

**FREE WILL AS ETIOLOGICAL
SELF-SUFFICIENCY**

by

Lok-Chun K. Gustin Law

B. A. in Philosophy, University of California, Berkeley, 2008

Submitted to the Graduate Faculty of
the Kenneth P. Dietrich School of Arts and Sciences in partial
fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH
THE KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Lok-Chun K. Gustin Law

It was defended on

1st June, 2018

and approved by

John McDowell, MA, Distinguished University Professor

Michael Thompson, PhD, Professor

Stephen Engstrom, PhD, Professor

Karl Schaffer, PhD, Professor, University of California, Irvine

Kieran Setiya, PhD, Professor, Massachusetts Institute of Technology

Dissertation Advisors: John McDowell, MA, Distinguished University Professor,

Michael Thompson, PhD, Professor

Copyright © by Lok-Chun K. Gustin Law
2018

FREE WILL AS ETIOLOGICAL SELF-SUFFICIENCY

Lok-Chun K. Gustin Law, PhD

University of Pittsburgh, 2018

To argue for the pessimism that free will is impossible in a deterministic world, a powerful way has been to invoke a case in which a behavior has a cause that is intuitively excusing and the cause is deterministic, beyond the agent's control, and opaque to the agent. The pessimist then argues: '(i) The agent in such a case is not fully responsible for the behavior. And (ii) if our world is deterministic, every behavior in it must also have a cause with the above characteristics. Therefore, we could not be fully responsible for our behavior in a deterministic world.' I do not dispute premises (i) and (ii) but contend that the conclusion does not follow. For if the behavior is right, the agent can be free in it and fully responsible for it even if it has a causal lineage of the said kind. In fact, if the behavior is wrong, the agent cannot be fully free in it whether or not it has such a lineage. This asymmetry follows from my conception of free action, which is in terms of a feature present in every intentional action: when someone is acting in a certain way intentionally, they thereby understand why they are so acting not as a fact they discover. Furthermore, if what they so understand is in a specific sense sufficient to explain why they are acting in the given way rather than not in that way, then they bear free will in the behavior. What the above etiological sufficiency takes is for the behavior to be an unimpeded manifestation of the agent's power of reason and for this power to be fully developed for considerations relevant to the practical situation. One feature that distinguishes this view from those of other free will optimists is the upshot that fully free action occurs much less commonly than it is allowed by those views. It also follows from the present view that, since an action is unfree in every way it is wrong, blame and resentment cannot be justified in a way in which recognition and gratitude can be.

Keywords: Spontaneity of rational representation, Self-consciousness in intentional action, Empirical explanations of conduct, Upbringing's impact on development, Susan Wolf's compatibilism, Anscombe on practical knowledge.

TABLE OF CONTENTS

PREFACE	x
1.0 INTRODUCTION	1
2.0 POTENTIAL RATIONALES FOR FREE WILL PESSIMISM	4
2.1 A CRITIQUE OF FRANKFURT-STYLE CASES	6
2.2 A TRADE-OFF BETWEEN MODESTY AND APPEAL	10
2.3 THE LACK OF ALTERNATE POSSIBILITIES IS INSUFFICIENT TO VINDICATE THE APPARENT GRIP OF RESPONSIBILITY INCOMPAT- IBILISM	11
2.4 THE DETERMINISTIC ASPECT OF A CAUSALLY DETERMINISTIC WORLD IS ALSO INSUFFICIENT TO VINDICATE THE APPARENT GRIP OF RESPONSIBILITY INCOMPATIBILISM	13
2.4.1 Resisting Pereboom’s inference to the best explanation	13
2.4.2 Argument from the case of someone who knows it all by observation	15
2.5 THE W ARGUMENT, DEBUNKING FOUR PERCEIVED BLOCKERS OF RESPONSIBILITY	17
3.0 THE INTELLECTUAL RELATION ONE BEARS TO THEIR IN- TENTIONAL ACTION	21
3.1 THE NAÏVE VIEW OF THE AGENT’S KNOWLEDGE OF THEIR ACTION	24
3.2 LIMITED INDEPENDENCE FROM MERELY DISCOVERABLE INVOLVE- MENTS	27
3.3 IN POSITION TO RECOGNIZE ONE’S DOING NOT SOLELY BY DIS- COVERY	34

3.4	TWO DESIDERATA ABOUT THE NON-THEORETICAL INTELLECTUAL RELATION THAT ISN'T KNOWLEDGE	37
3.5	DESIDERATUM 1: HOW DISAVOWAL OF KNOWLEDGE REJECTS 'WHY?'	39
3.5.1	The payoff of finding desideratum 1	39
3.5.2	Desideratum 1, found: the agent's non-knowledge relation is a practical thought of doing A	40
3.5.2.1	An objection to <i>Description-aware</i>	43
3.6	DESIDERATUM 2: THOUGHT THAT REMAINS NON-THEORETICAL EVEN WHEN IT AMOUNTS TO KNOWLEDGE	44
3.6.1	Desideratum 2, found: the agent must have <i>constitutive</i> practical thought of their intentional doing	44
3.6.2	The agent must bear their constitutive intellectual relation to some why	45
3.7	<i>SPONTANEOUS PRACTICAL THOUGHT</i> POINTS IN THE DIRECTION OF <i>KNOWLEDGE</i>	49
3.7.1	The 'split view'	49
3.7.2	Why must the agent's spontaneous practical thought of doing-A-because- <i>p</i> be conceptually anchored in spontaneous practical knowledge that embraces the doing	50
3.7.2.1	Premise one: My spontaneous practical thought of doing-A- because- <i>p</i> is also a constitutive intellectual relation to the fact that I am doing A.	51
3.7.2.2	Premise two: Constitutive intellectual relation to the fact that I am doing A, in a paradigm case of intentional action, involves what is transparently (to me) the activity of my capacity to do A in a context familiar in every way relevant to doing A.	53
3.7.2.3	Premise three: What is transparently the activity of my capac- ity to do A in a context familiar in every way relevant to doing A is spontaneous knowledge that I am doing A.	56
3.7.2.4	Conclusion	56

3.7.3	How the fringe cases are derivatives of the paradigm	57
3.7.4	The split view is wrong even in a fringe case of intentional action	57
3.7.5	Two suspicions concerning <i>spontaneous possession of why-answers</i>	58
4.0	AN IMAGE OF FREEDOM	60
4.1	DISCOUNTING CONSIDERATIONS AