square*: she does not believe that were she to choose one, she would watch this film. We should perhaps replace 'having judged an option to be best' with 'having chosen an option'. This is to cover cases where we cannot judge that an option is best—perhaps our reasons are incommensurable—and cases where we decide against our judgement about what it would be best to do (think about the amoralist). In these cases, our deliberation seems efficacious in the right way: we decide to do and act on this basis.

So, we have a case for deliberation compatibilism and a case for deliberation incompatibilism. The case for deliberation incompatibilism is that

---

[15] This is clearly recognised by advocates of principles of deliberative efficacy like Clarke (1992, 103), Nelkin (2004, 223), and Pereboom (2008a, 298).

[16] See Kapitan (1986, 247, 1996, 437), Clarke (1992, 103), and Nelkin (2004; 2011).

Sandrine's irrationality is best explained by BAT. The case for deliberation compatibilism is that Sandrine's irrationality is best explained by Deliberative Efficacy. Each principle is a competing doxastic requirement on rational deliberation in that it explains what agents must believe in order to deliberate rationally.

3.3 A Correction From Provision

I now move to my second goal in this paper, i.e. correcting the set-up of the deliberation question and the main case for each side. My first correction is that both deliberation incompatibilists and compatibilists must amend their explanatory principles to respect the provisional character of deliberation. Take the following case of rational deliberation. James wonders whether to take Oxford Street or Divinity Street to go to work. He does not believe that any is blocked or available; he simply assumes—reasonably—that they are available for the purpose of practical deliberation, hoping for the best. BAT tells us that James should have a *belief* that either street is available, which he lacks. Deliberative Efficacy tells us that James should have a *belief* that his choice will be efficacious, which he also lacks. Thus, both BAT and Deliberative Efficacy give us the wrong verdict, i.e. that James' deliberation is irrational.

It is tempting to first claim that the provisional character of deliberation implies that we can only place constraints in terms of the agents' *assumptions* rather than on his *beliefs*. Thus, the deliberation compatibilist could, in the spirit of Deliberative Efficacy, ask that the agent *assumes* that her deliberation is

efficacious. And the deliberation incompatibilist could, in the spirit of BAT, ask that the agent *assumes* that she has leeway.

But deliberation is not only provisional but rational. And there are rational constraints on what an agent can assume in deliberating. I can deliberate about whether to buy a jet *if I win the lottery*; but this is certainly not the same as deliberating about whether to buy a jet, nor is it the same as deliberating about whether to buy a jet *provided that things go well* or *insha'Allah*. Deliberating provided unreasonable assumptions is just daydreaming or irrational deliberation.

Now, these rational constraints on what we can assume depend on what we know or believe. It is because I know that I cannot buy a jet without winning the lottery that I cannot deliberate rationally about whether to do it. If I had other pieces of knowledge or other beliefs, this would perhaps not be impossible. To capture these restraints in doxastic terms (like BAT and Deliberative Efficacy)), I suggest the following. The deliberation incompatibilist should endorse the Disbelief in Inability Thesis (DIT):

> **DIT:** If *S* deliberates rationally about whether to Φ or Ψ, then *S* does *not* believe that <either Φ or Ψ (not knowing which)> is *un*available for her to perform.[17]

Or, differently put: to deliberate rationally, one must lack the belief that one lacks alternatives. DIT is strictly weaker than BAT and somewhat stronger

[17] *S* does not infringe DIT automatically in virtue of believing (truly) that she can only do Φ or Ψ *since they are exclusive*. This would involve a *de re*/*de dicto* confusion like claiming that I could not do otherwise because I 'cannot do what I don't'.

than Specific Disbelief.[18] That *S* does not believe that she lacks leeway is very close to a requirement on rational assumptions. If Sandrine believes (or knows) that there is one of the two films that she cannot see (not knowing which), she cannot simply assume that it is showing.

As for the deliberation compatibilist, I suggest that she amends Deliberative Efficacy as follows:

> **Deliberative Non-Inefficacy:** If *S* deliberates rationally about whether to Φ or Ψ (where Φ and Ψ are incompatible actions or omissions), then *S* does *not* believe that, under normal circumstances, if she were to decide to Φ, she would *not* or that, if she were to decide to Ψ, she would *not*.

In order to deliberate rationally, an agent must *refrain* from believing that her deliberation is *in*efficacious. This satisfies the provisional character of deliberation while setting a requirement on what an agent can assume. She cannot assume that her deliberation will be efficacious if she believes that one of the options is blocked in the way that it is in Cinema.

3.3 A Correction From Rationality

We saw that practical deliberation was a rational activity in that it was constrained by norms of rationality. But the way in which deliberation might be irrational is not very extended. In particular, someone who has false or irrational beliefs can still conduct rational deliberation (perhaps even if some of these beliefs are involved in deliberation).

---

[18] To be strictly stronger than Specific Disbelief, it would need not to contain the 'now knowing which' clause.

I have started this paper by claiming that the deliberation question asks whether leeway sceptics *or determinists* can deliberate rationally. Then, I left determinism aside completely. For I think that while it is plausible to claim that leeway sceptics cannot deliberate rationally, it is implausible to say this of all determinists. (Remember: as I indicated in note 1, leeway sceptics include free will sceptics who believe that free will requires leeway and hard determinists. Friends of leeway scepticism include those who, like the majority of semi-compatibilists, are happy to deny the existence of leeway.) A vast array of philosophers[19] have, by contrast, presented BAT/DIT as a challenge for determinists alone. And they have explicitly defined deliberation incompatibilism as the view that a belief in determinism is incompatible with practical deliberation.

But this is as an incautious way of setting things up. Believing in determinism *is* compatible with rational deliberation. This is firstly because a subject can rationally believe both that determinism is true and that she can do otherwise. That's just what many compatibilists do.[20] If some determinists cannot rationally deliberate, it is only in virtue of a further belief, i.e. the belief that the truth of determinism precludes their ability to do otherwise. So, even if DIT and incompatibilism were both true, it would not follow that determinists could not deliberate. A rational but false belief in compatibilism

[19] See Taylor (1966, 181–84), Waller (1985, 48), Nelkin (2004, 215–18), Pereboom (2008a, 288–89), Henden (2010, 313), Nielsen (2011, 283–84), and Cohen (2018, 87).

[20] See, for instance, Vihvelin (2013) and Dorr (2016).

would allow them to satisfy the condition specified in DIT. They would disbelieve that they lack leeway.[21]

We can add, further, that our conception of irrational deliberation allows for the view that *even if it were irrational to be a compatibilist* (which seems plain false), it would not follow that the truth of DIT would imply that determinists cannot deliberate. First thing, it is unclear that compatibilism would itself play a role in the deliberation of agents. Take James. He might lack the belief that he lacks leeway regarding his way to get to work. Even if this lack of belief is partly explained by the fact that he believes in compatibilism, compatibilism does not feature explicitly in James' deliberation. Furthermore, our conception of irrational deliberation does not imply that every time an agent uses an irrational belief in deliberation, the deliberation thereby becomes irrational. A deliberation is irrational when the activity of deliberating *causes* or *leads to* some irrationality such as the formation of a belief in a contradiction, not when some of the involved beliefs *preceding* deliberation are.

At any rate, it is unproductive to set up the debate as one between those who believe that a belief in determinism precludes rational deliberation and those who do not. The implausible stipulations of those who set up the debate in this way are evidence of this. For instance, Pereboom stipulates (2008a, 288; 2014a, 106)—merely for the purpose of the account—that it is 'evident' that if determinism is true, leeway scepticism is true. And Cohen (2018, 87) stipulates that the relevant form of rationality implies having a true belief about the

[21] See Coffman and Warfield (2005, 29) for a similar point.

compatibility question between determinism and leeway. We can avoid these dubious assumptions by phrasing the deliberation question as follows: can leeway sceptics deliberate rationally about what to do?

4. Polemic: An Argument From Decision

Some might be tempted to say 'no' with the deliberation incompatibilists. For they might say, with E.J. Coffman and Ted Warfield (2005, 42), that the deliberation compatibilist has merely provided a rival explanatory principle for the irrationality of characters like Sandrine, but that they have not given a clear counterexample to DIT. After all, both DIT and *Deliberative Efficacy* could be true. As a matter of fact, several philosophers believe that we need more than one doxastic requirement on rational deliberation.[22]

However, we must not forget why we might have admitted DIT in the first place. It is an inference to the best explanation of Sandrine's irrationality. If a weaker principle than DIT can explain just as much, then we have little reason to accept DIT, unless we find an argument supporting it.[23] The absence of a convincing counterexample to DIT thus fails to establish DIT and to trouble deliberation compatibilists and leeway sceptics.

This is not to say that we should disregard how difficult it is to identify a counterexample to DIT. I believe that this difficulty indicates that an argument in favour of DIT is available, as I will now argue. Instead of trying to show that it is the best explanation of some other case of irrationality in deliberating,

[22] See e.g. Taylor (1966, 171), Nelkin (2004, 218), and Pereboom (2008a, 300–302).

[23] See Nelkin (2004, 220) for a similar remark.

I will give a direct argument in its favour, an argument that is based on the conception of deliberation presented in §2. Here it is:

> **(1)** If *S* deliberates rationally about whether to Φ or Ψ, then either *S* can rationally decide to Φ or *S* can rationally decide to Ψ.
>
> **(2)** If *S* can rationally decide to Φ or rationally decide to Ψ, then *S* does not believe that <either Φ or Ψ (not knowing which)> is unavailable for her to perform.
>
> **DIT:** If *S* deliberates rationally about whether to Φ or Ψ, then *S* does not believe that <either Φ or Ψ (not knowing which)> is unavailable for her to perform.

We can recast the argument specifically for the leeway sceptic. To deliberate rationally, the leeway sceptic must be able to decide rationally to perform one of her options (in accordance with premise 1). Yet the leeway sceptic cannot decide rationally to perform one of her options—unless she already believes what she will do (in accordance with premise 2). If she already believes what she will do and *then* decides to do it, she has perhaps not been irrational (this depends on whether *Specific Disbelief* is true), but she has failed to engage in practical deliberation. Instead, she has engaged in introspection or theoretical deliberation—she has asked the question 'what will I do', not 'what shall I do.'[24] Then, she has simply surrendered to the inevitable.

[24] Thus, Tony Honoré (1999, 159) writes: If we are sceptics, 'we should […] come to reinterpret the experiences we call "making up our minds"; and "reaching a decision". We should come to think of them more like "becoming aware of our intention".'

I shall now focus on the argument for DIT. (2) is more contentious than (1). I will defend each, although my aim is merely to give to DIT some general plausibility.

4.1 Aiming at a Decision

We (1) can simplify by the principle that to deliberate about two options, we must be in a position to rationally decide to go along with one of them. The general case for (1) rests on the view that practical deliberation is a decisional activity (see §2). If the aim of theoretical deliberation is truth or knowledge (that it is an alethic activity) and I am aware that *p* is false—that is that my aim is unachievable—then my believing *p* will be irrational. Similarly, if the aim of deliberation is making a decision and that I am aware that I cannot make a decision, then my deliberating is irrational. The same goes if I am aware that I cannot form a *rational* belief or a *rational* decision.

What if I am not in a position to rationally decide to Φ and to Ψ but that I am unaware of this? Some might suspect that in such cases, I can deliberate rationally about what to do, *contra* (1). We can imagine Sophie deliberating about which of her children to save and who, at some point in the deliberation, realises that she will literally be incapable of setting her mind or committing herself to a course of action. This does not make her previous deliberation irrational, because when she weighed reasons, she thought or reasonably assumed that she might be able to make a decision.[25]

[25] Thanks to [REDACTED] for this helpful example.

I do not want to push back against this suspicion. Rather I will point out that this sort of counterexample to (1) is inconsequential for our purpose. We could amend the argument to restrict its application to agents who are aware of, or should be aware of, their inability to decide. So, an agent who deliberates in the (acceptable) ignorance of her inability to rationally decide to Φ and Ψ might do it rationally. But it does not help our leeway sceptic very much. For one thing, if the leeway sceptic gets off the hook, it is because he is blissfully ignorant of some facts about rationality (expressed by (2)). Our sceptic is not, like Sophie, unable to decide to Φ and to decide to Ψ. He is, according to (2), unable to decide *rationally*. But ignorance of norms of rationality might always be irrational (see Littlejohn 2015). Alternatively, even if the blissfully ignorant leeway sceptic gets off the hook, this is not great news for leeway sceptics in general. For ignorance of the norms of rationality is independently a significant failure.

Let us leave aside these characteristically tricky issues about ignorance. My point is that in normal cases I can only deliberate rationally if my deliberation can in principle succeed, i.e. it can lead me to a rational decision. Let us move to the more controversial premise (2).

4.2 A Constraint on Decisions

I now turn to (2). To give it some plausibility, I start with a simpler principle:

> **Realistic Decision:** To rationally decide to Φ, one must refrain from believing that Φ is unavailable

Once we have established it, we can show that we should endorse (2) for similar reasons.

Right off the bat, let us note that *Realistic Decision* takes into account the provisional nature of deliberation. If, as I will argue, it is impossible for a subject to decide rationally to do what she believes she cannot do, it remains possible for a subject to decide to do what she *lacks* the belief that she *can* do as the case of James establishes.

We cannot decide to Φ if we believe Φ is unavailable—or to use a classic phrase, we cannot rationally 'intend to do the impossible' (see Hampshire 1960). This strikes me as intuitive enough in itself to warrant a defensive stance, where the proponent of *Realistic Decision* merely replies to objections (see §5.4). But I have something positive to say for it.

Imagine that I think, 'I shall swim this afternoon, though I cannot' or suppose that I assert, 'I hereby decide to swim this afternoon, although I am unable to.'[26] This thought and this assertion are irrational or funny—that is funny because irrational. They are similar to 'Moore-paradoxical' assertions or thoughts like, 'It's raining but I don't know it is.' While the expressed propositions are not contradictions, there is something wrong with their assertions, whether it is epistemically wrong or contravenes some other norms of assertion. There is also perhaps something wrong with forming a Moore-paradoxical belief. A similar case is that of promises, where 'I promise to Φ, although I can't' is also infelicitous.

[26] This does not work for the sense of 'can' which is sensitive to obligations (see Hyman 2015, chap. 4). There is nothing paradoxical about: 'I shall Φ but I cannot, i.e. it is prohibited.'

Why are these thoughts or assertions infelicitous? Take Moore-Paradoxical assertions. The propositions that they express cannot be known; but a good belief is a piece of knowledge; and a bad belief should not be asserted (see Williamson 2000). Of course, the explanation is much more controversial than the judgement that Moore-Paradoxical assertions are infelicitous. The same goes with decisions. The explanation of the above infelicity, to which I now turn, gives plausibility to the judgement even though it is controversial.

Let us try to imagine someone who fails to make decisions in her life. This is impossible, unless we imagine someone who is permanently unconscious or who lacks a human mental life. But we can succeed to imagine someone whose decisions play a relatively minor role in her life. This person never decides anything *in advance*. She lets the events unfold and she allows her immediate desires to guide her on the spot. She never mentally commits to a course of action after deliberation—she just does things (intentionally). In a word, she never plans anything.

Similarly, we can imagine someone who fails to make promises in her life. This is possible, at least if we only consider the case of explicit promises. Our fellow is not necessarily a bad person, but she must be lonely. For she cannot coordinate with other people. She can communicate her desire to see her friend tomorrow at 10:00, but unless there is a promise to meet, this kind of desire is doomed to often remain frustrated. In a word, our fellow never plans anything in coordination with others.

So, *planning* is a key role of decisions and promises (motivation being another one). This explains why 'I shall Φ although I can't' and 'I promise to Φ, although I am unable to' are infelicitous and why the corresponding decisions and promises are irrational. The plan that they recommend is imaginary. A promise or a decision to Φ in these circumstances should not make us plan other things differently. If my friend claims to be unable to meet me at Bethnal Green at 20:00, his promise to be there should not influence my plans. We could in general accept:

> **Walk the Talk:** If *S* can rationally decide or promise to Φ, she does not believe that she is unable to Φ.

This principle implies *Realistic Decision*, which we can endorse on the grounds of the infelicity of certain claims and of the fact that, just like promises, the primary role of decision is planning.

If we accept *Realistic Decision*, we should also accept:

> **(2)** If *S* can rationally decide to Φ or rationally decide to Ψ, then *S* does not believe that <either Φ or Ψ (not knowing which)> is unavailable for her to perform.

Why? Consider similar thoughts to, 'I decide to Φ although I can't Φ.' Take, first, 'I decide to Φ, although it is very unlikely that I can't Φ' and, 'for all I know I can't Φ, but I decide to Φ.' These thoughts seem almost as unreasonable as the first. But if my thought is, 'I decide to Φ although there is a slim chance that I cannot', this does not strike us as infelicitous. Now, consider, 'I either cannot Φ or Ψ, I don't know which, but I decide to Φ'. Again, this seems infelicitous.

Now, I concede that it is difficult to draw the cut-off point: is *suspecting* that Φ is unavailable sufficient for making it irrational to decide to Φ? Or imagine that I have 100 options and that I know that one is unavailable not knowing which—surely, in such a case I *can* rationally decide to Φ. Be it as it may, we do not need for our purpose to draw the cut-off point. But a leeway sceptic who considered more than two options would be in deeper troubles. Deliberating about five exclusive options, four of which are unavailable, seems more clearly irrational than deliberating about two options, one of which is unavailable.

Again, we can provide a justification for these additional judgements of infelicity. We make decisions, or promises, to coordinate or plan. But a plan involving impossible course of actions, or probably impossible courses of action is hardly a good plan. Now the idea is not simply to dispel uncertainty about whether I will Φ when I promise or decide to Φ. For I can rationally decide or promise to Φ even though I fear I will not, *so long as I believe that I can.* I can decide to work even though I think I am lazy (this might motivate me). But if I think I am very likely to have a heart attack in the next hours, I cannot decide to work *simpliciter*.

Let me now bring the premises together again. Rational deliberation requires the possibility to rationally decide in favour of at least one option. The possibility to rationally decide in favour of at least one option requires that the deliberator not be in the following situations. The deliberator ought not to think that her options are unavailable, or that they are probably

unavailable, or that one of two is unavailable not knowing which. But then it follows that DIT is true.

Take Sandrine again. She deliberates about whether to go see *Isle of Dogs* or *The Square*. She believes that one of these options is closed to her, but she does not know which. Sandrine cannot rationally decide to go see *Isle of Dogs*, since she believes that either *Isle of Dogs* or *The Square* is not showing, not knowing which. The same goes for *The Square*. We can gather evidence for this from our judgement that 'I shall go see *The Square*' and 'I promise to go see *The Square*' are infelicitous. Our judgements are vindicated by the fact that this decision and this promise fail to contribute to good planning.

Let me reformulate this for the sceptic. Skye deliberates about whether to go for a run or for a swim. Being a leeway sceptic, she believes that one of these options is closed to her, but she does not know which.[27] Skye cannot rationally decide to go running, so long as she believes that either running or swimming is unavailable, not knowing which. The same applies for swimming. And so until she learns what she will do, she cannot rationally deliberate about what to do.

Now, Skye's case is different from Sandrine's case. For it seems that Skye could rationally decide or promise to go see *The Square* and plan accordingly. But why? Simply because Skye can predict that she will in fact go. From this prediction, she believes, in accordance with her scepticism, that she can. So, if Skye's decision or promise is rational, it is only because she has changed her

[27] She might sometimes not have this opinion because she is not a universal deducer. But she will often have this opinion if she is a consistent and 'active' leeway sceptic.

beliefs about what she *will* do. By contrast, let us suppose that Skye has *not* changed her beliefs about what she will do—she cannot predict what she will do and thus what she can do (since she is a leeway sceptic). The possibility that she cannot run remains too vivid to be think, assert, or promise 'I shall run'. In such circumstances, she cannot decide rationally, and thus deliberate rationally.

Consider the timing of Skye's deliberation. If she deliberates practically, then at *t1*, she decides to Φ; at *t2*, she *might* form the belief that she will Φ; and at *t3* that she can Φ. Skye satisfies DIT at *t2* (or at *t3*). But between *t1* and *t2* there is a gap where she has decided to do something which might well be unavailable. Skye is like a sprinter who can never respect the starting signal. That's not to say that she cannot or should not run. But there is something she does wrong.

This gap is much wider when the deliberator is not confident about her decisions. Imagine that Skye decides (or promises) to Φ next weekend but remains convinced until Friday that she will not carry through. For non-sceptics, this decision is rational. But for Skye, this means that she spends several days having planned to Φ despite believing or strongly suspecting that Φ is unavailable for her to perform. Even if we ignored the short moments of irrationality involved in sceptical deliberation, one thing would remain: rational sceptical deliberation could only end with a confident decision.

This is problematic since rational decisions need not be confident. The fact that the sceptic relies so heavily on her belief that she will act in

accordance with her decision suggests a more general problem, i.e. that she confuses theoretical and practical deliberation. Answering the questions, 'what shall I do?' and, 'what will I do?' are two separate things. For one thing, it is perfectly intelligible and rational to decide to do something in spite of the fact that we suspect that we will fail to do what we decided to do. But the sceptic cannot do this. She fails to see that good decision does not imply prediction. And for another thing, it is perfectly intelligible and rational to predict that I will do something without deciding to do it. Perhaps I find it regretful that I will act in this way, or perhaps I am not ready to decide to do it because I have yet to identify the right reasons to act.[28] But again, since the leeway sceptic believes that she has only one path, there is little point in trying to resist something she has predicted she will do.

5. Four Objections

5.1 The Efficacy Objection

The deliberation compatibilist might have accepted my account of deliberation and my set-up for the deliberation question. But she might insist that she has the upper hand because even though what we can rationally deliberate about depends on what we can rationally decide to do, we can rationally decide to do something just if we believe this decision to be efficacious. So the point is that the leeway sceptic is not irrational in deciding to Φ when she does not know whether she will Φ or Ψ and thus whether she can Φ. But how should we specify this belief in efficacy? Perhaps the deliberation compatibilist will say:

[28] See Waller (1985, 50), Cowan (1969, 59), Clarke (1992, 108), and Pereboom (2008a, 293).

> If *S* decides rationally to Φ, then *S* believes that, under normal circumstances, if she were to decide to Φ she would.

This proposal raises an immediate difficulty. It is not irrational for us to decide to do something even though we lack the belief that we will execute our decision. We can decide to work even though we suspect that we will fail, provided that we believe that we can work. Perhaps we must *lack* the belief that we will fail:

> If *S* decides rationally to Φ, then *S* does not believe that, under normal circumstances, if she were to decide to Φ she would not.

This is contentious. Is it clearly irrational for me to decide to work long hours if I believe that I will fail, provided that I believe that I have the ability to work late? I doubt it. Perhaps to rationally deliberate I need to believe that there is a particular set of possible worlds in which I have decided to work and where I do. But this raises the difficult question of which set; and it seems that the answer will be given by what I am *able* to do. This indicates that a principle about abilities is better suited to cover these cases. I cannot deliberate rationally about whether to work if I believe that I cannot work, or decide to work, or act on my decision.

A second difficulty with this proposal relates to assertions and thoughts like 'I shall Φ though I cannot.' Either the efficacy condition explains the infelicity, or it does not. If it does not, it is fair for the deliberation incompatibilist to claim that DIT is indispensable. If it does explain the infelicity, then this seems to be in virtue of the fact that the conditional claim is the right analysis of 'cannot.' But then, this means that classical compatibilism is true, and so leeway scepticism is false. I now explore this sort of problem, i.e. the risk for the leeway sceptic and deliberation compatibilist to give too much faith in what looks like a compatibilist analysis of 'can'.

5.2 The Equivocation Objection

The deliberation compatibilist might accept that everything I have said so far is correct, but only *insofar as there is just one notion of leeway*. That is, she might argue that there are *two* notions of having the ability to do otherwise, the first whose existence the leeway sceptic denies, but which is irrelevant to deliberation but relevant to freedom; the second whose existence the leeway sceptic accepts, but which is relevant to deliberation.[29]

The claim that there is a compatibilist and that there is an incompatibilist conception of the ability to do otherwise, which are not coextensive, is common.[30] The compatibilist conception is roughly the ability to do what we choose. And the incompatibilist conception is something like the ability to do different things, holding fixed the laws and the past. According to the Equivocation Objection, the compatibilist ability to do otherwise is necessary for deliberation but irrelevant to—or at least very insufficient for—freedom, whereas the incompatibilist ability to do otherwise is perhaps sufficient for freedom, but certainly irrelevant to deliberation.[31]

I believe that The Equivocation Objection fails for three reasons. The first is that choices or decisions are crucial for freedom. They are the locus of our freedom, at least if we are free. But to deliberate practically is the activity

---

[29] On a first glance, this seems to be Nelkin's suggestion. On multiple occasions (2004, 225, 228, 233), she points out that the kind of alternatives relevant to rational deliberation might differ from the kind of alternatives relevant to freedom. Yet, Nelkin's concluding remarks suggest that our view on deliberation and on freedom should be linked.

[30] See Dennett (2015), who agrees, and Kenny (1978), who does not.

[31] This is the sort of distinction that is found in Dennett's work (2015) and which van Inwagen (2017, chaps 10, 13) considers to be an important mistake.

which we engage in in order to make choices or decisions—it is a decisional activity. So, it is quite natural to think that the kind of ability which is relevant to deliberation is precisely the kind of ability which is relevant to freedom. Therefore, it seems *ad hoc* if not abstruse to maintain that the incompatibilist ability is necessary for freedom but unnecessary for deliberation.

The second reason why the objection fails is that we cannot decide to do what we believe we are *unfree* to do. Belief in unfreedom and belief in inability restrict our deliberation in the very same way. And so, if we believe that freedom requires the incompatibilist conception of the ability to do otherwise and that we lack this freedom, we cannot deliberate rationally. This explains why assertions or thoughts like, 'I shall swim, although I am not free to' are similarly paradoxical to, 'I shall to swim this afternoon, although I cannot.' Throughout the paper, the reader could replace 'able to do otherwise' with 'free to do otherwise' and get equally plausible arguments.[32]

The third reason why the Equivocation Objection fails is that if it is correct, it undermines leeway scepticism. It is crucial that leeway sceptics and incompatibilists make plausible their view that leeway or freedom is the ability to add something to the given past, holding the laws of nature fixed.[33] One of the only ways they have found to do this is by examples of rational

[32] But see note 26.

[33] See Ginet (1990, 102–3) and Fischer (1994, 88).

deliberation. Whenever we deliberate rationally, we hold the laws and past fixed.[34] For instance, John Fischer asks us to consider the following example:

> **Icy Patch:** Sam saw a boy slip and fall on an icy patch on Sam's sidewalk on Monday. The boy was seriously injured, and this disturbed Sam deeply. On Tuesday, Sam must decide whether to go ice-skating. Suppose that Sam's character is such that if he were to decide to go ice-skating at noon on Tuesday, then the boy would not have slipped and hurt himself on Monday. The situation is puzzling. It seems that Sam is able to decide to go […] ice-skating on Tuesday. And it also appears plausible that if he were to decide to go skating on Tuesday, the terrible accident would not have occurred on Monday. So it appears that Sam *ought* to decide to go ice-skating on Tuesday. And yet, given that Sam knows that the accident did in fact take place on Monday, it also seems *irrational* for Sam to decide to go ice-skating on Tuesday on the basis of a reason flowing from the truth of the backtracker. […] To do so would seem to exemplify something akin to wishful thinking. (Fischer 1994, 95)

Fischer's idea is that when we deliberate rationally about what to do, we must hold the past fixed on pain of irrationality; so we use the incompatibilist notion of ability when we deliberate.[35] But this is precisely to reject the Equivocation Objection. Thus, either Fischer's idea is correct and the objection fails because the incompatibilist ability is relevant to deliberation; or Fischer's idea is incorrect and the Equivocation Objection is unmotivated

---

[34] See van Inwagen (1983), Fischer (1994, 94–98), and Fischer and Pendergraft (2013, 586–90).

[35] Differently put, the counterfactuals relevant to the ability to do otherwise are those which hold the past and laws fixed. And we can see this when we consider cases of practical deliberation with backtrackers.

because the incompatibilist conception of the ability to do otherwise is not sufficiently motivated.[36]

5.3 The Alternative Deliberation Objection

In a similar spirit, the leeway sceptic might argue that there are other forms of deliberation which I have not shown she cannot engage in and which can replace practical deliberation. Although the leeway sceptic is free to come up with new accounts of deliberation, the prospects are gloomy.

First, the leeway sceptic cannot replace deliberation about what to do with deliberation about what to *try to do*. This is because she believes not only that no one can ever *do* otherwise, but that no one can ever *try* to do otherwise (unless one has in fact tried). Non-sceptics typically believe that there is an asymmetry between *trying to act*, which we can almost always do, and *acting*, which we often cannot do. (It is this asymmetry that I argued the Stoics overplay.) For the leeway sceptic, there is no such asymmetry. It follows that my main argument applies to deliberation about trying (replace 'Φ' with 'try to Ψ' in §4).

Second, the leeway sceptic cannot replace practical deliberation with conditional deliberation. She can engage in the latter, since it does not have strict rational requirements. She can, if she wants, assume that she can do otherwise just like I can assume that I will win the lottery. But as I have argued before, there is a good reason why we think practical deliberation is not just

[36] Fischer's strategy might fail for independent reasons as Vihvelin (2013, chap. 7) and Dorr (2016) argue.

like any conditional deliberation (i.e. it is not daydreaming): it issues decisions which are themselves pillars in planning.

Third, the sceptic cannot replace practical deliberation with theoretical deliberation. She can deliberate about what it would be best to do (what should I do?) or about what she predicts she will do (what will I do?). But this is insufficient because theoretical judgement often does not lead to action. In Dworkin's words: 'You cannot lift yourself above yourself just to watch how you choose. You must choose.' (2011, 223)[37] This makes sense of Sartre's idea that we are condemned to freedom. Something like practical deliberation is inescapable. For there will always be cases where the leeway sceptic cannot predict what she will do; where introspection seems pointless; and where she has to favour a path although she believes it is very likely to be unavailable.

5.4 The Rational Irrationality Objection

A fourth and last objection is that there are independent counterexamples to DIT such as the following.

> If, like the rock climber Aron Ralston, I find that my hand is stuck under a boulder in the Utah Canyons, and I determine, after days of waiting for help, that my only means of survival is cutting off the hand with a pocket knife, I will probably doubt whether I have the courage and resilience to execute the plan. But given what will happen if I don't, it would be irrational of me not to entertain this possibility, despite my doubts about my ability to follow through. (Nielsen 2011, 413–14)

[37] For a rare opposition to this claim, see Blackmore (2013).

This is a good counterexample to BAT: deliberation does not strictly require a belief that our options are available.[38] For it to be a real counterexample to DIT, we need to add that, in Nielsen's example, she in fact *believes* that she cannot cut her own hand (and not merely suspect it, for instance).

A similar kind of counterexample, namely a case of desperate action, was raised by those argued in the sixties that we can in fact 'intend to do the impossible' (against *Realistic Decision*). We might think that it is possible that a lifeguard intends to save a swimmer he believes to be dead (Thalberg 1962, 54) or that a prisoner intends to resist torture even though he is convinced that he cannot (Anscombe 1963, 94). The thought is that if the prisoner succeeded in resisting torture or if the lifeguard succeeded in saving the man, this resisting and this saving would have been intentional. So, the lifeguard could intend to save the man and the prisoner could intend to resist torture. The thought can be put in terms of decisions: we might think that the lifeguard has decided to bring the swimmer back to life or that the prisoner has decided to resist. If they succeeded, we would say that this came from their decisions; it was deliberate.

At least three answers are available to Thalberg's, Anscombe's, and Nielsen's purported counterexamples. The first is to redescribe the action that the subject decides in favour of. This is Stuart Hampshire's solution (1960, 134): all that the prisoner could decide is to resist torture as long as possible. Then, it happens to be longer than he thought. This means that his action of

[38] For a similar case, see James (1912, 96–97).

resisting until the torturer stops was deliberate, but under a different description. Similarly, when deliberating about what to do, Nielsen or Ralston might provisionally entertain a different kind of possibility such as 'make just *this* cut.' This solution is elegant, but it might not apply easily to cases like that of the lifeguard (see Hampshire 1962, 54).

A second response consists in using the notion of pragmatic encroachment.[39] In fact, the stakes are very high in all the examples; and it seems like any counterexample to *Realistic Decision* or DIT will have to be similarly dramatic. We could insist that we cannot decide or deliberate rationally to do what we believe to be impossible but that the level of disbelief that is required depends on pragmatic considerations. Since the stakes are very high in the case of the tortured and of the lifeguard, their belief in their inability might not be strong enough to make their decisions irrational. To put the point in terms of assumptions, the stake of the scenario might allow that we assume all sorts of things which would have otherwise been irrational—to leave aside our beliefs. In low-stakes scenarios, this would not be true. And even in high-stakes scenarios like when I am stuck on the rock face of a mountain, some of my beliefs, e.g. my belief that I cannot jump on the other side of a valley, would prevent me from deciding rationally to do so or to deliberate about it seriously (even if unbeknownst to me, a gust of wind would carry me slowly to the other side if I jumped.)

[39] See e.g. Hawthorne (2004).

This brings me to a third response, which is to say that these cases feature a 'rational irrationality.' We can admit that in such cases there *is* a rational decision, intention, or deliberation but that they are irrational in another sense. That is, we could insist that Anscombe, Nielsen, and Thalberg play instrumental rationality against another kind of rationality. The lifeguard, the prisoner, and the hiker might be in fact rational *to* decide or *to* deliberate as they do and yet irrational *in* deciding or *in* deliberating. As we have discussed in §2, this latter kind of rationality could be epistemic or constitutive. This opposition mimics a widely accepted distinction,[40] i.e. the opposition between practical reasons to believe or assert and epistemic or constitutive reasons to believe or assert. Differently put, the counterexample might invoke the wrong sort of reasons in favour of the rationality of some forms of decision. In general, we would happily become dunces of 'theoretical' or 'internal' rationality in order to be heroes of instrumental rationality. But it is only in these heroic or high-stake cases that we can even begin to imagine what it would be like to decide to do what we think is impossible or to deliberate about it.

This raises a connected objection. Even if deliberation compatibilism was false, the leeway sceptic could insist that she only cares about practical rationality: if it is rational and perhaps even essential for her *to deliberate*, then that is the end of the story. (Of course this does not help the deliberation compatibilist.) I can only address this worry very briefly since it touches on general issues about the normativity of rationality.

[40] An exception is Rinard (2017).

Let me first insist that leeway sceptics *should* deliberate, because it is a good means-to end device.[41] And they *do* deliberate, which shows that they have more sense than doctrinal coherence. Since it would be irrational to stop deliberating completely, the only way to be adequately rational is to reject leeway scepticism or to suspend judgement.

Now, whether it matters to be rational or not depends more generally on whether we should care about respecting epistemic norms or constitutive norms of the activities that we engage in. There is a tradition, personified in James (1912), according to which it does because we are *passionate* about rationality—we desire it and we should weigh this desire with others. If this is right, then the sceptic has a cause for regret. (In passing, it is quite frequent to see this passion of rationality in leeway, especially when they insist that leeway is *but an illusion*.) There is also a tradition, personified in Clifford (1887), according to which these norms matter on their own and perhaps so much that they are properly ethical. If, however, these norms do *not* matter for their own sake, it does not follow that they do not matter instrumentally here. Perhaps in disregarding DIT, leeway sceptics will acquire bad deliberative habits which will cause them further troubles. I shall leave this stone unturned.

6. Conclusion

I had three aims: a constructive aim, i.e. to present a fresh and unified account of deliberation (§2); a corrective aim, i.e. to correct the set-up of the deliberation question (§3); and a polemic aim, i.e. to argue for deliberation

[41] See Dennett (2015, 114).

incompatibilism (§4–5). Again, it is likely that my account and correction have convinced more than my polemical argument.

But I think I have mounted a case for it that deserves some second thoughts from deliberation compatibilists and from leeway sceptics. Leeway sceptics must figure out what to do next. If their doctrine is true and that I am right, then, as Castañeda claimed, the world is in a sense 'ugly' (1975, 134): we are forced to choose between a false belief and an irrational practice.

Being myself sympathetic to leeway scepticism, I want to suggest three escape alternatives (quite ironically). The first is to go against the grain[42] and claim that since all deliberation is provisional, Sandrine can in fact deliberate about which film to go see, despite knowing that one of her options is blocked. But then, it is difficult to block the claim that I can rationally decide *to buy a jet tomorrow* in virtue of *having decided to buy a jet were I to win the lottery*.

The second alternative for the leeway sceptic is to claim that we can rationally decide to do what we believe to be unavailable. But then she will have to explain the outlandishness of assertions like 'I shall swim, though I cannot.' This strikes me as a lost cause, but then again perhaps it's possible to rationally decide to win a lost cause.

The third solution is to do what I have hinted at in §5.4, namely to argue that rationality does not matter. If she takes this last route, the leeway sceptic will be playing a dangerous game. For, then, it seems easy to reply that the

[42] This means opposing Taylor (1966, 177–78), van Inwagen (1983, 154), Waller (1985, 49), Kapitan (1986, 247), Clarke (1992, 104), Nelkin (2004, 219), and Pereboom (2008a, 297).

only way to admit in all cases irrationality in deliberating is to endorse a form of general pragmatism. However, pragmatism is a double-edged sword for scepticism. My belief that I have the ability to choose amongst real alternatives makes me happy, so why should I take leeway scepticism seriously?

VII. FROM COERCION TO MANIPULATION

Abstract: Strong manipulation, like coercion, undermines blameworthiness. This general thesis depends on three others, for which I argue in turn. Firstly, structuralism, the view that whether an agent is blameworthy for a wrong action φ supervenes on her mental structure when φing, is false. Secondly, historicism, the view that whether an agent is blameworthy for a wrong action φ supervenes on the combination of her mental structure when φing and its non-interpersonal history, is fragile. Thirdly, interpersonalism, the view that whether an agent is blameworthy for a wrong action φ depends *also* on the intervention of others (like coercion or manipulation) is viable. This yields important consequences for so-called manipulation arguments for incompatibilism.

1. Introduction

It is widely accepted that coercion generally exculpates; that is, having been coerced removes or diminishes an agent's blameworthiness for the (wrong) action that resulted from coercion. It seems natural to think that manipulation is just the same, since both manipulation and coercion are violent ways of controlling others.

Yet several philosophers (e.g. Frankfurt 2002) hold that manipulation is sometimes not exculpatory at all. In a weak sense of manipulation, that of *influence*, everyone should agree: I can 'manipulate' people by bribing them and this surely is not exculpatory. However, in a strong sense of manipulation, that of *brain control*, we should admit that manipulation is exculpatory.

Whether this is right turns on a more general question: what is relevant to blameworthiness? Here, three schools of thought disagree. Structuralists (e.g.

Frankfurt 2002, 28; McKenna 2014) believe that, once we have evaluated the subject's action, we should focus on her mental structure. More precisely, *structuralism* claims (or so I will understand it) that whether an agent is blameworthy for a wrong action φ depends solely on her mental structure when φing.[1] For instance, a structuralist will judge Maria blameworthy if she lies to Fred because she desires to lie to him despite being aware that it would be wrong to do so. Maria's past is irrelevant.[2]

Historicists (e.g. Fischer and Ravizza 1998; Mele 2006) believe that we should also examine how the mental structure of a subject was acquired prior to blaming her. More precisely, *intrapersonal historicism* (henceforth, *historicism*) claims that whether an agent is blameworthy for a wrong action φ depends on her mental structure when φing *as well as* on the history of this mental structure, described in intrapersonal (that is non-interpersonal) terms. For instance, an historicist might judge a brutal warmonger non-blameworthy for extorting money from innocent villagers if this is the clear result of his having been turned into a child-soldier. Yet the historicist will insist that what matters is that the warmonger never had the opportunity to revise his values or that his capacities were impeded, or simply that he was *harmed*. What the historicist will not care about is the fact that the warmonger was *wronged* by someone else.

---

[1] Structuralism is typically defined as a view about *moral responsibility*. Employing 'blameworthiness' is more precise and allows an asymmetry between praiseworthiness and blameworthiness (see e.g. Hyman 2016). I am aware of the possibility to be a structuralist regarding *responsibility* or *free will* and an interpersonalist regarding *blameworthiness*.

[2] There are non-historicists who focus on response to reasons rather than on mental structures (see e.g. Arpaly 2002). The controversial nature of reasons would make our discussion more difficult; thus I leave this kind of view aside.

The warmonger was carried in a never-ending swirl of violence; but whether this violence was human or not is irrelevant.[3]

Interpersonalists think that historicists do not go far enough. What matters for blameworthiness is not only the mental structure of the agent and its intrapersonal history but its *interpersonal* history. It could matter whether the agent's mental structure resulted from a violent past or whether it resulted from her genetic baggage. In particular, manipulation *qua* interpersonal could matter for blameworthiness. Interpersonalists thus extend the spectrum of what is relevant to blameworthiness to irreducibly interpersonal factors (like manipulation and coercion).

Interpersonalism is not often considered seriously, when it is considered at all (see Levy and McKenna 2009; Pereboom 2005; van Inwagen 2017, 124–127). I believe that this is a mistake, one which is due to the lack of attention philosophers have paid to the parallel between manipulation and coercion. As I already indicated, both are fundamentally interpersonal events.

This paper has three primary tasks: to undermine structuralism; to put a dent in historicism; and to reveal the plausibility of interpersonalism. To lighten this heavy programme, I temporarily leave aside the typically *incompatibilist* perspective on blameworthiness (which claims that if determinism is true, no one is ever blameworthy). Still, interpersonalism has implications for *manipulation arguments* (see e.g. Mele 1995, 2006; Pereboom

[3] Some historicists might want to accept meaning externalism where meaning is *inter*personal. I leave this view aside, which disagrees with historicism in words more than in spirit.

2008b, 2014a; Todd 2013, 2016), which are in fact arguments for incompatibilism. Presenting them will be my fourth task.

Here is the plan. §2 lays out my own argument against structuralism. §3 supports each step of the argument. §4 runs a modified version of the argument against historicism. §5 briefly discusses the plausibility of interpersonalism. §6 draws an important lesson for manipulation arguments from interpersonalism. §7 considers a final objection.

2. Against Structuralism

Here is a rough version of my argument against structuralism. There are cases where coercion exculpates. Now, amend these cases such that the coercer gives the coerced the option of being strongly manipulated into doing the same wrong instead. Since the coerced would be equally blameworthy to give in to coercion and to give in to manipulation, the coerced would not be blameworthy to give in to manipulation. But surely to remove the 'coercion option' (and thus simply manipulate) would not add any blameworthiness. So strong manipulation, like coercion, exculpates. This is the case even if structuralist conditions are satisfied. Therefore, structuralism is false.

2.1 The Coercion Argument

Here is the long version of the argument, which is inspired from Derk Pereboom's manipulation argument (see Pereboom 1995).

Four-Case Coercion Argument:

> **Background:** An evil genius, Grey, wants to ensure that by Christmas Day his enemy—Black—is punished by seeing his

home burned down. (Black and his family are on a holiday cruise.) Black has a good-natured neighbour named Brandy, who would never harm Black under ordinary circumstances.

Coercion: Brandy is coerced into setting his neighbour's house on fire. If she does not comply, Grey will kill her. Reluctantly, Brandy complies with Grey's demand.

Choice 1: Brandy is offered two options. Either she is to set Black's house on fire or she can refrain to do so. In this latter case, Grey will turn Brandy (during her sleep that night) through a foolproof manipulation device into a malevolent character who will set Black's house on fire the next day. Grey assures Brandy that if she is manipulated, her malevolent character will progressively disappear in such a way that no one other than Black will suffer from the whole process. If Brandy attempts not to fulfil one of these conditions, she will be killed. Brandy takes the first option: she reluctantly takes the gas tank outside and makes Black's house disappear in flames.

Choice 2: Same as **Choice 1**, but Brandy picks the second option. She goes to bed as usual. The next day, after the manipulation, she deliberates carefully and then sets Black's house on fire. Structuralist conditions on blameworthiness are satisfied: her action expresses her newly acquired character, is the result of a reason-responsive mechanism, and so on.

Manipulation: On Christmas Eve, Grey breaks into Brandy's house. He has a foolproof manipulation device, which he uses to alter Brandy's character overnight, turning her into a malevolent individual who will wake up and set Black's house on fire the next day. (She will satisfy the same conditions on blameworthiness as in **Choice 2**.)

My *Four-Case Coercion Argument* makes the two following claims:

(I) Brandy is blameless in the first case.

(II) For each pair of cases in the sequence, there is no (relevant) difference in *whether* Brandy is blameworthy.

They together entail:

(III) Brandy is not (relevantly) blameworthy in the last case.

I take it that (I) is obvious. I will argue for (II) in detail in the next section.

To assent to (III) is just to reject structuralism, since Brandy satisfies all structuralist compatibilist requirements in **Manipulation**. We can name a few of these requirements while duly noting that more can easily be added: Brandy's effective desire to act can conform to her second-order desires (Frankfurt 1971). She has the ability to do the right thing and has a fair opportunity to do so (Hart 1968, 45; Kenny 1976, 151; 178, 34). At least, she can have these capacities and opportunities if they are given a compatibilist interpretation. The mechanisms that produce Brandy's action can be sufficiently responsive to reasons (Fischer and Ravizza 1998; Sartorio 2016). She can grasp moral reasons and use them to regulate her actions (Wallace 1994). Finally, Brandy's action does not result from her ignorance (Hyman 2013, 685-686).

A word of warning: I am aware that there might be an important difference in the *reason why* Brandy is not blameworthy from case to case. There are difficult questions concerning whether she is *justified* or *excused*; she

might be justified when coerced,[4] but excused when manipulated. But this consideration is orthogonal to my view since I am merely interested in whether blameworthiness itself depends on irreducibly interpersonal factors. (If a certain form of manipulation exculpates, it does; and so structuralism is false.)

Two notes on the nature of the manipulation. First, we can suppose that it is of the strongest kind: it is physically invasive, intentional, and perfectly effective (assume determinism if you want). It makes a difference in that, absent this manipulation, Brandy would have acted differently. Second, Brandy has been *globally manipulated*, that is, a manipulator has changed her *character* or *values* without her informed consent. Global manipulation is opposed to *local manipulation*, which occurs when a manipulator intrusively changes a subject's *actions*, *feelings* or *thoughts* without her informed consent. When a subject is strongly and locally manipulated to perform a wrong action (for instance, the subject is given the irresistible desire to jump through hypnosis), she often does not satisfy structural requirements for blameworthiness. On the contrary, a globally manipulated subject can more easily satisfy these requirements, as we have just pointed out.

2.2 Why Start with Coercion?

Brandy is obviously blameless in **Coercion**. Having been coerced is widely taken to be an exculpatory factor which is strongly implemented in our moral

[4] This is contentious, for it relies on the assumption that coerced acts are simply cases of choosing the lesser of two evils.

practice;[5] and virtually everyone accepts this as legitimate (see e.g. Ayer 1954; P.F. Strawson 2008; Frankfurt 1969, 4–5; Fischer 1994a, 151). At least, this is true of coercion, like Grey's, which involves the threat of rights infringement.[6] Why, then, start with **Coercion**?

Coercion is useful to the argument, for two reasons. The first concerns our method of moral investigation, which resembles the way we seek the right punishment for a crime. We start from a case that we think we have a good grip on and we evaluate whether a second case is similar enough to it to warrant the same kind of blame (or punishment). But to do this, we need a common starting point; one which cases of manipulation do not give us. The 'dialectical stalemate' produced by conflicts of intuition in the manipulation literature show that this is a necessary precaution (see e.g. Fischer 2016, 52–56; McKenna 2014, 473; Sartorio 2016, 167). A clear case of coercion is ideal for this purpose.

The second reason to include **Coercion** in the series is to emphasise a feature of manipulation, which is present across the cases. Although the reason why Brandy is blameless in the fourth case might not be *identical* to the reason why she is blameless in the first, there is an important continuity. At each step, Brandy remains under the illegitimate control of Grey. And her action is explained by an event which constitutes an important wrong. After all, people manipulate and coerce others in order to control them; the means

[5] Coercion (or 'compulsion') is an excuse in the Odyssey, the *Code of Hammurabi*, the *Mānava-Dharmásāstra*, the Deuteronomy, and the Qin dynasty provisions (see Hyman 2015, 78).

[6] Michael Garnett (2018) calls it deontic coercion.

are different, the end is the same. This gives us good reason to think that manipulation is very similar to coercion: it is a sort of coercion without the possibility to heroically resist. (More on this in §5.)

This gives us a *prima facie* reason for thinking that Brandy remains non-blameworthy across the cases. We have a common explanation for Brandy's non-blameworthiness. This is of course not *necessary* for our purpose because (III) already implies that structuralism is false. That is, structuralism is false if Brandy is not blameworthy in **Manipulation**, whatever the reason why Brandy is exculpated for her wrong action. But a common explanation which plays on the similarity between coercion and manipulation warrants the claim that in fact (II) is true. That is, it warrants the claim that there is no difference in whether Brandy is blameworthy across the cases because in each Brandy is under the malevolent control of Grey. If you are not convinced by this inference, despair not: I will provide other principled reasons for thinking that Brandy remains in fact non-blameworthy at each step of the argument. (The possibility to supplement the *Four-Case Coercion Argument* with an inference to the best explanation makes it even more similar to manipulation arguments (see §6).)

2.3 Worthy of Blame?

I want to address a final worry before I justify each step of the Coercion Argument. It is easy to confuse two questions: whether someone is blameworthy and whether we should blame them. There are obviously practical considerations which are relevant to the second question but which are perfectly irrelevant to the first. On top of this, there might be *standing*

conditions to blame (like whether the wrong act is the blamer's business, see e.g. Todd 2017). Again, these concern only the question of whether we should blame someone—not whether they are blameworthy.

Thus, the structuralist might feel comfortable admitting that the coerced and manipulated *are* blameworthy, since she could add that, nevertheless, we should not blame them. We should not blame them since we have a better target to blame, i.e. the coercer and the manipulator. This, I think, is a mistake: for there is no practical reason—at least in our case—to only blame one person or to only blame the worst agent. (To see this, we can simply imagine that Grey dies shortly after the events.) But then, the idea that we should not blame Brandy because there is a better candidate for blame seems to rely on the doubtful idea that we have a limited quantity of blame to distribute per crime. But, as cases of coordinated assassination show, this is absurd—two assassins do not get half the blame because they cooperated.

3. Justifying Each Step of the Coercion Argument

Since the structuralist happily exonerates Brandy in the first case, he has to draw the line at another step in the four-case argument. This is in itself a significant challenge to structuralism. So, in a sense, my argument against structuralism could stop here.

Instead, I want to turn this challenge into a real puzzle. To do so, I will provide principled reasons to judge that Brandy's is equally blameworthy across each case. These principled reasons are controversial, and I do not expect the reader to accept them all. After all, they depend on more general

ethical disputes. But these reasons show that to refute the Coercion Argument, i.e. to demonstrate that there is a pair of steps such that Brandy is not blameworthy in the first but is blameworthy in the second, is difficult.

3.1 From Coercion to Choice 1

What distinguishes **Coercion** and **Choice 1** insofar as blameworthiness is concerned? Only one insignificant element: in **Choice 1**, Brandy is granted an alternative. But this alternative is not relevant to her blameworthiness for burning Black's house, because it is not a relevantly more moral route (and Brandy is aware of this). I say 'relevantly' because even if the most moral decision is to resist, it does not follow that Brandy is blameworthy in **Choice 1**; at least it does not follow that Brandy is blameworthy in the way an uncoerced arsonist would be. Brandy does not deserve blame *for arson*.

I believe that this is what should be maintained by both consequentialists and deontologists. I can only give a rough picture here. I understand consequentialism and deontology as the two broad families of ethical theories which take *action* (and omission) to be the primary object of the moral evaluation. While consequentialism claims that whether an action is right is a function of the value of its consequences or of its reasonably expected consequences, deontology denies this (and typically focusses on intentions).

For the consequentialist, Brandy's decision is close to indifferent. If consequentialism considers that the right action is the one that actually has the best consequences or reaches a certain threshold of good consequences, then choosing to commit the arson right away is perfectly acceptable. This is the

case, whether one takes 'good consequences' to refer to pleasure and absence of pain, to satisfaction of preferences, to non-infringement of rights, to happiness, etc. If consequentialism considers that the right action is the one that has the best reasonably *predicted* consequences, then the second option is perhaps slightly better, but not significantly enough to make Brandy seriously blameworthy. Nothing prevents us from filling the details of the story to make it obvious to Brandy that Grey *is* a proficient and resolute manipulator.

For the deontologist, things are slightly (and typically) more complicated. Brandy is perhaps exculpated because she becomes the tool of Grey; she is in some sense under his control. Under coercion, Brandy's decision is ultimately Grey's decision (cf. Nozick 1997, 38); but now she has not been given the chance to stop being the tool of Grey, for she knows that she will either be manipulated or coerced into doing something wrong that will fulfil Grey's ill intents.

Perhaps Brandy has a duty not to co-operate with Grey and thus perhaps the second option is better. But we usually think that this sort of heroism is not *required* of subjects, especially when the subject knows that resistance is futile. In other words, it is one thing to ask for heroism; it is another to ask for useless heroism. Anyway, the difference between indulging in useless heroism and giving way to coercion cannot account for the difference between the blameworthiness of Brandy in **Coercion** and **Choice 1** and that of a regular arsonist, which is the blameworthiness that the structuralist wants to attribute to Brandy at some point (at least by the time we have reached **Manipulation**).

Lastly, perhaps taking the first option can be justified by self-defence because manipulation is intrusive. Brandy might even have a duty to herself not to allow others to manipulate her. Being manipulated is after all to be wronged. Even if the effects of manipulation wear off after a few days, Brandy will have been wronged. But agents are generally not blameworthy for actions that are performed in self-defence.

Some will simply think that Brandy is not blameworthy because she could not have done otherwise (e.g. Widerker 2000, 191). But my judgement about **Choice 1** does not rely on the idea that absence of alternatives exculpates. Brandy's action does not manifest a deep commitment or desire to commit arson. She does not want to set the house on fire; she does it reluctantly. In fact, the reason why Brandy burns the house down in **Choice 1** is that she is threatened, added to the fact that she (reasonably) does not see a way out of this. To use a principle that even Frankfurt suggests as necessary for responsibility and thus for blameworthiness (1969), Brandy acts the way she does *because* she lacks relevant alternatives and thus is blameless.

Bottom line, even the structuralist should at least admit that coercion exculpates Brandy (in **Coercion**) and that giving her a morally irrelevant alternative will not change this. Therefore, in **Choice 1**, Brandy does not deserve to be blamed for giving way to coercion.

3.2 From Choice 1 to Choice 2

The passage from **Choice 1** to **Choice 2** is tricky to justify with precision. Allow me then to proceed slowly. Firstly, we have:

(A) In **Choice 1**, Brandy is not (relevantly) blameworthy.

This is justified by what we have said in the last section.

In both **Choice 1** and **Choice 2**, Brandy makes a moral decision about whether she will set the house on fire now or whether she will refuse to co-operate, knowing that the next morning, having acquired a new personality, she will perform the same action. In the last section, we have started to show that this decision was morally indifferent, mostly by giving reasons to think that refusing to co-operate was a useless sacrifice. This is our second premise:

(B) In **Choice 1** and **Choice 2**, Brandy's *decision* is morally indifferent (*or the difference is irrelevant*) and she is aware of this.

We have already argued that deciding to immediately set the house on fire was an acceptable option because the alternative, to refuse to co-operate and to foresee manipulation, was not much *better*. We now need to make it clear that this alternative is not much *worse* either in order to justify (B).

(B) can again be justified both from a consequentialist and a deontologist perspective. The consequentialist will, as we have pointed out earlier, not consider that there is a significant difference in consequences between the decision involved in **Choice 1** and the decision involved in **Choice 2**. Apart from the uncertainty, which favours **Choice 2** over **Choice 1**, and the possible trauma of being manipulated, which can be removed from the story, there is not much that pulls the consequentialist to one side or another.

What about the deontologist perspective? If anything, Brandy's decision is better in **Choice 2** than in **Choice 1**. In **Choice 2**, her decision is to refuse to

co-operate. One could object as follows. To commit a crime whilst having the character, intentions, or attitudes of a criminal is worse than being a good person that is forced to commit a crime. This is because character, intentions, and attitudes, change the moral value of an action, regardless of its consequences. These claims might be correct, but it does not follow that the right decision is to give in to coercion immediately. The reason for this is that Brandy's decision in **Choice 2** is *to refuse to co-operate, thus foreseeing that she will be manipulated*; not *to accept to be manipulated*. For, as we well know, we can intentionally act without intending the consequences of this action. And this distinction is typically important for the deontologist (see e.g. Hills 2003, McMahan 2009).

Now, even if the morally best *decision* was in fact to give in to coercion in order to avoid manipulation, it would not follow that Brandy is relevantly blameworthy to *decide* to resist. This decision is, again, nothing like the decision of an uncoerced and unmanipulated arsonist. At the moment of the decision (the decision in (B)), Brandy is very reluctant to set Black's house on fire. And the anticipated consequences for others are the same. This reasoning does not presuppose any contentious consequentialist principle. This is because the claim is *not* that the two courses of actions are (roughly) morally equivalent for Brandy because the consequences are identical. Rather, the claim is that the two decisions are (roughly) morally equivalent for Brandy because they both (i) have the same known consequences for others *and* (ii) result from a morally acceptable intention.

It follows that Brandy would not be any more blameworthy *for making one decision* rather than the other:

> (C) If Brandy's decision is morally indifferent (and she is aware of this), then if she is not blameworthy for deciding one way, she is not blameworthy for deciding the other.

From (A), (B), and (C), we conclude that:

> (D) Brandy is not blameworthy *for her decision* in **Choice 2**.

So far, all I have talked about is Brandy's decision, and not her action of setting a house on fire.

Our result so far is not surprising. Imagine that Brandy is nervous whenever she has to choose actions that lead to similar outcomes. (This is why she avoids restaurants.) Picture a stressed Brandy who is deliberating in **Choice 2** or in **Choice 1**. She starts panicking but the fact that the case is difficult makes it worse than usual. Sophia helps her to calm down: 'Listen Brandy. The authorities will understand regardless of what you do; your options are morally equivalent. Or, even if there is a distinction, it will *not be a significant one*; at least *certainly not one that will put you in prison*.' I am strongly tempted to take Sophia's words as words of wisdom. Thus, we have a confirmation of (D): Brandy is not (relevantly) blameworthy for *deciding* according to **Choice 2** rather than according to **Choice 1**.

Now, structuralists want to say that Brandy is seriously blameworthy *for setting the house on fire* in **Choice 2**. But this, alongside (D), entails that Brandy is seriously blameworthy for setting Black's house on fire in the morning,

although she is not blameworthy for choosing an option she knew would result in her setting Black's house on fire. This is very difficult to admit. How could you not be blameworthy for initiating a sequence of events whose end you know will ensue, but at the same time be seriously blameworthy for an event that is part of this sequence? This seems problematic, and so we should grant:

> (E) If Brandy is not blameworthy for her decision in **Choice 2**, she is not blameworthy for her action in **Choice 2**.

This vindicates my passage from **Choice 1** to **Choice 2**.

(E) is also plausible because it is the instantiation of a more general principle which seems indispensable. When someone *is* blameworthy for an action, then they are *also* blameworthy for knowingly putting themselves in a position which will result in them performing this action. Imagine the following situation. Meredith has to hand in an essay tomorrow morning. She knows that if she gets intoxicated, she will be incapable of writing it. She thus indulges in drinking. Her mother intervenes: 'How could you put yourself in such a mess! You have an essay to write!'. Meredith responds: 'Don't blame me for drinking, mum; blame me tomorrow when I don't hand in my essay'. The argument is then that without our general principle, Meredith's mother could make no satisfactory response. But this is absurd. So, we have good reasons to grant the principle. The challenge for the structuralist is to find a non-*ad hoc* way of amending it, so that it could be used by Meredith's mother while not vindicating my judgement about **Choice 2**.

3.3 From Choice 2 to Manipulation

Finally, we are warranted to claim that Brandy is identically blameworthy in **Choice 2** and in **Manipulation**. This can be explained by yet another moral principle. Again, suppose I am a middle manager and higher management demands that I either fire or suspend an employee. Suppose that I would be blameless in suspending my employee. Then, if higher management had only given me the option to suspend him, I would, yet again, be blameless. The underlying principle is that if an agent is not blameworthy for an action in a sequence *T* and that *T** is just like *T* except that the agent in it has fewer options, then the agent is blameless in sequence *T**. In a word, removing options does not add blameworthiness. If anything, sometimes *adding an option* makes the agent blameworthy; and sometimes *removing an option removes* blameworthiness.[7] At least, this is true when removing an option does not change the subject's intentions; but Brandy's intentions are identical in **Choice 2** and in **Manipulation** when she acts.

Some might point out that the two situations are more different than I claim: in **Choice 2**, there is coercion while in the other there is not. Adding an intermediary case between **Choice 2** and **Manipulation** shows that this is not a serious difficulty:

> **Announced Manipulation:** Like in **Manipulation**, but Brandy is informed (perhaps with a threatening tone) that she will be manipulated into burning Black's house.

[7] This might be because removing an option cannot add control. But if the agent does not have enough control in **Choice 2** to deserve blame, she cannot have enough control to deserve blame in **Manipulation.**

It should be obvious that there is no difference in Brandy's blameworthiness between **Announced Manipulation** and **Manipulation**. Now, there is in fact no other difference than a *decision* between **Choice 2** and **Announced Manipulation**.

A last objection: you might be worried that **Manipulation** does not feature the same *action* as **Choice 2**. If this were true, the principle I have suggested to support my judgement would be irrelevant. We may consider two options. A first possibility claims that the action in **Choice 2** is *setting a house on fire at t2 in order to avoid setting a house on fire at t1*; and that the action in **Manipulation** is *setting a house on fire at t2*. But this action-individuation is incorrect, for in **Choice 2**, at the moment of the action, Brandy no longer sets the house on fire *in order to avoid setting it* on fire at *t1*. This 'in order to' clause carries an intention which Brandy no longer has. A second possibility is that, in **Choice 2**, the action is *to set a house on fire at t2 as a result of having avoided setting a house on fire at t1*. This individuation is acceptable because 'as a result of' does not imply that Brandy has a specific intention at t2 in **Choice 2**. But if this is the correct individuation of action in **Choice 2**, the correct individuation of action in **Manipulation** should be similar: *to set a house on fire at t2 as a result of manipulation*. Thus, Brandy is no more blameworthy in **Manipulation** than she is in **Choice 2**.

On a side note, it might be useful to emphasise that the *Four-Case Coercion Argument* does not rely on the mere fact that Brandy receives a prediction about her future actions. Suppose that the manipulator was replaced with a clairvoyant who told Brandy that she would commit arson the next day (with

the desire to do so) unless she committed arson today. Brandy could be shocked because right now she does not have a desire to commit arson. This would not (at least, not obviously) excuse Brandy for either course of action. We expect people not to turn into arsonists (leaving aside conditions, which could affect structuralist requirements for blameworthiness). However, we do not expect people to resist strong manipulation. (Additionally, whether Brandy is told that she is going to be manipulated or not seems perfectly irrelevant to whether she will deserve blame for arson.)

This concludes my vindication of the Coercion Argument against structuralism. Note again that this is *one way* to justify the judgement that Brandy is just as non-blameworthy in the first case as she is in the last. Others might prefer other moral principles. Yet others might flip the argument, and take the Coercion Argument as a reason against one of the plausible principles which I endorsed. In all cases, we make moral progress.

4. Against Historicism

So far, my argument shows that structuralism is in hot water (at least it has made the water *hotter*). I now turn to historicism. Some historicists, who are initially reluctant to follow Frankfurt in attributing blameworthiness to manipulated agents, end up maintaining that strongly and globally manipulated agents are only blameworthy *sometimes*. Their position is more comfortable than Frankfurt's (see McKenna 2014, 154) because the manipulated agents whom they judge as blameworthy not only have the right capacities or mechanisms but also satisfy some historical conditions on blameworthiness.

We can name a few of these historicist conditions for blameworthiness while duly noting that if there are more to be added, there will be nothing in the way of ensuring that they are satisfied in **Manipulation**. A globally manipulated agent may need to have progressively endorsed and taken responsibility for certain actions (endorsing them as theirs and accepting blame or praise for them) (Fischer and Ravizza 1998, 241–243). She may need to have been able to revise her commitments and character over time (Mele 1995, 2006). She may be required not to have issued an action from unsheddable values which she did not have the opportunity to critically assess, endorse, and sustain (McKenna 2016, 97).

Historicism is not as comfortable as it looks. This is not to say that historicism is as unacceptable as structuralism: my claim is that my argument will only show an unexpected *cost* of historicism. The argument amends the *Four-Case Manipulation argument* in a such a way that Brandy satisfies all the historicist requirements on blameworthiness across all four cases. Here, I will simply amend my set of cases so that they satisfy one particular kind of historical condition on blameworthiness: *that an agent takes responsibility for her actions* (Fischer and Ravizza 1998). Whether this controversial historical condition is correct does not matter for my purpose: my revised argument should work in the same way with any sort of historical conditions on blameworthiness.[8]

[8] There might be criteria for which it is trickier to apply. But this could indicate that the criteria are actually interpersonalist criteria. I have in mind Alfred Mele's bypassing requirement (1995, 144–176; 2006, 164–173; 2008, 267–268, 271).

Fischer and Ravizza (1998, 241–243) maintain that for an agent to be responsible (and *a fortiori* to be blameworthy) for an action produced by mechanism *x*, the agent must have taken responsibility for the behaviour issued by *x*. This general criterion is supposed to explain how children become progressively responsible for their actions. They slowly develop a sense of agency and then take responsibility for their actions: they first understand, test, and endorse the mechanisms that issue some actions; second, they understand that sometimes they are the fair target of praise and blame for these actions. Take adolescents whose bodies become stronger and clumsier and whose desires change radically. In order to become fully responsible, they need to exercise, test, acknowledge, endorse these new aspects of themselves. A similar analysis of Brandy in **Manipulation** leads us to the judgement that, like the adolescent, she needs some time before she is fully responsible for the actions which result from her new mechanisms, dispositions, desires, character traits, capacities, etc. and that she can only be blameworthy for her arson after she has had the opportunity to acknowledge, test, and endorse these new mechanisms, dispositions, desires, traits, and capacities.

This invites us to replace our cases with historical versions:

> The first case is **Coercion**.
>
> In **H-Choice 1** and in **H-Choice 2**, which replace **Choice 1** and **Choice 2**, Brandy's options are changed. She may either act now

---

Roughly, this requirement claims that an agent is blameworthy only if her action is not the result of attitudes which were given to her in a way that bypassed her mental life in general and her capacity to deliberate in particular. It is unclear whether *bypassing* is an interpersonal or an intrapersonal notion.

> under coercion or refrain, knowing that she will be manipulated eventually. Brandy is informed that if she picks the second option, Grey will use a foolproof manipulation device which will slowly give Brandy the egoistic character that will cause her to set her neighbour's house on fire. She will come to take responsibility for the mechanisms that issue her actions. In other words, she will accept as *hers* her new dispositions, emotions, reasons, and she will accept that she is the adequate object of praise or blame for actions issued by these mechanisms. At first, she will have egoistic impulses that are foreign to her, but like a teenager who progressively endorses an unsettling sexual orientation or an unconventional professional aspiration, she will eventually become the kind of Brandy that will unreluctantly set her neighbour's house on fire and who will feel that *this is her action*. None of this transformation of character would occur were Brandy not to be manipulated.
>
> In **Manipulation**, Brandy is not manipulated overnight; the process of developing a new character is identical to the one in **H-Choice 2**.

The progression is about as natural from one historical case to the next as from a non-historical case to the next. There is no significant difference about whether Brandy is blameworthy between each step.

The very same arguments as before apply. Again, in **Coercion**, Brandy is blameless. **H-Choice 1** does not feature a more blameworthy Brandy than **Coercion** because **H-Choice 1** (just like **Choice 1**) merely adds a bad escape from a coercive situation. If Brandy should be excused for setting Black's house on fire in **Coercion** and in **Choice 1**, then she should also be excused in **H-Choice 1**. Again, the passage from **Coercion** to either **Choice 1** or **H-**

Choice 1 does not rely on the idea that it is always possible to do the right thing. We might sometimes find ourselves in positions where whatever we do, we will be blameworthy because we have done something wrong; for instance, we have not sufficiently cultivated our skills or moral virtues; we have not been careful enough, etc. But this is not such a situation because were the manipulator absent, Brandy would have had a good action available to her. Grey makes it the case that Brandy has two equally bad options which Brandy would strongly prefer to avoid. This should be enough to exculpate her for setting the house on fire immediately in **H-Choice 1**, a decision which results from her not having adequate alternatives.

Brandy is equally blameworthy in **H-Choice 2** and in **H-Choice 1** because both options offered to her are approximately morally equal (and Brandy knows this); as a result, she would not be seriously more blameworthy for picking one over the other. And because of this, she cannot be seriously blameworthy for carrying out an action which she foresaw resulted from a decision that she was not seriously blameworthy for making (see §3 for more details).[9] Lastly, **H-Choice 2** and **Manipulation** feature equally blameworthy agents because the distinction between the two is simply a matter of removing an option, which cannot make blameworthiness appear (assuming that the intention does not change).

If the transition from each case of the historical *Four-Case Coercion Argument* is less convincing, one must consider how the transition could be

[9] There might be borderline cases which disprove the validity of this inference, but they are irrelevant here because the principle of inference could easily be amended.

warranted in the initial argument but not here. True: Brandy has more alternatives or opportunities (under a compatibilist interpretation) to depart from Grey's set path in the historical cases than in the non-historical cases. But what matters should not be the *number* of opportunities she has. Compare: in **Manipulation**, Brandy technically has the opportunity to refrain from setting her neighbour's house on fire; in **H-Manipulation**, she has simply more opportunities to do so. This is perhaps too quick. Brandy might have a different *kind* of opportunity in **H-Manipulation** compared to what she has in **Manipulation**: not just the opportunity to do the right thing (not burn Black's house) but the opportunity to revise her commitments and character.

Where can the historicist draw the line then? The historicist judges that Brandy is blameless in **Coercion**, but blameworthy in **H-Manipulation**. She should judge Brandy blameworthy in **H-Choice 2** because the only difference between **H-Manipulation** and **H-Choice 2** is that Brandy is given an extra option in the latter. This means that there are two places to draw the line: between **Coercion** and **H-Choice 1** or between **H-Choice 1** and **H-Choice 2**.

Yet each possibility gives rise to a paradox. Suppose, on the one hand, that the historicist draws the line between **Coercion** and **H-Choice 1**: Brandy is not responsible when coerced, but she is blameworthy for co-operating in **H-Choice 1**. In general, we do not want to say that *S*, who is blameless in situation X, could be blameworthy in situation Y where Y is like X except that *S* is given an extra option that is unacceptable. In particular, if Brandy is blameless in **Coercion** but blameworthy in **H-Choice 1**, the extra option

should be relevant or acceptable. But it is paradoxical to maintain with the historicist that this option is acceptable since she is claiming that accepting it would make Brandy blameworthy. (This option in fact has to lead to a blameworthy action for the historicist because it is obvious to him that Brandy is blameworthy in **H-Choice 2**.)

Suppose, on the other hand, that the historicist draws the line between **H-Choice 1** and **H-Choice 2**. Then we run into the problem that I mentioned in §3.3. We are forced to say that Brandy is blameworthy for her action, but blameless for knowingly choosing to put herself in a situation which will result in this action. Yet, this problem is not as acute for the historicist as it is for the structuralist. That is because the historicist can insist that there is a distance between Brandy's decision to be manipulated and her action under manipulation. It makes it more palatable to admit an asymmetry between the blamelessness of her choice and the blameworthiness of her action. This is why I said that my argument merely puts a dent in historicism.

5. For Interpersonalism

Interpersonalism is not a very popular view. Consider the following:

> Try as they might, compatibilists of any stripe will eventually have to admit that it is at least in principle possible to imagine a case of manipulation that includes everything a compatibilist would demand (barring of course a mere *ad hoc* amendment that no global manipulation is allowed at all). (Levy and McKenna 2009, 110; see Pereboom 2005 for a similar suggestion)

In other words, there are just two sides. There are compatibilists who are either structuralists or historicists, but who all need to admit that sometimes strong manipulation does not matter, morally speaking. As John Martin Fischer once put it, manipulation is their 'dirty little secret' (2000). And there are incompatibilists, who have a way of saying that strong manipulation is always exculpatory (strong manipulation is like determinism). Interpersonalism is not seriously considered; it is counted as the mere *ad hoc* suggestion that 'manipulation is not allowed'.

I believe this suggestion is hasty. There are at least two ways[10] to hang onto compatibilism and legitimately bar the possibility that a strongly and globally manipulated agent be blameworthy.

The first way consists in rejecting Neil Levy and Michael McKenna's methodology. Manipulation is an irreducible exculpatory factor, but this is not problematic. To put it bluntly: if this view counts as *ad hoc* then *ad hoc* is fine. Blameworthiness is understood negatively in terms of a (long) list of things that eliminate or affect it rather than as the combination of positive conditions on responsibility and the presence of a wrong action (or intention, decision, character, etc.). It may be suggested, for instance, that actions that result from coercion, ignorance, compulsion, and strong global manipulation should be excused. To use one of J.L. Austin's handy expression, it would be exculpation (and excuse) rather than blameworthiness (and responsibility) that would 'wear

[10] I leave aside the suggestion that if manipulation were possible, our current practice of attributing blame would stop making sense. On this, see Waismann (1965, 214–216).

the trousers' (Austin 1962, 70). That is, the second would be understood in terms of the first. This is perhaps too intellectually pessimistic an approach to the problem of blameworthiness, one that entertains no hope of finding the necessary and sufficient conditions of blameworthiness. This approach might not satisfy the most metaphysically-inclined of us but it has had serious defenders (e.g. Austin 1956, Ayer 1954, Strawson 1992).

The second way accepts that to simply claim that manipulation is an irreducible exculpatory factor is theoretically inadequate. But it maintains that there is some more general interpersonal requirement for blameworthiness—a requirement whose satisfaction is incompatible with strong global manipulation. For instance, being under the malevolent control of someone else could be an exculpatory factor. This avoids the charge that judging Brandy non-blameworthy in **Manipulation** or in **H-Manipulation** is ultimately *ad-hoc*—and which give a common explanation to coercion and manipulation. Sometimes what explains why someone acts in a certain way is that someone else has done something to her (e.g. threatened or deceived her). Perhaps whether she has been *wronged* through this previous action (e.g. if she is globally manipulated rather than convinced)[11] has an irreducible impact (one that could not be described in terms of intrapersonal properties) on her blameworthiness for her current action. Or perhaps being the source of our actions is both necessary for blameworthiness and irreducibly interpersonal. That is, whether Brandy is the source of her action depends on whether she was manipulated or not. Perhaps the presence of a manipulator influences

[11] This a suggestion that Nozick (1997) makes about coercion.

sourcehood in a way that other physical phenomena do not (see Deery and Nahmias 2016, 1273).

I claimed in the introduction that I would primarily undermine structuralism, weaken intrapersonal historicism, and suggest that interpersonalism is worthy of our consideration. I additionally said that I would show how these three points impact the treatment of manipulation arguments. I have already completed the first two tasks and have shown how interpersonalism could be formulated. By arguing against historicism and structuralism, I have given an indirect argument in favour of interpersonalism but more needs to be said. In order to give interpersonalism a rough defence, I need to firstly focus on manipulation arguments. I thereafter examine a response to manipulation arguments which stands in the way of interpersonalism. I will thereby have achieved my four goals.

6. Manipulation Arguments and Interpersonalism

Structuralism is false; historicism is fragile. This has consequences not only for the treatment of manipulation, but for manipulation arguments. In this section, I firstly present manipulation arguments; secondly, I show the links between my paper and manipulation arguments; thirdly, I address a criticism of manipulation arguments and of interpersonalism.

6.1 Manipulation arguments

Manipulation arguments are arguments for incompatibilism which use the same method as I used in my *Four-Case Coercion Argument*. They use a series of cases to argue that strongly manipulated and physically determined subjects are

identically non-blameworthy. Here is an example of a manipulation argument, adapted from Derk Pereboom's famous example (see Pereboom 2014a, 71–103):

> ***Four-Case Manipulation Argument:***
>
> **Background:** An evil genius, Grey, wants to ensure that by Christmas Day his enemy—Black—is punished by seeing his home burned down. (Black and his family are on a holiday cruise.) Black has a good-natured neighbour named Brandy, who would never harm Black under ordinary circumstances. (To exclude libertarian concerns, one may assume that determinism is true.)
>
> **Manipulation:** On Christmas Eve, Grey breaks into Brandy's house. He has a foolproof manipulation device, which he uses to alter Brandy's character overnight, turning her into a malevolent individual who will wake up and set Black's house on fire the next day.
>
> **Education:** Years in advance of Black's murder, Grey starts subtly brainwashing Brandy to believe that Black deserves to have his house burned down. By Christmas Day, this brainwashing has been so effective that Brandy has acquired a different character: one that will lead her to burn Black's house to the ground.
>
> **Determinism:** Brandy was born in a deterministic world. The laws of nature and the state of the world at her birth entailed that she would eventually develop an extremely malevolent character, which, on a certain Christmas Day, she would manifest (without being manipulated) by burning down Black's house.

Like in my own argument, this manipulation argument makes the following claims:

(I) Brandy is blameless in the first case.

(II) For each pair of cases in the sequence, there is no (relevant) difference in whether Brandy is blameworthy.

These claims together entail:

(III) Brandy is not (relevantly) blameworthy in the last case.

Yet to assent to (III) is just to endorse incompatibilism, since **Determinism** is a normal case of seemingly blameworthy action in a deterministic world. Thus our inclination to excuse manipulated agents lands us with incompatibilism; equivalently, compatibilists, who deny that determinism exculpates, must also deny that manipulation exculpates (see Todd 2013, 202–3).

There are two main types of manipulation arguments. The first includes an inference to the best explanation. It holds that what explains the fact that Brandy is not blameworthy across the cases is a common fact: the truth of determinism (or the fact that Brandy's actions are causally determined by factors beyond her control) (see Pereboom 2014a, 79). The second type of manipulation argument is more modest, since it does not include an explanatory claim. It only claims that *if determinism is true, no one can be blameworthy*, rather than that determinism is directly relevant to blameworthiness (see Mele 2016, Mickelson 2016). This is the type of argument that I have just laid out. In these arguments, Brandy could in principle be non-blameworthy for different reasons across the cases—and

sometimes for multiple reasons at once. (This is reflected in my own presentation of *coercion* arguments, see §2.2.)

There are two lines one can take in response to a manipulation argument, a soft and a hard line.[12] *To take the soft line* against a manipulation argument is to deny that (II) is true of that argument. This means that there are relevant differences across the cases of manipulation arguments (e.g. Demetriou 2010; Waller 2013; Schlosser 2014; Deery and Nahmias 2016). *To take the hard-line* against a manipulation argument is to deny that (I) is true of that argument (e.g. Frankfurt 2002; McKenna 2008; Haas 2013a). One may take the hard-line for some manipulation arguments and the soft line for others. Taking the hard-line is generally justified by appealing to structuralism or historicism. On reflection, our best theory tells us that we should judge the manipulated agent as blameworthy; so much the worse for our initial reluctance to do so.

6.2 Interpersonalism and Manipulation Arguments

I have argued that interpersonalism might be true because strong global manipulation is exculpatory. A fundamental advantage of interpersonalism is precisely that it can vindicate the view that strong manipulation can never be brushed away.

[12] Philosophers tend to talk of the soft line and of the hard-line as general stances and even as *kinds of compatibilism*. I suspect that this is misleading: for one thing, someone can be an incompatibilist and nevertheless believe that manipulation arguments fail. For another, there are several ways of splitting the hard and the soft lines which depend on the sort of manipulation arguments which are deemed relevant. I will thus avoid this lexicon.

So, if (the right kind of) interpersonalism is true, we should refuse to take the hard-line against manipulation arguments. At least, we should not take the hard-line against *well-formed manipulation arguments*, that is those whose first case satisfies all the standard compatibilist intrapersonal conditions of blameworthiness. While in some cases like the example I offered in §6.1, historicism provides sufficient justification to reject the hard-line, there will be other cases where the historicist takes the hard-line. In these cases, the interpersonalist is justified in refusing to follow her.

For well-formed manipulation arguments, the hard-line is a bad line. We get this result via two ways. First, for most manipulation arguments, we can simply add a coercion argument at its beginning. As I put it in this paper, my *Four-Case Coercion Argument* can simply be added at the front of my *Four-Case Manipulation Argument* (since **Manipulation** is the same in each). This makes the hard-line much harder to take. Second, my *Four-Case Coercion Argument* gives us reasons to endorse interpersonalism; but if interpersonalism is true, taking the hard-line is unjustified. Hard-liners cannot appeal to structuralism or historicism to vindicate their judgement.

This result is perhaps surprising. Many philosophers sympathetic to manipulation arguments have confessed that they do not have much more against the hard-line than an appeal to their intuitive judgements that globally and strongly manipulated agents do not deserve to be blamed for their actions (Todd 2013, 199; Pereboom 2014a, 79; Tognazzini 2014, 367); and their opponents have agreed (McKenna 2012, 149; Cyr 2016, 1075). Although principled explanations for this exculpatory judgement have been given (e.g.

Tognazzini 2014; Pereboom 2014a; Todd 2016), hard-liners can easily reject these explanations as relying on fundamentally incompatibilist assumptions or unsound intuitions (see e.g. Fischer 2016, 52–56). They can also reverse the steps of the manipulation arguments to show that global manipulation does not exculpate or that our intuitions about such cases are not reliable (McKenna 2008, 2014; Fischer 2011). There would then be a 'stalemate' between hard-liners and incompatibilists which would force incompatibilists—because they are the proponents of the argument (McKenna 2014, 473)—to 'go back to the drawing board' (Sartorio 2016, 167).

In this paper, I have risen to this challenge. I have proposed an argument which *ends* with a case of strong global manipulation. This argument provides a reason independent from our intuitive judgement about manipulation itself for thinking that taking the hard-line against well-formed manipulation arguments is incorrect. Crucially, my argument did not rely on incompatibilist intuitions. While manipulation arguments allegedly reveal structural moral similarities between strongly globally manipulated actions and normal actions, my four-case argument revealed structural moral similarities between coerced actions and globally manipulated ones. The upshot of my argument is that because coercion exculpates, so too does strong global manipulation.

Understandably, proponents of manipulation arguments (namely some incompatibilists) might think that this is not very helpful, for if interpersonalism is true, then then they have nothing to support the claim that there is no morally relevant difference between a 'deterministic Brandy' and a manipulated Brandy. My purpose's paper is not to help incompatibilists, but

we should insist that in my argument for interpersonalism, I set aside incompatibilism. My precise conclusion was 'if compatibilism is true, then interpersonalism is viable'. And *this* does not undermine manipulation arguments.

A note on the flexibility of my argument. I suggested that the last case of my *Four-Case Coercion Argument* could in principle be the first case of any well-formed manipulation argument. But this excludes *original design arguments* (see e.g. Mele 1995, 2006; Todd 2013, 2016).[13] In the classic original design case, the goddess Diana creates a foetus because she wants the resulting person, Ernie, to perform a wrong action when he is thirty. My *Four-Case Coercion Argument* cannot be appended to an original design argument because I cannot assume that a foetus is coerced or given an option.

This is not a problem for my argument. First, as I indicated, it undermines structuralism and weakens historicism in general. This means that taking the hard-line is much harder to *justify*, regardless of whether I can in fact append a coercion argument in front of *every* manipulation argument. Second, while original design arguments might be the clearest and most effective of manipulation arguments (see Deery and Nahmias 1258), it is hard to admit that they are the only manipulation arguments which work. On reflection, it

[13] Some, e.g. Fischer (2016) and Mele (2016) think that original design arguments are fundamentally different to manipulation arguments while others, e.g. Deery and Nahmias (2016) disagree. I am happy with either possibilities: in the first, I maintain that my argument is directly relevant to all manipulation arguments and that original design arguments are not inherently superior to manipulation arguments; in the second, I claim that my argument only works directly on non-original-design manipulation argument, but that this is for accidental reasons.

seems that for one who refuses to allow that serious global manipulation always (at least partly) exculpates, there would seem to be little reason to draw a distinction (relevant to our question) between manipulation in one's distant past and at one's creation.[14] It seems that we can turn original design arguments into 'early-stage-of-life' (say adolescence) manipulation arguments. Diana may tweak young-Ernie's biology so that, knowing precisely what will happen in Ernie's life, she can predict that he will perform a specific action during his late adulthood. This is the kind of case for which I can easily adapt a different historical coercion argument.

7. A Last Objection

Before we bring this discussion to a close, we should address an obvious criticism against interpersonalism. We can extract it from John Martin Fischer's response to manipulation arguments featuring the Diana case (2011, 267-270; 2016, 48-62), which we just described. Fischer firstly asks us to replace Diana with 'normal' parents who, when engaging in the act of procreation of Ernie, have the intention of fashioning a killer. As it happens, Ernie becomes a killer, even though he was raised normally. Clearly, Ernie's parent's intentional agency is irrelevant to Ernie's blameworthiness. It seems reasonable to infer from this that there is no reason why Diana's intentional

[14] Demetriou (2010) has argued that we should distinguish 'Create and Tweak' (where scientists permanently manipulate a subject and thus undermine her agency) from 'Create and Release' manipulations (where scientists perform a one-time operation which does not undermine agency) and that we should only be troubled with the second (for which we have to take the hard-line). My case is not a case of creation, but it is a case of 'release' and so it admits Demetriou's wise words. However, if Demetriou finds it acceptable to take the hard-line against 'release' cases, she seemingly will have to provide a response to my *Four-Case Coercion Argument*.

agency in the regular Diana story should be relevant to Ernie's blameworthiness. As a consequence, manipulation arguments featuring the Diana case are in fact simply deterministic scenarios to which the incompatibilist has misleadingly added intentional agency. Therefore, says Fischer, manipulation arguments featuring the Diana case do not give us any new reasons to be incompatibilists. The intuition that Ernie is blameless is debunked.

We can make Fischer's point even stronger. First, the mere difference in *knowledge* between Diana and Ernie's parents cannot be directly relevant to Ernie's blameworthiness. This follows from a more general principle, that is *reliable predictability* is *compatible with blameworthiness*. I can reliably foresee that the group of armed men charging me in a back alley will mug me; they can nevertheless be blameworthy. Second, *biological determinism* is false. As a consequence, the reason why Diana's Machiavellian plan is successful is not because she has implemented genes which determine alone (that is, regardless of Ernie's environment) the future of Ernie. In this regard, she is just like Ernie's parents.

However, I believe that there *is* a difference between Ernie in the two scenarios (with Diana, without Diana). This is true even if biological determinism is false; reliable predictability does not exculpate; mere intentional agency of the manipulator is not in itself exculpatory; and compatibilism is true. The difference is one which perhaps implies prediction and intentional agency but cannot be reduced to them. Firstly, Diana exercises a form of

control over Ernie.[15] This is not the sort of control that Ernie's parents exercise. And, secondly, Diana's control wrongs Ernie. Ernie's parents have not wronged him.[16] Intentional actions are already hard to reduce to a sum of intentions and actions; it is no surprise that wronging is irreducible to a sum of ill intentions, actions, and bad consequences. This general suggestion begs for some further refinement and discussion. But I think it at least *prima facie* dispels some worries about interpersonalism and thus about the soft line against manipulation arguments.

8. Conclusion

A strongly manipulated agent does not deserve to be blamed for the action they were manipulated to perform. It follows that structuralism about blameworthiness is false. While this perhaps can be explained by the similarity of coercion and manipulation, it certainly can be supported by the *Four-Case Coercion Argument*. There is more: historicism about blameworthiness cannot easily deal with strong global manipulation which satisfies historical conditions on blameworthiness. This discomfort is sustained by the historical version of the *Four-Case Coercion Argument*.

This leaves the compatibilist with interpersonalism. While I have not proposed a specific interpersonalist theory, I have sketched two ways in which an interpersonalist may deal with cases of manipulation. The first is intellectually pessimistic: it maintains that the best we can hope for is a

[15] Deery and Nahmias (2016) make a similar point in terms of sourcehood.

[16] This is true even if they might have done something wrong, which is different to *wronging*. Wronging requires a victim.

multicoloured list of precise conditions which exculpate, a list which includes manipulation. The second is more systematic: it gives interpersonal conditions on blameworthiness, and it accepts that these conditions need to be more general than 'absence of manipulation'.

Lastly, if interpersonalism is true, then we should never take the hard-line against well-formed manipulation arguments. This does not mean much for the success of these arguments because it leaves two options open: first, manipulation arguments are sound and so is incompatibilism; second, manipulation arguments fail because manipulation but not determinism is incompatible with blameworthiness.

Interpersonalism is more appealing than its looks. After all, determination by others is more worrying than determination by nature. Even compatibilists feel uncomfortable about the idea of determination by other minds or by systems created by other minds. Manipulative advertisement can sometimes be more frightening than the laws of nature. Interpersonalists have a story to explain this sentiment. To be in control is partly not to be under someone else's control, at least when this control wrongs us.

VIII. EXCUSES AND ALTERNATIVES

Abstract: The principle of alternate possibilities (PAP) says that one is only responsible for actions which one could have avoided. Much of the philosophical discussion about PAP concerns Frankfurt's widely discussed counterexamples to it. In these scenarios, an agent seems responsible for an unavoidable action because what makes it unavoidable has no influence or effect on the actual course of events. Fifty years of discussion has led philosophers to refine the counterexamples and to polish their rebuttals. Attempts to rally everyone seem hopeless as (largely metaphysical) controversies rage. Still, we can make progress on the issue by asking what our reason for believing PAP is. The best answer is this: lacking alternatives is a good excuse. If this is right, lacking *eligible* alternatives—alternatives whose cost is not too high to consider—is a good excuse. Yet, this last principle is subject to counterexamples similar to Frankfurt's—modified cases where an agent would (unbeknownst to him) be shot if he refused to commit a minor offence. In these scenarios, the agent lacks eligible alternatives and yet seems blameworthy. This is true independently of the metaphysical controversies surrounding Frankfurt's cases. It follows that our reason for believing PAP does not withstand scrutiny.

1. Introduction

Several philosophers still endorse the principle of alternate possibilities (PAP), according to which one is only responsible for actions one was able to avoid, that is, actions to which one had alternatives. Why believe in this principle? If anything, it is because lacking alternatives is a good[1] excuse: it is unfair to hold

[1] I talk about 'good excuses' to make it clear that I am not simply interested in the social fact that some states of affairs are *accepted* as excuses. By a good excuse, I will only refer to *full* excuses; excuses which get one off the hook.

people responsible for outcomes lying outside of their control.[2] More precisely, it is lacking an *eligible* alternative, one whose cost is not too high, that is a good excuse. This is reflected in our quotidian interactions: 'don't be harsh, I had no choice but to stay late at work', 'sorry that I did not call you, but I really could not' are common phrases if any.[3] This common practice seems to presuppose PAP, which can give confidence to the PAP defender.

I want to argue that this perhaps banal explanation is important. It has a crucial bearing on whether philosophers have successfully defended PAP. In essence, I contend that an adequate defence of PAP should not only defeat Harry Frankfurt's putative counterexamples to PAP (1969), but secure our reason for believing PAP. Yet the major defences of PAP of the past fifty years, which I will explain in due time, have failed at this second task.

Instead, the debate over PAP was more narrowly focussed on Frankfurt's cases themselves. This has raised several metaphysical controversies over the nature of alternatives, desires, and actions. However interesting, these controversies can be bypassed by directly asking an ethical question: is lacking eligible alternatives a good excuse? My answer is negative, which is bad news for the PAP defender.

Here is my precise argument. (1) Our reason for believing PAP is that lacking eligible alternatives is a good excuse. (2) Yet, on close examination, lacking eligible alternatives is *not* a good excuse, unless the eligibility of an

[2] This is controversial and depends partly on what it is to blame someone. See e.g. Smith (2005) and Scanlon (2008).

[3] The interpretation of the 'I could not have done otherwise' excuse is subject to controversy; see Strawson (1974), Frankfurt (1969), and Dennett (2015).

alternative is relativised to the agent's epistemic situation. (3) And, in this latter case, that *lacking eligible alternatives is a good excuse* provides no support for PAP and thus is not a good reason for believing it. *Therefore*, we have no good reason for believing PAP.

In §2, I argue for the first premise. I then move on to the second premise. To demonstrate its truth, I must show how it cuts across the metaphysical controversies surrounding Frankfurt cases. In §3, I present these controversies. In §4, I argue that we can depart from these controversies in establishing the second premise. In §5, I consider a different sense of 'eligibility' and I argue that the third premise is true. In §6, I explore some options for the PAP defender.

2. Excuses and Eligibility

PAP is under-described. It leaves open the question of whether lacking alternatives is an *excuse* or an *exemption* (or part of either). This distinction, famously theorised by Peter Strawson (1974) and Gary Watson (2004), goes as follows. Excuses apply to specific *conduct*: for instance, we excuse actions and omissions which are the result of ignorance or coercion. Exemptions, on the other hand, apply to *agents*: for example, we exempt young children and those suffering from severe dementia. We exempt them from the kind of interpersonal relationship we have with other adults and which presuppose that we are all in general responsible for our actions. There might be borderline cases between exemptions and excuses: we could say that we excuse the kleptomaniac for theft but not for lies; and we might say that we

exempt some autistic people from blame for antisocial attitudes and actions, but not for laziness; or vice-versa. At any rate, both exemptions and excuses render blame and responsibility attribution unwarranted.

It seems quite clear that lacking an alternative is (if anything) an excuse, not an exemption. We understand that a man who slipped on an icy patch could not help bumping into the person he hit, so we excuse him. And we do not blame our friend for failing to call us if there was no signal on the train. While it might be true that the exempted also lack alternatives, this is not the way that we describe them. We might claim that they lacked emotional or motor control or comprehension, but not mere alternatives. Thus, lacking alternatives is not an exemption. And if it is implied by a condition that does qualify as an exemption, this is far from obvious.

These remarks have important implications for the kind of alternatives which are relevant to PAP. For, if lacking any alternative at all is a good excuse, then lacking an *eligible* alternative is also a good excuse. Again, an eligible alternative is an alternative whose cost is not too high to reasonably consider or choose—and thus one that an informed third party could usually expect one to consider or choose. Now, leaving aside the occasional slipping case, lacking alternatives is almost always used as an excuse when there are in fact alternatives, but only ineligible ones. I live in London and at midday all the trains to Oxford have been unexpectedly cancelled. I *could* hire a taxi for £200. But if I take the bus and arrive at 13:45 for my 13:00 class, my students should accept my excuse. I did not have an eligible alternative, i.e. one whose cost is not too high (and thus one they could reasonably expect me to take). I *had* an

alternative in the taxi case, but it was too costly. The same goes when someone is excused because they were blackmailed or coerced into doing something. But things would have been the same if I had no alternative at all, for instance, if there had been no taxi to hire. I would have lacked an eligible alternative all the same.

As the taxi case shows, what counts as an eligible alternative is decided in part by making a value judgement. Imagine that instead of teaching, I was supposed to perform some urgent surgery at 13:00 in Oxford. In that case, the taxi ride *was* an eligible alternative to arriving late. To take another example, I would be rightly excused for handing over money to a burglar—I had no eligible alternative. But I perhaps cannot be rightly excused for handing secret nuclear plans to an enemy—there I had eligible alternatives.[4] Value thus permeates the question of whether subjects have eligible alternatives and thus relevant alternatives for PAP.[5]

By contrast, the kind of alternative that would be required if lacking alternatives was an exemption rather than an excuse would perhaps not be value-dependent. Suppose a child has eaten half of the clafoutis that was supposed to be served to the guests. Asking whether refraining from eating the dessert was an eligible alternative is not the way to determine whether we should exempt him. We should instead ask whether he lacks the relevant capacity for control or whether he understands how this affects the dinner

[4] Hart (1968, 16) makes a similar point about coercion.

[5] As I will briefly consider in §7, the same might apply to 'can' in "'ought" implies "can"".

plans. In general, if a mentally disabled person or a child is exempt from responsibility, this will not be due to a lack of alternatives: holding fixed their general capacities they would be exempt regardless of their specific situation (i.e. independently of opportunities). Thus, again, lacking alternatives does not seem like an exemption.

When we try to be more specific about PAP, by asking whether lacking alternatives is an excuse or an exemption, we come to realise that it must be the first, if anything. And if lacking alternatives is an excuse, so is lacking eligible alternatives. So we can suggest a more specific variant of PAP, the 'principle of eligible alternate possibilities':

> **PEAP:** One should be *excused* for one's conduct if one lacks eligible alternatives, viz. alternatives whose cost is not too high to reasonably consider or choose.

This is not to claim that PAP is false. On the contrary, it is to explain the grounds we have for thinking that it is true. But it follows that if we discover that PEAP is false, then we should no longer accept PAP—that is, unless and until we can find some other reason to believe it. In a nutshell, our reason for believing PAP is PEAP. Setting acronyms aside, our reason for believing that we are not responsible for unavoidable actions is that lacking eligible alternatives is a good excuse.

In this section, I have vindicated the first premise of my main argument:

> (1) Our reason for believing PAP is that lacking eligible alternatives is a good excuse.

To make it clear, this claim should not be read too strictly: my actual reason for believing PAP might be slightly different from yours. What matters is that our reason for believing PAP is undermined if we discover that PEAP is false.

In the next section, I turn to the second premise, which claims that in fact lacking eligible alternatives is *not* a good excuse, unless we understand eligibility differently than we have so far. Remember that I claimed an eligible alternative is an alternative whose cost *is not* too high to reasonably consider or choose. Things would be different if an eligible alternative was an alternative whose cost *was not reasonably perceived as* too high to reasonably consider or choose. Which conception of eligibility is preferable? I will come back to this question in §5. For now, I will use the non-epistemicised conception of eligibility.

Before we get to our second premise, let me raise a quick objection, following Steward (2009, 2012). Steward argues that lacking alternatives is incompatible with *action* (agency). One does not φ, where φing is a kind of act, unless one can refrain from φing. Thus, we might think, lacking alternatives is not an excuse, and something different from the exemptions we considered. It is not an excuse because it is simply inadequate to ask in the first place whether I have an excuse. Suppose that I come back home, where my partner blames me for having spilt some wine on the carpet. Yet it turns out that the cat was responsible. I do not have an *excuse*, but I should not be held responsible.

This objection is unconvincing for two reasons. The first reason, which I cannot deploy fully here, is that Steward's metaphysics of agency is subject to difficult cases such as the case of a man throwing a dart on a bullseye. It seems that, as Anthony Kenny (1976) argued, it does not follow that the man was able to throw the dart on a bullseye. And so, action does not require ability, not even to do the performed action. The second reason to doubt the objection is that even if we concede that lacking alternatives is incompatible with agency, it does not follow that lacking alternatives is not an excuse. For, from the fact that I did not act, it does not follow that I am in a situation analogous to the situation where I am unfairly accused of having spilt the milk. In fact, I can be responsible for *E* even if *E* is not an action and if I did not act. *E* could be an omission or an attitude for which I am equally responsible (see Clarke 2014). In fact, my relation to the event that is, according to Steward, not an action could be the same relation that I have with my omissions or some of my attitudes, for instance.[6] This is a feature that we see in the law, where criminal liability is thought to include relations such as possession or at least omissions (see Hart 1968; Duff 2007).[7]

3. PAP and the Flicker of Freedom Feud

Is lacking an eligible alternative a good excuse? As I will argue in §4, the answer is negative. And it is negative regardless of whether Frankfurt's

[6] One might object that this is irrelevant to PAP because PAP is about actions, not omissions. Yet, I fail to see the motivation for an asymmetrical approach to omissions and actions—this is why I phrased PEAP in terms of conduct.

[7] Michael S. Moore (2010), perhaps the main advocate of the act requirement, ends up advocating a *control requirement* which encompasses both actions and omissions.

controversial cases are counterexamples to PAP. In fact, asking whether lacking an eligible alternative is a good excuse is a way to *bypass* the metaphysical controversy surrounding these cases. Yet, the controversy is instructive in answering the question. This is why I turn to it in this section.

3.1 PAP: Counterexamples and Defences

Consider Frankfurt's counterexample to PAP. Jones is about to commit a crime for his own reason and without any interference. But Black wishes to make sure that Jones will in fact commit the crime, so he implants in secret a device in Jones' brain. The device will be activated if Jones chooses not to commit the crime. If activated, the device will reliably make Jones change his mind and commit the crime. As it happens, Jones decides to commit the crime on his own. So, he is responsible for committing a crime even though he lacked alternatives, *contra* PAP.

Frankfurt claims that there is a broader lesson that we can learn from this case, namely, that what matters for responsibility for an action is just what actually contributes causally to it and nothing else.[8] Since lacking alternatives does not necessarily contribute causally to action, it is not always relevant to moral responsibility. In fact, there is a difference between causing an agent to φ and depriving her of alternatives regarding φ. When these two things come apart, we see that alternatives are not required for responsibility.

This is an appealing but controversial lesson. Some have argued that it is metaphysically impossible to make an action unavoidable without causally

[8] See Fischer (1994, 158) for a clear statement of this lesson.

influencing it.[9] What this means is that even if Frankfurt's lesson is correct, PAP remains true. Additionally, some have objected directly to the lesson as being subject to counterexamples.[10] Thus, we do better in assessing Frankfurt's counterexample and its successors independently of the lesson.

This is what PAP defenders have done for fifty years. For this, they have mounted one of two main defences. Each tries to exploit the notion that the agent had alternatives in Frankfurt cases. According to both defences, Frankfurt has failed to provide a case where an agent is responsible for φ and yet cannot refrain from φing. In other words, he has failed to provide a true counterexample to PAP. Let me briefly explain the defences in turn.

The *pure defence* (Steward 2009; Alvarez 2009) consists in arguing that the agent in Frankfurt cases did in fact have alternatives to φing and so is responsible for φing. Here is their reasoning. If you move my fingers on a piano keyboard, I am not playing the piano. If an earthquake makes you kick and bounce, you are not dancing the Charleston, even if your body follows the correct footwork. Similarly, if Jones is moved by Black in the counterfactual scenario, Jones is not acting; he is simply moving.[11] Black might instead be merely *influencing* Jones; but in such a case, Jones had the ability to resist. In any case, Jones had an alternative, namely refraining from action (despite perhaps being moved by Black).

---

[9] Unless perhaps if determinism is true, which causes further dialectical complications; see e.g. Widerker and Goetz (2013).

[10] See van Inwagen (1997) and Widerker (2000), mentioning an unpublished example of Ginet.

[11] Hart, citing Austin, makes a similar point about a different case (1968, 144).

The *fine-grained defence* (Ginet 1996; van Inwagen 1997; Capes and Swenson 2017) consists in arguing that Jones is not responsible for φing *simpliciter*. Rather, he is responsible for φing in a certain way, e.g. by executing a decision to φ or by φing at a different moment.[12] True, Jones could not have refrained from committing a crime, but he could have refrained from *committing a crime on his own* or he could have refrained from *committing a crime at that precise instant*.

3.2 The Flicker of Freedom Feud

Are these responses adequate? John Fischer (1994), one of the most ardent critics of PAP, was amongst the firsts to recognise that Jones did in fact have alternatives in Frankfurt cases. By forcing Black to show his hands, Jones could have committed a crime at a different time (Ginet 1996), he could have committed a crime under the influence of Black (Capes and Swenson 2017), or something of the sort (van Inwagen 1997). Fischer dubbed these alternatives *flickers of freedom*. But Fischer thought that even if Jones had these alternatives in Frankfurt cases, Frankfurt had shown that PAP was false.

Why? After all, strictly speaking, for PAP to be true, all that needs to be the case is that there is no situation where someone is both responsible and lacks alternatives, whether robust or not. So, flickers of freedom *do* disqualify putative counterexamples to PAP. However, this is short-sighted for two related reasons. The first is metaphysical: if PAP is true, it is in virtue of the

[12] This is Ginet's (1996) preferred solution. Suppose that Jones commits a crime at *t1*. If Jones had not committed a crime at *t1*, Black would have intervened at *t2* to make sure that Jones did it by *t3*. Ginet claims that Jones is not responsible for committing a crime *by t3* (something he could not have avoided), but only for committing a crime *at t1* (something he could have avoided).

fact that one cannot be responsible without *certain relevant* alternatives, not just any. Suppose that PAP is true and consider someone who lacks alternatives to φ. But now, give her one alternative: she could refrain from φing by suffering a heart attack; or perhaps by knocking herself out. These additions cannot be relevant to moral responsibility. Similarly, imagine a person who only has these alternatives and remove them: this subtraction cannot make a difference to their moral responsibility. Thus, the existence of alternatives *simpliciter* is not relevant to moral responsibility; it is the existence of *some relevant alternatives* which is. The second reason is epistemological: our reason for believing PAP is that lacking certain alternatives is incompatible with moral responsibility, not just any. We have no reason for believing that the addition or the subtraction of just any alternative like having a heart attack or knocking oneself out is relevant to moral responsibility.

Fischer said something similar when he argued that an adequate defence of PAP should feature *robust* alternatives, alternatives which, in his terms, have enough 'oomph'. Less metaphorically, a robust alternative is an alternative which *grounds* the responsibility of an agent (Fischer 1994, 140) or *explains* her blameworthiness (Pereboom 2014a, 10). If we can have a scenario where an agent lacks robust alternatives and is nevertheless responsible, then we should abandon PAP.[13]

[13] If we prefer the metaphysical reason that I gave in the previous paragraph, we will claim that we should abandon PAP because it is false; if we prefer the epistemological reasons, we will claim that we should abandon PAP because have no reason to believe it.

Back to Frankfurt cases: do the pure and the fine-grained defences show that Frankfurt cases feature robust alternatives? Their proponents generally say 'yes' while Fischer, Pereboom, and others says 'no'. This is the real metaphysical controversy, which I call 'the Flicker of Freedom Feud'. This is the controversy which I will argue in the next section we can avoid.

But let's dabble in it for a moment, starting with the fine-grained defence. Jones is responsible for what he could have avoided, namely *committing a crime on his own* or *committing a crime at this time*. He could have avoided committing a crime on his own or at this time by committing a crime under Black's influence.[14] Is this alternative robust, namely does it ground or explain Jones' responsibility or blameworthiness? Perhaps. There is certainly a case for the view that it is *more* robust than the heart attack alternative, for instance.

But is this enough? The answer depends on the nature of robustness. Again, the *concept* of robustness is such that a robust alternative explains or grounds responsibility or blameworthiness. But there are different *conceptions* of robustness. According to Fischer's and Pereboom's conception, Jones lacks robust alternatives. Very roughly, they claim that a robust alternative must involve voluntary behaviour salient to the agent.[15] (We should add that a robust alternative must include the manifestation of an ability: throwing a basketball into a net from fifteen metres might not be metaphysically impossible, but it is not a robust possibility because I lack the ability to do this.

[14] We might think that independently of robustness, this solution fails because people are neither responsible for doing such and such at such and such a time; nor for *doing-something-on-their-own*; see Steward (2009, 69).

[15] See Pereboom (2014a, 13) for a more detailed definition of robustness.

Such a possibility is, at least usually, morally irrelevant.) Fischer's and Pereboom's view is that, in Frankfurt cases, the action of 'committing a crime as the result of Black's intervention' is either not salient to Jones or involuntary.

Now, the fine-grained defender might respond that what is salient and voluntary is the *choice to refrain from committing a crime*. However, Fischer (2012), Hunt (2007), and Pereboom (2014a) have come up with refined Frankfurt cases where Black will intervene earlier, i.e. if he predicts that Jones is about to decide to do otherwise, to deliberate about what to do, or something of the sort. It seems that the only alternative open to Jones is then to *try to decide to refrain* from committing a crime or something of the sort. But this alternative does not seem robust. It seems insufficient to explain Jones' responsibility or blameworthiness—at least, it seems insufficient to explain why Jones is as blameworthy as if Black was absent. This might lead the fine-grained defender to argue for a different conception of robustness. Welcome to the land of metaphysical controversies.

Briefly turning to the pure defence. Remember, the pure-defender claims that Jones has alternatives because if he is moved by Black, he is not *acting* but simply *moving*. As such, Jones *refrains from committing a crime*. But refraining from committing a crime seems like a robust alternative. Thus, the pure-defender has shown that Frankfurt cases fail because they fail to see that Jones has robust alternatives.

There are two typical ways to reject the pure defence. The first is to cast doubt on the claim that if Black causes Jones to commit a crime, Jones is moving, not acting. Couldn't Black instil in Jones the irresistible desire to commit a crime? The general question this raises is whether irresistible desires are metaphysically possible.[16] So it is, after all, controversial whether the pure defence manages to identify robust alternatives for all Frankfurt cases. Again, we are trapped in metaphysical controversies.

The second response is Fischer's and Pereboom's: even if the pure defender is right that Jones could have refrained from committing a crime, it is not a robust alternative because this alternative does not feature voluntary behaviour. Jones does not voluntarily refrain from committing a crime; it is Black that makes him refrain by, *qua* the pure defence, making him move. To make things clearer, Fischer introduces 'Fischer-style cases':

> In this kind of variant on the Frankfurt-type scenario, if it becomes clear to the counterfactual intervener—he is an excellent judge of such things—that the agent is going to decide to do something else, he will use his machine to destroy the agent's brain and thus kill him instantly! (Fischer 2003, 242)

Therefore, says Fischer, we see that PAP is false. For we have a clear case where the agent lacks robust alternatives and yet is responsible.

Further responses have been given but let me press the 'pause' button on our discussion of metaphysical controversies surrounding robustness. Whether the reader's sympathies lie with the pure defender, the fine-grained defender,

[16] Why would irresistible desires be impossible? Perhaps because desires are only manifested in actions and that actions are always avoidable (Steward 2012).

or with the PAP-foe does not matter at this point. My goals in this section were the following: first, introduce the reader to Frankfurt's counterexamples; second, show that the success of his counterexamples depends on metaphysical controversies; and third, demonstrate that although the idea that we should focus on robust alternatives is important, it fails to settle the debate between the friends and the foes of PAP. I now turn to my own approach to the question, which distances itself from these debates.

4. Beyond the Feud: Robustness Includes Eligibility

It would seem, then, that after fifty years the jury is still out on whether Frankfurt cases disprove PAP. For it is difficult to determine whether the agent in Frankfurt cases has robust alternatives, i.e. alternatives that explain why the agent is responsible or blameworthy for her conduct.

Yet we can still make progress independently of who is right about this issue. We can make progress by asking, again, what reason we have to accept PAP. In § 2, I argued that our reason for believing PAP is PEAP, i.e. that lacking eligible alternatives is a good excuse. Now, this principle is easier to assess because there are cases where an agent uncontroversially lacks eligible alternatives and yet seems blameworthy.

The general thought is then that *robustness* of alternatives, as Fischer and Pereboom understand it, is insufficient. According to their conception, it is (mainly) a metaphysical and epistemic property. (A robust alternative includes

voluntary[17] behaviour salient to the agent.) We need to add an ethical—and thus value infused—dimension: *eligibility*. A robust alternative should be an eligible one. For otherwise, it does not adequately explain the agent's responsibility or blameworthiness. An alternative that is too costly for an agent should be irrelevant to her responsibility, just like an alternative that is inaccessible to her. Thus, PAP defenders cannot respond to Frankfurt cases by invoking the presence of ineligible alternatives in them.

If we can however identify scenarios—whether Frankfurt cases or others—where an agent is blameworthy yet lacks eligible alternatives, this will be a problem for PAP defenders. In particular, it will undermine their reason for believing PAP: that lacking eligible alternatives is a good excuse (PEAP). Consider again Fischer-style cases:

> In this kind of variant on the Frankfurt-type scenario, if it becomes clear to the counterfactual intervener—he is an excellent judge of such things—that the agent is going to decide to do something else [shooting Smith], he will use his machine to destroy the agent's brain and thus kill him instantly! This rather ghoulish Fischer-type example is just as potent in calling into question PAP as is the standard sort of Frankfurt-type example. […] Of course, in the Fischer-type case the agent in some sense has access to another path—the path on which he is instantaneously killed (not a very promising path, by the way!) But this path does not contain any voluntary behaviour by the agent or anything with sufficient 'oomph' to ground moral responsibility. (Fischer 2003, 242)

[17] It is crucial to note that, for Fischer and Pereboom, voluntary behaviour is *not* an ethical concept (cf. Hyman 2015, chap. 4). But if voluntariness *is* an ethical concept in the way Hyman thinks (*contra* Fischer and Pereboom), we do not need to amend Fischer and Pereboom's view on robustness.

I included the full quote so that we could assess Fischer's reasoning and notice his bracketed comment.

His reasoning is partly mistaken (for a reason we did not yet encounter because I pressed the 'pause' button). If PAP implies that to be responsible, one needs to be able to *perform another action* (as opposed to merely refraining from performing the first), then a Fischer-style case is a counterexample to PAP. But as Alvarez (2009) correctly notes, PAP only implies that we need to have an alternative in order to be responsible, whether this alternative consists in performing an action or omitting to perform one. She writes:

> It is true that Black's presence in the background prevents the possibility of a mental act of deciding, or choosing, not to shoot Smith. But, […] in Fischer-style cases at the moment when Jones shoots Smith […] it is up to Jones whether he shoots Smith or not: if he decides to shoot he shoots him, and if he does not decide to shoot, he does not shoot him.[…] Fischer-type cases are cases where the agent has access to pathways with 'relevantly different contents'—for shooting Smith is as different from not shooting Smith as the contents of two pathways can be […]. (Alvarez 2009, 69–70)

Alvarez seems right about the robustness of Jones' alternatives.

Yet, there is more to Fischer-style cases than this. Fischer notes in passing that Jones' alternative—being shot by Black—is 'not very promising' but he does not make much of this comment. Yet his comment raises the crucial question of the eligibility of Jones' alternatives. Something like Fischer-style cases are problematic for PAP, but not because the alternative does not feature a voluntary action as Fischer thinks. Fischer-style cases are problematic

for PAP because they suggest that someone can be responsible in the absence of eligible alternatives and thus that PEAP is false. Differently put, Fischer cases suggest that lacking eligible alternatives is *not* a good excuse and thus that we have no reason to believe PAP. This is where Fischer can be shown to have the upper hand against the PAP defenders.

Are Fischer-style cases counterexamples to PEAP? Well, we might think that it is reasonable to demand that Jones sacrifices himself instead of shooting Smith. After all, the law does not recognise duress as a defence for murder.[18] If so, then he has an eligible alternative and so PEAP is safe.

However, it is easy to amend Fischer-style cases to cause troubles for PEAP. Jones' action can be chosen as a minor offence, for instance, stealing Smith's bag. If he refrains from stealing, he will be shot, unbeknownst to him. Now, the option of dying so as to refrain from stealing Smith's bag is too costly, considering the threatening (though invisible) presence of Black. And there we have, it seems, a counterexample to PEAP: Jones is blameworthy, but he had no eligible alternatives. If PEAP falls then so does PAP: PAP might still be true, but we then lose our reason for believing that it is. For we can no longer plausibly claim that lacking alternatives is an excuse. And we did not want to say that lacking alternatives was an exemption.

At this point, you might question my account of eligibility and think that whether an agent has eligible alternatives depends her epistemic situation. This

[18] See Abbott v The Queen (1977), AC 755; Howe & Bannister (1987), 2 WLR 568; R v Gotts (1992), 2 AC 412.

is true. But remember that the premise I intended to defend in this section was this:

> (1) Lacking eligible alternatives is not a good excuse, unless the eligibility of an alternative is relativised to the agent's epistemic situation.

I think that the amended Fischer case establishes this. And it establishes this independently of the metaphysical controversies about robustness that I presented in the last section. Let me quickly show how.

Consider first the pure defence. In Frankfurt cases, the agent might have alternatives for φ and it's contentious whether the agent has robust alternatives. But in Fischer-style cases with the amendment I introduced (featuring theft), it is uncontroversial that Jones lacks eligible alternatives and thus robust alternatives. At least, this is clear so long as we do not try to reinterpret eligibility. And if that's true and we want to maintain our intuitive judgement that Jones is blameworthy, we will have to reject the claim that lacking eligible alternatives is a good excuse. We perhaps thought for a while (ten years) with Alvarez and Steward that we could always appeal to robust alternatives when the agent was blameworthy, but now we realise that sometimes this is false, i.e. when the alternatives of an agent are all ineligible. Jones has no excuse and no eligible alternatives.

Consider next the fine-grained defence. In Frankfurt cases, Jones might have alternatives for the more specific action, e.g. acting on his own. And this might explain why Jones is as blameworthy as he would have been in the absence of Black. But again, consider amended Fischer-style cases. This time,

the agent lacks eligible alternatives. If he does anything differently, he will be shot. For instance, were he to *not* act on his own, or to *not* act at *t1*, he would be shot. This is not an eligible alternative; and, again, Jones has no excuse.

If PAP defenders accept that their reason for believing PAP is that lacking eligible alternatives is an excuse (1), they should accept that their strategy fails. It becomes impossible to make sure that seemingly blameworthy agents have eligible alternatives, whether by arguing that some counterfactual interventions would change the nature of the agent's conduct (the pure defence) or by arguing that the agent is blameworthy for something more specific than the initial action (the fine-grained defence). At least, we will have shown this once we have established that redefining eligibility in epistemic terms is of no help to the PAP defenders.

5. Relative Eligibility

We now consider the view that eligibility should be understood as relative to the agent's epistemic situation. If it is, then it seems that we can save PEAP. We can say that lacking an eligible alternative is an excuse, when an eligible alternative is one whose cost is such that it is reasonable for the agent to consider or choose it, *considering what he knows or should reasonably believe in the circumstances*.

If this is the correct understanding of PEAP—call it 'relativised-PEAP'—then we could say the following. In the revised Fischer-style case (with bag-stealing), whether Jones is blameworthy depends on whether he knew or reasonably believed that he would be killed if he did not steal. If he knew, he

did not have an eligible alternative (from his point of view, or from ours[19]), and should not be blamed. Whereas if he did not know he did have an eligible alternative (from his point of view, though not from ours), and should be blamed. The principle that he should be blamed only if he had an eligible alternative is therefore saved. PEAP stands.

We can make this important point differently. Keeping an invariant (non-relativised) version of eligibility, we can amend PEAP as follows:

> **PEAP*:** One should be excused for one's conduct if *one knows or reasonably believes that* one lacks eligible alternatives, viz. alternatives whose cost is not too high to reasonably consider or choose.

I have a slight preference for this second way,[20] but it does not matter much for our purpose. Everything I say about one applies to the other.

Is this epistemic move a problem for my general argument? No, because:

(1) If eligibility is relativised to the agent's epistemic situation, that *lacking eligible alternatives is a good excuse* is not a good reason for believing PAP.

---

[19] Where 'our' point of view is the point of view of the omniscient (or at least well-informed) third party.

[20] One advantage of PEAP* over relativised-PEAP is that it allows us to distinguish more clearly two ways in which excuses are relative to an agent's epistemic situation. First, the excuse is relative in the sense that it depends on whether the agent believes or should believe that she has an eligible alternative, *qua* PEAP*. Second, the excuse is relative in the sense that it depends on what the agent values—for instance, it depends on her religious and political beliefs. For what the agent values can impact what counts as eligible to her. Relativised-PEAP, but not PEAP*, lumps these distinctions together.

Differently put, PEAP* does not constitute a good reason for believing PAP. This is not because it is false, but because it provides no support for PAP.

Why? PEAP* states an *epistemic* excusing condition (and thus an epistemic requirement on moral responsibility). But PAP does *not* state an epistemic requirement on moral responsibility; it states a *metaphysical* requirement on moral responsibility. Yet, a principle about an epistemic requirement is not a good reason to endorse a principle about a metaphysical requirement. Therefore, (3) is true.

To illustrate this, consider how differently PAP and PEAP* stand to the free will debate. If PAP is true and if we all lack alternatives at all times, as some free will sceptics maintain, then we are never responsible. This is true *regardless of what we believe or should believe about the matter*. By contrast, if PEAP* is true, no such thing follows. It only follows that, in worlds where free will exists, some but not all subjects are excused; and the same goes for worlds where free will does not exist. So, unlike PAP, PEAP* could not be part of a sceptical or incompatibilist argument about responsibility.

Leaving this illustration aside, it is clear that PEAP but not PEAP* could be a reason for believing PAP. In fact, they stand in a very different relation to PAP. Like PEAP, PEAP* excuses some people that PAP does not excuse. But unlike PEAP, PEAP* fails to excuse some people which PAP excuses.[21] Put another way, PEAP extends the domain of the excused that PAP sets; PEAP*

[21] To be precise, PAP is not about excuses, but we need not complicate the situation by adding a new principle.

is about a completely different set of excused agents. Thus, PEAP* is not a reason for believing PAP. The two principles are related in name alone.[22]

(PEAP* does not support PAP; but it might support PAP*, Frankfurt's (1969) and Strawson's (1974) principle according to which one is only responsible for actions that she performed *because* she lacked alternatives. For, if Jones acted because he lacked alternatives, he must at least believe[23] that this is true. This raises a further question: should we simply accept PAP* or should we also grant PEAP*? If both principles are true, then an agent is excused if she acted *with* the knowledge that she lacked eligible alternatives, even if she did not act *on* that knowledge. According to PEAP*, a driver is excused for bumping into someone else's car if he was aware that there was no eligible alternative, such as braking, even if his reason for bumping is malevolent. If PAP* but not PEAP* is true, this is not necessarily the case.[24])

Let's recap. (1) Our reason for believing PAP is that lacking eligible alternatives is a good excuse. (2) Yet, on close examination, lacking eligible alternatives is *not* a good excuse, unless the eligibility of an alternative is

---

[22] Some might be tempted to look for mixed principles as support for PAP. Two come in mind. The first principle states that lacking eligible *or* epistemically accessible alternatives is an excuse. This principle is false because it is false that merely lacking eligible alternatives is an excuse, as I have argued. The second principle states that lacking eligible *and* epistemically accessible alternatives is an excuse. This principle might be true, but it is not a reason for believing PAP. For there are cases that PAP excuses but which it does not excuse, i.e. cases where an agent lacks all alternatives unbeknownst to her. (Consider again the free will problem.)

[23] Hyman (2015, chap. 6) argues convincingly that for *S* to act because of *p*, *S* must know that *p*. See also Williamson (2000).

[24] Frankfurt (1969) prefers the second avenue. For he argues that a coerced agent ($Jones_3$) who acts for his own reason is responsible for the action that he was coerced to perform.

relativised to the agent's epistemic situation. (3) And, in this latter case, that *lacking eligible alternatives is a good excuse* provides no support for PAP and thus is not a good reason for believing it. This strategy allows us to bypass the metaphysical debates which gripped our attention in §3 and gripped the attention of many philosophers for five decades. Or, perhaps, more charitably, it allows us to show that robustness has been misunderstood (for two decades)—it should include eligibility as well.

I do not expect PAP defenders to surrender. They might claim that PAP is irreducibly intuitive and that I was wrong in the first place to ask *why* one should believe in it. In the next section, I explore a different concession that they might try to negotiate.

6. The Price of PAP

Faced with my argument, the PAP defender has one option, which is to make an ethical concession. For this, she must take the following line. She can accept almost everything that I have said so far except for my judgement about the Fischer-style case. This forces her to conclude that, in this case—where Jones would, unbeknownst to him, be shot unless he stole a bag—he is morally lucky. For Jones has a good excuse for his crime: he had no eligible alternative. She can continue and argue that Jones is responsible for something else, like harbouring some reprehensible desires. But then she should expect that I modify the Fischer-style case again. I can have Black preside over Jones' desires—he would shoot him if Jones had a good will towards the bag owner.

There, the PAP defender has, at least in this revised version of the case, to excuse Jones entirely.

This seems like an exorbitant and unusual concession. It is exorbitant because we can imagine a plethora of counter-intuitive cases. Imagine two thieves working for the Mafia. The first will be shot if she does not fulfil a certain job—unbeknownst to her. The second will *not* be shot if he fails, owing to the fact that he reminds the Mafia boss of his lost son. Suppose that both thieves execute their order. If we grant PEAP (in its invariant version) and PAP, we will be committed to claiming that the second thief, but not the first, is blameworthy.

Notwithstanding the number of counter-intuitive cases, this price seems exorbitant because it introduces an unusual form of moral luck. PEAP introduces a form of moral luck because it predicts that some agents are off the hook simply in virtue of circumstances that they were unaware of and which they have no control over (e.g. Black's existence, the Mafia boss's presence). Therefore, two identical agents with the same flawed motives and conduct are not necessarily equally responsible. This is typical moral luck, the same that distinguishes a driver who is guilty of dooring a cyclist (by accident) from a driver who is merely guilty of opening her car's door recklessly.[25] However, PEAP introduces an *unusual* form of moral luck (different from the moral luck of the two drivers) because, here, e.g. in the case of the two thieves, not only are the conduct and motives identical, but the *consequences* of the

[25] Responsibility that fluctuates according to the consequences of one's action is the standard case of moral luck. See Williams (1981) and Nagel (2012).

conduct are identical. The only difference between the two cases is the modal properties of the agents—one had eligible alternatives, the other one did not.

This concession is especially unusual considering the stance that PAP defenders have taken in the past fifty years. They have almost always assumed that they could respond to Frankfurt cases at no moral cost.[26] In particular, they thought that they could agree with Frankfurt's verdict about the blameworthiness of individuals like Jones and challenge Frankfurt's explanation of that verdict.[27] But this middle ground seems closed. The PAP defenders cannot have their cake and eat it, namely PAP and all the intuitive judgements about all the Jones of Frankfurt cases and Fischer-style cases.

That's not to say that the concession is completely unmotivated. For instance, the PAP defender could marshal the following argument, inspired from David Widerker (2000) and David Copp (2008):[28]

> *S* is only blameworthy for φing if *S* has the obligation to refrain from φing.
>
> *S* only has the obligation to refrain from φing if *S* can refrain from φing. (This is a variation on "ought' implies 'can")

---

[26] One notable exception is Levy (2008), who argues that we should not rely on our intuition about Frankfurt cases because we have an internalist prejudice: we believe that Jones' capacities cannot be affected by Black's mere presence.

[27] They might refuse to attribute responsibility to Jones in deterministic Frankfurt cases. This brings up the 'dilemma defence' (see e.g. Pereboom 2014a for a recent discussion). The dilemma defence tries to force a dilemma on Frankfurt: 'is determinism true or false in your cases? If it is true, you beg the question against incompatibilists; if it is false, you cannot make sure that Jones lacks alternatives.' But the defence faces Fischer's flicker response.

[28] Both of these principles are controversial. See Saka (2000) for arguments against the first and Scanlon (2008) for arguments against the second.

> *Therefore, S* is only blameworthy for φing if *S* can refrain from φing.

This gives us reasons to endorse PEAP and PAP, as long as 'S can refrain' implies that refraining is an eligible alternative, in the given circumstances. (It seems that our reason for believing that "ought' implies 'can" is precisely that the existence of an obligation implies the existence of an *eligible alternative*.)

The PAP defender might add that when we do something which we think is wrong but then later realise that we could not have done otherwise, we are typically relieved. We tell ourselves that we did the best we could in the circumstances, even if this amounts to something quite nasty. We might take this as a depressing lesson about our flaws. But in a way, we think we are off the hook. Now, if this natural tendency tracks anything about whether we *should* be excused in such circumstances, it supports PEAP.

Whether we decide in the end to travel with Fischer, Pereboom, and Frankfurt, on the one hand, or with Copp and Widerker, on the other hand, might depend on broader issues about blame and blameworthiness. If to blame is to communicate to someone that our relationship with them has been altered (see e.g. Scanlon 2008), then it makes sense to go with the first triad. This is because relationship amendments with someone can be appropriate even in the absence of an obligation infringement. Sometimes an action which is in conformity with someone's obligations reflects poorly on her. It is hard to see why someone's relationship with the first thief would be more affected than her relationship with the second simply in virtue of the fact that the second but not the first had eligible alternatives. If, however, to blame is

(partly) to desire that the agent have acted differently (see e.g. Sher 2005), then it makes sense to claim that absence of eligible alternatives is an excuse (where eligibility is invariant). For we do not and often should not wish that we had chosen ineligible alternatives. If, finally, to blame is to express certain negative emotions by the way of reactive attitudes (see e.g. P.F. Strawson 1974; Wallace 1994), then the question is open. It depends on what reactive attitudes are and what they respond to: are they a retributive response to obligation infringements or a communicative response to poor quality of will? In the second case, we should go with Frankfurt; in the first, we might want to go with Copp (although we do not have to). These remarks are quite impressionistic, but they indicate to the PAP enthusiast where to go next.

7. Conclusion

I have revisited the debate about Frankfurt's counterexamples to the principle of alternate possibilities by asking why the principle is true—if in fact it is. My answer was that lacking eligible alternatives is a good excuse. But, I argued, this can only be plausible if eligibility depends on the agent's epistemic situation. Yet, if this is the line we take, we need to concede that we have lost our reason for believing PAP. For PAP is not an epistemic, but a metaphysical principle. The two traditional responses to Frankfurt cases, the pure defence and the fine-grained defence, cannot help us avoid these counter-intuitive consequences. In fact, the motivation behind my approach was precisely to avoid the metaphysically murky grounds on which these defences stood.

If, on the contrary, PAP must be saved, it seems that we must do something that virtually no PAP defender has done in the past fifty years.[29] We need to accept that the existence of an invisible and non-intervening counterfactual agent can make a difference for moral responsibility and blameworthiness. This means rejecting Frankfurt's lesson: blameworthiness for an action does not merely depend on its causes. It also means admitting an unusual form of moral luck, one that it would be important to assess independently. Whether exorbitant or not, this is the price of PAP. For myself, I prefer not to pay it.

[29] As mentioned in note 26, Levy (2008) is an exception.

BIBLIOGRAPHY

Alston, William P. 1989. *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca; London: Cornell University Press.

Alvarez, Maria. 2009. 'Actions, Thought-Experiments and the "Principle of Alternate Possibilities"'. *Australasian Journal of Philosophy* 87 (1): 61–81. https://doi.org/10.1080/00048400802215505.

———. 2016 'Reasons for Action: Justification, Motivation, Explanation.' *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 22, 2019. https://plato.stanford.edu/entries/reasons-just-vs-expl/

Annas, Julia, C. C. W. Taylor, and D. N. Sedley. 1983. *Oxford Studies in Ancient Philosophy*. Oxford: Clarendon Press.

Anscombe, G. E. M. 1963. *Intention*. 2nd ed. Library of Philosophy and Logic. Oxford: Basil Blackwell.

Austin, J.L. 1956. 'A Plea for Excuses: The Presidential Address'. *Proceedings of the Aristotelian Society* 57: 1–30.

———. 1962. *Sense and Sensibilia*. Oxford: Clarendon Press.

Ayer, A.J. 1963. *The Concept of a Person: And Other Essays*. London: Macmillan.

———. 1954. 'Freedom and Necessity'. In *Philosophical Essays*. Oxford: Oxford University Press.

Bailey, Cyril. 1926. *Epicurus: The Extant Remains*. Oxford: Clarendon Press.

Baumann, Peter. 2012. 'Knowledge and Dogmatism'. *The Philosophical Quarterly* 63 (250): 1–19.

Beck, Lewis White. 1951. 'Kant: Foundations of the Metaphysics of Morals. What Is Enlightenment? And a Passage from the Metaphysics of Morals'.

Bergamaschi Ganapini, Marianna. 2016. 'Why We Can Still Believe the Error Theory'. *International Journal of Philosophical Studies* 24 (4): 523–36.

Blackmore, Susan. 2013. 'Living without Free Will'. In *Exploring the Illusion of Free Will*, edited by Gregg D. Caruso, 161–76. Lanham, MD: Lexington Books.

Bobzien, Susan. 2000. 'Did Epicurus Discover the Free-Will Problem?' *Oxford Studies in Ancient Philosophy* 19: 287–337.

Boghossian, Paul A. 1990. 'The Status of Content'. *The Philosophical Review* 99 (2): 157–84.

Brennan, Tad. 2001. 'Fate and Free Will in Stoicism'.

Broome, John. 2013. *Rationality Through Reasoning*. Oxford: Wiley-Blackwell.

Burnyeat, Myles. 1984. 'The Sceptic in His Place and Time'. *Philosophy in History*, 225–54.

Byrd, Jeremy. 2010. 'Agnosticism about Moral Responsibility'. *Canadian Journal of Philosophy* 40 (3): 411–32. https://doi.org/10.1080/00455091.2010.10716729.

Campbell, Charles Arthur. 2013. *On Selfhood and Godhood*. Routledge.

Capes, Justin A., and Philip Swenson. 2017. 'Frankfurt Cases: The Fine-Grained Response Revisited'. *Philosophical Studies* 174 (4): 967–81. https://doi.org/10.1007/s11098-016-0726-z.

Carr, Jennifer Rose. 2015. 'Don't Stop Believing'. *Canadian Journal of Philosophy* 45 (5–6): 744–66. https://doi.org/10.1080/00455091.2015.1123454.

Caruso, Gregg D. 2013. *Exploring the Illusion of Free Will and Moral Responsibility*. Lanham, MD: Lexington Books.

———. 2018. 'Skepticism About Moral Responsibility'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/skepticism-moral-responsibility/.

Caruso, Gregg D., and Stephen G. Morris. 2016. 'Compatibilism and Retributivist Desert Moral Responsibility: On What Is of Central Philosophical and Practical Importance'. *Erkenntnis* 82 (4): 1–19. https://doi.org/10.1007/s10670-016-9846-2.

Castagnoli, Luca. 2007. 'Everything Is True, Everything Is False: Self-Refutation Arguments from Democritus to Augustine.' *Antiquorum Philosophia*. 1: 11–74.

———. 2010. *Ancient Self-Refutation: The Logic and History of the Self-Refutation Argument from Democritus to Augustine*. Cambridge, UK: Cambridge University Press.

Castañeda, Hector-Neri. 1975. *Thinking and Doing: The Philosophical Foundations of Institutions*. Philosophical Studies Series ; v. 7. Dordrecht: Reidel.

Chevarie-Cossette, Simon-Pierre. 2017. 'Responsabilité morale (A)'. In *L'Encyclopédie Philosophique*. https://encyclo-philo.fr/responsabilite-morale-a/.

———. 2019a. 'Self-Defeating Beliefs and Misleading Reasons'. *International Journal of Philosophical Studies* 27 (1): 57–72. https://doi.org/10.1080/09672559.2018.1548636.

———. Forthcoming. 'Knowing about Responsibility: A Trilemma'. *American Philosophical Quarterly*.

———. 2019b. 'Is Free Will Scepticism Self-Defeating?' *European Journal of Philosophy* 15(2): 55–78.

———. Conditional acceptance. 'Sceptical Deliberations' *Pacific Philosophical Quarterly*.

Chisholm, Roderick. 1964. *Human Freedom and the Self*. Kansas City.

———. 1977. *Theory of Knowledge (2nd Edition)*. Prentice-Hall.

Christensen, David. 2009. 'Disagreement as Evidence: The Epistemology of Controversy'. *Philosophy Compass* 4 (5): 756–67.

Churchland, Patricia Smith. 1981. 'Is Determinism Self-Refuting?' *Mind* 90 (357): 99–101.

Clarke, Randolph K. 1992. 'Deliberation and Beliefs About One's Abilities'. *Pacific Philosophical Quarterly* 73 (2): 101–13.

———. 2014. *Omissions: Agency, Metaphysics, and Responsibility*. New York, NY: Oxford University Press.

Coffman, E.J. 2016. 'Incompatibilist Commitment and Moral Self-Knowledge: The Epistemology of Libertarianism'. *Philosophical Issues* 26 (1): 78–98. https://doi.org/10.1111/phis.12066.

Coffman, E. J., and Ted A. Warfield. 2005. 'Deliberation and Metaphysical Freedom'. *Midwest Studies In Philosophy* 29 (1): 25–44.

Cohen, G.A. 1983. "The Structure of Proletarian Unfreedom." *Philosophy & Public Affairs* 12 (1): 3–33.

Cohen, Yishai. 2018. 'Deliberating in the Presence of Manipulation'. *Canadian Journal of Philosophy* 48 (1): 85–105.

Copp, David. 2008. "'Ought" Implies "Can" and the Derivation of the Principle of Alternate Possibilities'. *Analysis* 68 (1): 67–75. https://doi.org/10.1093/analys/68.1.67.

Côté-Bouchard, Charles. 2017. *Belief's Own Metaethics? A Case against Epistemic Normativity*. KCL doctoral thesis.

Cowan, J. L. 1969. 'Deliberation and Determinism'. *American Philosophical Quarterly* 6 (1): 53–61.

Cuneo, Terrence. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

Cyr, Taylor W. 2016. "The Parallel Manipulation Argument." *Ethics* 126 (4): 1075–89.

Davidson, Donald. 1985. 'Replies to Essays X-XI'. In *Essays on Davidson: Actions and Events*, edited by Bruce Vermazen and Merrill B. Hintikka, 242--252. Oxford: Oxford University Press.

Dennett, Daniel C. 2015. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

Denyer, Nicholas. 1981. *Time, Action & Necessity: A Proof of Free Will*. Duckworth.

Descartes, René. 1724. *Les Méditations Métaphysiques*. Vol. 1. chez Robert-Marc d'Espilly.

Diderot, Denis. 1875. 'Œuvres Complètes, Ed'. *J. Assezat and M. Tourneux, Paris, Garnier Frères* 7.

Dorr, Cian. 2016. 'Against Counterfactual Miracles'. *The Philosophical Review* 125 (2): 241–86.

Double, Richard. 1996. *Metaphilosophy and Free Will*. Oxford: Oxford University Press.

Dretske, Fred I. 1970. 'Epistemic Operators'. *Journal of Philosophy* 67 (24): 1007–1023.

———. 2005. 'The Case Against Closure'. In *Contemporary Debates in Epistemology*, edited by M. Steup and Earnest Sosa, 13–25. Malden, Ma: Blackwell.

Duff, Antony. 2007. *Answering for Crime: Responsibility and Liability in the Criminal Law*. Legal Theory Today. Oxford; Portland, Or: Hart.

Dworkin, Ronald. 1996. 'Objectivity and Truth: You'd Better Believe It'. *Philosophy & Public Affairs* 25 (2): 87–139.

———. 2011. *Justice for Hedgehogs*. Harvard University Press.

———. 2013. 'Replies'. *Analysis* 73 (1): 139–46. https://doi.org/10.1093/analys/ans156.

Deery, Oisín, and Eddy Nahmias. 2016. "Defeating Manipulation Arguments: Interventionist Causation and Compatibilist Sourcehood." *Philosophical Studies*, 174 (5): 1255-1276.

Demetriou, Kristin. 2010. "The Soft-Line Solution to Pereboom's Four-Case Argument." *Australasian Journal of Philosophy* 88 (4): 595–617.

'Egypt zoo accused of painting donkey to look like a zebra', 2018. *BBC*, July 26. https://www.bbc.co.uk/news/world-africa-44968509

Feldman, Richard. 2004a. '12 Freedom and Contextualism'. In *Freedom and Determinism*, edited by and D. Shier M. O. 'Rourke J. K. Campbell, 255. MIT.

———. 2004b. 'Comments on DeRose's "Single Scoreboard Semantics"'. *Philosophical Studies* 119 (1–2): 23–33. https://doi.org/10.1023/B:PHIL.0000029348.33904.fa.

Fischer, John Martin. 1994a. *The Metaphysics of Free Will: An Essay on Control.* Blackwell.

———. 1994b. *The Metaphysics of Free Will: An Essay on Control*. Blackwell.

———. 2000. "Responsibility, history and manipulation." *The Journal of Ethics*, 4 (4): 385–391.

———. 2003. 'Responsibility and Agent-Causation'. In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*, edited by David Widerker and Michael McKenna. Ashgate.

———. 2005. 'The Cards That Are Dealt You'. *The Journal of Ethics* 10 (1–2): 107–129.

———. 2006. *My Way: Essays on Moral Responsibility*. Oxford: Oxford University Press.

———. 2011. "The Zygote Argument Remixed." *Analysis* 71 (2): 267–72.

———. 2012. 'Semicompatibilism and Its Rivals'. *The Journal of Ethics* 16 (2): 117–143. https://doi.org/10.1007/s10892-012-9123-9.

———. 2016. "How Do Manipulation Arguments Work?", *The Journal of Ethics* 20 (1–3): 47–67.

Fischer, John Martin, and Garrett Pendergraft. 2013. 'Does the Consequence Argument Beg the Question?' *Philosophical Studies* 166 (3): 575–95.

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

Forcehimes, Andrew T., and Robert B. Talisse. 2016. 'Belief and the Error Theory'. *Ethical Theory and Moral Practice* 19 (4): 849–56.

Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829–39.

———. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.

———. 2002. "Reply to John Martin Fischer" In: Sarah Buss and Lee Overton, *Contours of Agency: Essays on Themes from Harry Frankfurt*. Cambridge, MA: MIT Press.

Frede, Michael. 2011. *A Free Will: Origins of the Notion in Ancient Thought*. Sather Classical Lectures ; v. 68. Berkeley, CA; London: University of California Press.

Garnett, Michael. 2018. "Coercion: The Wrong and the Bad." *Ethics* 128 (3): 545-73.

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

———. 1996. 'In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing'. *Philosophical Perspectives* 10: 403–17. https://doi.org/10.2307/2216254.

Haas, D. 2013a. "In defense of hard-line replies to the multiple-case manipulation argument." *Philosophical studies* 163 (3): 797-811.

———. 2013b. "Merit, fit, and basic desert." *Philosophical Explorations 16*(2): 226-239.

Haji, Ishtiyaque. 2012. *Reason's Debt to Freedom: Normative Appraisals, Reasons, and Free Will*. Oxford: Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199899203.001.0001/acprof-9780199899203.

———. 2016. 'Luck's Extended Reach'. *The Journal of Ethics* 20 (1–3): 191–218. https://doi.org/10.1007/s10892-016-9224-y.

Hampshire, Stuart. 1960. *Thought and Action (1959)*. Notre Dame: University of Notre Dame Press.

Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.

Hart, H. L. A. 1962. *Punishment and the Elimination of Responsibility*. L.T. Hobhouse Memorial Trust Lecture; No. 31. London: University of London, Athlone Press.

———.1968. *Punishment and Responsibility*. Oxford: Clarendon Press.

Hawthorne, John. 2001. 'Freedom in Context'. *Philosophical Studies* 104 (1): 63–79. https://doi.org/10.1023/A:1010398805497.

———. 2004. *Knowledge and Lotteries*. Oxford: Clarendon Press.

Hawthorne, John, and Jason Stanley. 2008. 'Knowledge and Action', *Journal of Philosophy*, vol. 105, no. 10, pp. 571–90.

Henden, Edmund. 2010. 'Deliberation Incompatibilism'. *Dialectica* 64 (3): 313–33.

Hieronymi, Pamela. 2005. 'The Wrong Kind of Reason'. *The Journal of Philosophy* 102 (9): 437–57.

Hills, Alison. 2003. 'Defending Double Effect'. *Philosophical Studies* 116 (2): 133–52. https://doi.org/10.1023/B:PHIL.0000005744.80837.65.

Hinman, Lawrence M. 1979. 'How Not to Naturalize Ethics: The Untenability of a Skinnerian Naturalistic Ethic'. *Ethics* 89 (3): 292–97.

Hobart, R. E. 1934. 'Free Will as Involving Determination and Inconceivable Without It'. *Mind* 43 (169): 1–27.

Hoefer, Carl. 2016. 'Causal Determinism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2016. http://plato.stanford.edu/archives/spr2016/entries/determinism-causal/.

Hume, David. 1739. *A Treatise of Human Nature*. London: John Noon.

Hunt, David P. 2007. 'Black the Libertarian'. *Acta Analytica* 22 (1): 3–15. https://doi.org/10.1007/BF02866207.

Hurley, Susan L. 2003. *Justice, Luck, and Knowledge*. Cambridge, MA: Harvard University Press.

Husi, Stan. 2013. 'Why Reasons Skepticism Is Not Self-Defeating'. *European Journal of Philosophy* 21 (3): 424–49.

Holton, Richard. 2013. "From Determinism to Resignation; and How to Stop It." In Andy Clark, Julia Kiverstein, and Tillman Vierkant, *Decomposing the Will*, Oxford: Oxford University Press.

Honoré, Tony. 1999. *Responsibility and Fault*. Hart Publishing.

Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Hyman, John. "Voluntariness and Choice." *Philosophical Quarterly*, 63 (253): 683–708.

———. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

———. 2016. "Voluntariness and Intention." *Jurisprudence*, 7 (3): 692–709.

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

———. 1997. 'Fischer on Moral Responsibility'. Edited by John Martin Fischer. *The Philosophical Quarterly (1950-)* 47 (188): 373–81.

———. 2017. *Thinking about Free Will*. Cambridge, UK: Cambridge University Press.

Istvan, Michael Anthony. 2011. 'Concerning the Resilience of Galen Strawson's Basic Argument'. *Philosophical Studies* 155 (3): 399–420. https://doi.org/10.1007/s11098-010-9578-0.

James, William. 1912. The Will to Believe: And Other Essays in Popular Philosophy. Auckland, New Zealand: Floating Press.

Jordan, James N. 1969. 'Determinism's Dilemma'. *The Review of Metaphysics* 23 (1): 48–66.

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

Kapitan, Tomis. 1986. 'Deliberation and the Presumption of Open Alternatives'. *Southern Journal of Philosophy* 40 (April): 230–51.

———. 1996. 'Modal Principles in the Metaphysics of Free Will'. *Philosophical Perspectives* 10: 419–45.

Kavka, Gregory S. 1983. 'The Toxin Puzzle'. *Analysis* 43 (1): 33–36.

Kearns, Stephen. 2015. 'Free Will Agnosticism'. *Noûs* 49 (2): 235–52. https://doi.org/10.1111/nous.12032.

Kelly, Thomas. 2003. 'Epistemic Rationality as Instrumental Rationality: A Critique'. *Philosophy and Phenomenological Research* 66 (3): 612–40.

Kenny, Anthony. 1978. *Freewill and Responsibility*. Routledge & Kegan Paul Ltd.

———. 1976. *Will, Freedom and Power*. Oxford: Basil Blackwell.

Kiesewetter, Benjamin. 2017. *The Normativity of Rationality*. Oxford: Oxford University Press.

Kim, Jaegwon. 2007. *Physicalism, or Something near Enough*. Princeton: Princeton University Press.

Kolodny, Niko. 2005. 'Why Be Rational?' *Mind* 114 (455): 509–63.

Lamb, James W. 1977. 'On a Proof of Incompatibilism'. *The Philosophical Review* 86 (1): 20. https://doi.org/10.2307/2184160.

Lasonen-Aarnio, Maria. 2010. 'Unreasonable Knowledge'. *Philosophical Perspectives* 24 (1): 1–21.

———. Maria. 2014a. 'Higher-Order Evidence and the Limits of Defeat'. *Philosophy and Phenomenological Research* 88 (2): 314–45. https://doi.org/10.1111/phpr.12090.

———. 2014b. 'The Dogmatism Puzzle'. *Australasian Journal of Philosophy* 92 (3): 417–32.

Lillehammer, Hallvard, and Niklas Möller. 2015. 'We Can Believe the Error Theory'. *Ethical Theory and Moral Practice* 18 (3): 453–59.

Littlejohn, Clayton. 2015. 'Stop Making Sense? On a Puzzle about Rationality'. *Philosophy and Phenomenological Research* 96 (2): 257–72.

List, Christian. 2014. 'Free Will, Determinism, and the Possibility of Doing Otherwise'. *Noûs* 48 (1): 156–78. https://doi.org/10.1111/nous.12019.

Levy, Neil. 2005. 'The Good, the Bad and the Blameworthy'. *Journal of Ethics & Social Philosophy* 1: 1–16.

———. 2008. 'Counterfactual Intervention and Agents' Capacities'. *Journal of Philosophy* 105 (5): 223–239.

———. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.

Levy, Neil, and Michael McKenna. 2009. "Recent Work on Free Will and Moral Responsibility". *Philosophy Compass* 4 (1): 96–133.

Lewis, David. 1976. 'The Paradoxes of Time Travel'. *American Philosophical Quarterly* 13 (2): 145–52. https://doi.org/10.2307/20009616.

———. 1981. 'Are We Free to Break the Laws?' *Theoria* 47 (3): 113–21.

———. 1996. 'Elusive Knowledge'. *Australasian Journal of Philosophy* 74 (4): 549–67. https://doi.org/10.1080/00048409612347521.

———. 2000. *Papers in Ethics and Social Philosophy*. Cambridge, UK: Cambridge University Press.

Lockie, Robert. 2018. Free Will and Epistemology: A Defence of the Transcendental Argument for Freedom. New York: Bloomsbury Academic.

Long, A. A., and D. N. Sedley, eds. 1987. 'Epicureanism'. In *The Hellenistic Philosophers: Volume 2: Greek and Latin Texts with Notes and Bibliography*, 2:18–162. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139165907.003.

Long, Douglas C. 1992. 'The Self-Defeating Character of Skepticism'. *Philosophy and Phenomenological Research* 52 (1): 67–84.

Lucas, J. R., and Hans Sluga. 1967. 'Symposium: Freedom and Prediction'. *Proceedings of the Aristotelian Society, Supplementary Volumes* 41: 163–84.

MacIntyre, A. C. 1957. 'Determinism'. *Mind* 66 (261): 28–41.

Mackie, J. L. 1964. 'Self-Refutation--A Formal Analysis'. *The Philosophical Quarterly* 14 (56): 193–203.

———. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.

Maier, John. 2014. 'The Argument from Moral Responsibility'. *Australasian Journal of Philosophy* 92 (2): 249–67. https://doi.org/10.1080/00048402.2013.808679.

Malcolm, Norman. 1968. 'The Conceivability of Mechanism'. *The Philosophical Review* 77 (1): 45-72.

McKenna, Michael. 2008. "A Hard-Line Reply to Pereboom's Four-Case Manipulation Argument." *Philosophy and Phenomenological Research* 77 (1): 142–59.

———. 2012. "Moral responsibility, manipulation arguments, and history: Assessing the resilience of nonhistorical compatibilism." *The Journal of Ethics* 16 (2): 145-174.

———. 2014. "Resisting the Manipulation Argument: A Hard-Liner Takes It on the Chin." *Philosophy and Phenomenological Research* 89 (2): 467–84.

McKenna, Michael, and D. Justin Coates. 2016. 'Compatibilism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2016. Metaphysics Research Lab, Stanford University.

McMahan Jeff. 2009. "Intention, Permissibility, Terrorism, and War." *Philosophical Perspectives* 23 (1): 345-72.

Mele, Alfred R. 1995. *Autonomous Agents: From Self-Control to Autonomy*. Oxford University Press.

———. 2006. *Free will and luck*. Oxford: Oxford University Press.

———. 2016. "Moral Responsibility: Radical Reversals and Original Designs". *The Journal of Ethics* 20 (1–3): 69–82.

Mickelson, Kristin. 2015. "The Zygote Argument Is Invalid: Now What?" *Philosophical Studies* 172 (11): 2911-29.

Mill, John Stuart. 1979. *An Examination of Sir William Hamilton's Philosophy*. Edited by Alan Ryan and J.M. Robson. University of Toronto Press and Routledge.

Moore, G.E. 1939. 'Proof of an External World'. *Proceedings of the British Academy* 25 (5): 273–300.

Moore, Michael S. 2010. *Act and Crime*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199599509.001.0001.

Moretti, Luca. 2014. 'The Dogmatist, Moore's Proof and Transmission Failure'. *Analysis* 74 (3): 382–89. https://doi.org/10.1093/analys/anu060.

Nadelhoffer, Thomas, and Daniela Goya Tocchetto. 2013. 'The Potential Dark Side of Believing in Free Will (and Related Concepts)'. In *Exploring the Illusion of Free Will and Moral Responsibility*, edited by Gregg D. Caruso. Lexington Books.

Nagel, Thomas. 1986. *The View from Nowhere*. New York; Oxford: Oxford University Press.

———. 2012. *Mortal Questions*. Cambridge: Cambridge University Press. http://ebooks.cambridge.org/ref/id/CBO9781107341050.

Nahmias, Eddy. 2014. 'Is Free Will an Illusion? Confronting Challenges from the Modern Mind Sciences'. In *Moral Psychology, Vol. 4: Freedom and Responsibility*, edited by Walter Sinnott-Armstrong. MIT Press.

Nelkin, Dana Kay. 2004. 'Deliberative Alternatives'. *Philosophical Topics* 32 (1/2): 215–40.

———. 2014. 'Free Will Skepticism and Obligation Skepticism: Comments on Derk Pereboom's Free Will Agency, and Meaning in Life'. *Science, Religion & Culture* 1 (3): 209–17.

Nichols, Shaun. 2007. 'After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes*'. *Philosophical Perspectives* 21 (1): 405–28. https://doi.org/10.1111/j.1520-8583.2007.00131.x.

Nielsen, Karen Margrethe. 2011. 'Deliberation as Inquiry: Aristotle's Alternative to the Presumption of Open Alternatives'. *The Philosophical Review* 120 (3): 383–421.

Nolfi, Kate. 2018. 'Why Only Evidential Considerations Can Justify Belief'. In *Normativity*. Oxford: Oxford University Press.

Nozick, Robert. 1974. *Anarchy, State and Utopia*. Oxford: Basil Blackwell.

———. 1981. *Philosophical Explanations*. Oxford: Clarendon Press.

———. 1997. 'Coercion'. In *Socratic Puzzles*. Cambridge, MA: Harvard University Press.

Olson, Jonas. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Otsuka, Michael. 1998. 'Incompatibilism and the Avoidability of Blame'. *Ethics* 108 (4): 685–701. https://doi.org/10.1086/233847.

Owens, David. 2009. 'Freedom and Practical Judgement'. In *Mental Actions*, edited by Lucy O'Brien and Matthew Soteriou. Oxford: Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199225989.001.0001/acprof-9780199225989-chapter-6.

Page, Carl. 1992. 'On Being False by Self-Refutation'. *Metaphilosophy* 23 (4): 410–26. https://doi.org/10.1111/j.1467-9973.1992.tb00552.x.

Pascal, Blaise. 2004. *Pensées*. Edited by Michel Le Guern. Paris: Gallimard. Vol. 4054. Folio Classique.

Pereboom, Derk. 1995. 'Determinism Al Dente'. *Noûs* 29 (1): 21–45. https://doi.org/10.2307/2215725.

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

———. 2005. Defending hard incompatibilism. *Midwest Studies in Philosophy*, 29 (1): 228-247.

———. 2008a. 'A Compatibilist Account of the Epistemic Conditions on Rational Deliberation'. *The Journal of Ethics* 12 (3–4): 287–306.

———. 2008b. 'A Hard-Line Reply to the Multiple-Case Manipulation Argument.' *Philosophy and Phenomenological Research* 77 (1): 160–70.

———. 2014a. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

———. 2014b. 'The Disappearing Agent Objection to Event-Causal Libertarianism'. *Philosophical Studies* 169 (1): 59–69.

———. 2014c. 'Responses to John Martin Fischer and Dana Nelkin'. *Science, Religion & Culture* 1 (3): 218–25.

Pettit, Philip. 1989. 'Determinism with Deliberation'. *Analysis* 49 (1): 42–44.

Plantinga, Alvin. 2009. 'The Evolutionary Argument against Naturalism: An Initial Statement of the Argument'. *Philosophy After Darwin: Classic and Contemporary Readings*, 301–9.

Olson, Jonas. 2016. 'On the Defensibility and Believability of Moral Error Theory'. *Journal of Moral Philosophy* 13 (4): 461–73.

Page, Carl. 1992. 'On Being False by Self-Refutation'. *Metaphilosophy* 23 (4): 410–26.

Popper, Karl. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

———. 1983. 'Is Determinism Self-Refuting?' *Mind* 92 (365): 103–104. https://doi.org/10.1093/mind/XCII.365.103.

———. 2012. *The Open Universe: An Argument for Indeterminism from the Postscript to the Logic of Scientific Discovery*. Routledge.

Rawls, John. 2009. *A Theory of Justice*. Cambridge, MA: Harvard university press.

Raz, Joseph. 1999. *Practical Reason and Norms*. Oxford: Oxford University Press.

Rieber, Steven. 2006. 'Free Will and Contextualism'. *Philosophical Studies* 129 (2): 223–52. https://doi.org/10.1007/s11098-004-7829-7.

Rinard, Susanna. 2017. 'No Exception for Belief'. *Philosophy and Phenomenological Research*' 94 (1): 121-43.

Rysiew, Patrick. 2016. 'Epistemic Contextualism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/contextualism-epistemology/.

Saka, Paul. 2000. 'Ought Does Not Imply Can'. *American Philosophical Quarterly* 37 (2): 93–105.

Sartorio, Carolina. 2016. *Causation and Free Will*. Oxford: Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198746799.001.0001/acprof-9780198746799.

Sartre, Jean-Paul. 1946. *L'Existentialisme Est un Humanisme*. Nagel Paris.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA; London: Belknap Press of Harvard University Press.

———. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA; London: Belknap.

Schlosser, Markus E. 2014. "Manipulation and the Zygote Argument: Another Reply." *The Journal of Ethics* 19 (1): 73–84.

Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.

Searle, John R. 2007. 'Neuroscience, Intentionality and Free Will: Reply to Habermas'. *Philosophical Explorations* 10 (1): 69–76.

Sehon, Scott. 2013. 'Epistemic Issues in the Free Will Debate: Can We Know When We Are Free?' *Philosophical Studies* 166 (2): 363–80. https://doi.org/10.1007/s11098-012-0044-z.

Sgaravatti, Daniele. 2013. 'Petitio Principii: A Bad Form of Reasoning'. *Mind* 122 (487): 749–79. https://doi.org/10.1093/mind/fzt086.

Sharples, R.W. 1983. *Alexander of Aphrodisias on Fate: Text, Translation and Commentary*. London: Duckworth.

Sher, George. 2005. *In Praise of Blame*. Oxford: Oxford University Press.

Slagle, Jim. 2016. The Epistemological Skyhook: Determinism, Naturalism, and Self-Defeat. Routledge.

Slote, Michael. 1990. 'Ethics Without Free Will': *Social Theory and Practice* 16 (3): 369–83. https://doi.org/10.5840/soctheorpract199016319.

Smart, J. J. C. 1961. 'Free Will, Praise, and Blame'. *Mind* 70: 291–306.

Smilansky, Saul. 2000. *Free Will and Illusion*. Oxford: Oxford University Press.

Smith, Angela M. 2005. 'Responsibility for Attitudes: Activity and Passivity in Mental Life'. *Ethics* 115 (2): 236–71. https://doi.org/10.1086/426957.

Stace, W. T. 1953. *Religion and the Modern Mind*. London: Macmillan.

Stanley, Jason. 2005. *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Steward, Helen. 2009. 'Fairness, Agency and the Flicker of Freedom'. *Noûs* 43 (1): 64–93.

———. 2012. *A Metaphysics for Freedom*. Oxford: Oxford University Press.

———. Forthcoming. 'Free Will and External Reality: Two Scepticisms Compared. The presidential address'. *Proceedings of the Aristotelian society*.

Stewart, Aubrey. 1900. *Seneca: Of Anger (Book III)*.

Strawson, Galen. 1994. 'The Impossibility of Moral Responsibility'. *Philosophical Studies*, 75 (1/2): 5–24.

———. 2010. *Freedom and Belief*. Oxford: Oxford University Press.

Strawson, P.F. 1992. *Analysis and Metaphysics*. Oxford: Oxford University Press.

———. 2008. *Freedom and Resentment and Other Essays*. Routledge.

———. 1985. *Skepticism and Naturalism: Some Varieties*. Woodbridge Lectures ; 1983. London: Methuen.

———. 1974. *Freedom and Resentment and Other Essays*. Routledge.

Streumer, Bart. 2016. 'No, We Cannot'. *International Journal of Philosophical Studies* 24 (4): 537–46.

———. 2017. *Unbelievable Errors: An Error Theory About All Normative Judgements*. Oxford: Oxford University Press.

Taylor, Richard. 1966. *Action and Purpose*. New York: Humanities Press.

Thalberg, I. 1962. 'Intending the Impossible'. *Australasian Journal of Philosophy* 40 (1): 49–56.

Todd, Patrick. 2013. "Defending (a Modified Version of) the Zygote Argument." *Philosophical Studies* 164 (1): 189–203.

———. 2016. "Manipulation Arguments and the Freedom to Do Otherwise." *Philosophy and Phenomenological Research* 95 (2): 395-407.

Tognazzini, Neal A. 2014. "The Structure of a Manipulation Argument." *Ethics* 124 (2): 358–69.

Unger, Peter K. 1984. *Philosophical Relativity*. Oxford: Oxford University Press.

Vihvelin, Kadri. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford: Oxford University Press.

———. 2015. 'Arguments for Incompatibilism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.

Vilhauer, Benjamin. 2009. 'Free Will and Reasonable Doubt'. *American Philosophical Quarterly* 46 (2): 131–40.

Waismann, Friedrich. 1965. *The Principles of Linguistic Philosophy*. Macmillan and Co Ltd.

Wallace, R. Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.

Waller, Bruce N. 1985. 'Deliberating about the Inevitable'. *Analysis* 45 (1): 48–52.

———. 2004. 'Virtue Unrewarded: Morality without Moral Responsibility'. *Philosophia* 31 (3–4): 427–47. https://doi.org/10.1007/BF02385194.

———. 2011. *Against Moral Responsibility*. Cambridge, MA: The MIT Press. http://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262016599.001.0001/upso-9780262016599.

———. 2014. *The Stubborn System of Moral Responsibility*. Cambridge, MA: MIT Press.

Waller, Robyn Repko. 2013. "The Threat of Effective Intentions to Moral Responsibility in the Zygote Argument." *Philosophia* 42 (1): 209–22.

Walter, Henrik. 2011. 'Contributions of Neuroscience to the Free Will Debate: From Random Movement to Intelligible Action'. *The Oxford Handbook of Free Will*, July. https://doi.org/10.1093/oxfordhb/9780195399691.003.0027.

Walton, Douglas N. 2006. 'Epistemic and Dialectical Models of Begging the Question'. *Synthese* 152 (2): 237–84. https://doi.org/10.1007/s11229-005-3984-4.

Watson, Gary. 2003. 'The Work of the Will'. In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, 172–200. Oxford: Oxford University Press.

———. 2004. *Responsibility and the Limits of Evil: Variations on a Strawsonian Theme*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199272273.001.0001/acprof-9780199272273-chapter-9.

White, F. C. 1989. 'Self-Refuting Propositions and Relativism'. *Metaphilosophy* 20 (1): 84–92.

Widerker, David. 1991. 'Frankfurt on "Ought Implies Can" and Alternative Possibilities'. *Analysis* 51 (4): 222. https://doi.org/10.2307/3328762.

———. 2000. 'Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look'. *Philosophical Perspectives* 14: 181–201.

Widerker, David, and Stewart Goetz. 2013. 'Fischer against the Dilemma Defence: The Defence Prevails'. *Analysis* 73 (2): 283–95. https://doi.org/10.1093/analys/ant013.

Wiggins, David. 1969. 'Freedom, Knowledge, Belief and Causality'. *Royal Institute of Philosophy Supplements* 3 (March): 132–54. https://doi.org/10.1017/S1358246100000655.

———. 1973. 'Towards a Reasonable Libertarianism'. In *Essays on Freedom of Action*, edited by Ted Honderich, 31. Routledge and Kegan Paul.

Willaschek, Marcus. 2010. 'Non-Relativist Contextualism about Free Will'. *European Journal of Philosophy* 18 (4): 567–87. https://doi.org/10.1111/j.1468-0378.2009.00374.x.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers, 1973–1980*. Cambridge, MA: Cambridge University Press.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.

———. 2001. 'Comments on Michael Williams' Contextualism, Externalism and Epistemic Standards'. *Philosophical Studies* 103 (1): 25–33. https://doi.org/10.1023/A:1010368318927.

Wittgenstein, Ludwig. 1974. *On Certainty*. Oxford: Blackwell.

Wolf, Susan. 1981. 'The Importance of Free Will'. *Mind* 90 (359): 386–405.

———. 1993. *Freedom within Reason*. Oxford: Oxford University Press.

Worsnip, Alex. 2016. 'Moral Reasons, Epistemic Reasons and Rationality'. *The Philosophical Quarterly* 66 (263): 341–61.

Wright, Crispin. 2007. 'The Perils of Dogmatism'. In *Themes from G. E. Moore: New Essays in Epistemology*, edited by Nuccetelli and Seay. Oxford: Oxford University Press.

Zimmerman, Michael J. 2002. 'Taking Luck Seriously'. *The Journal of Philosophy* 99 (11): 553–76. https://doi.org/10.2307/3655750.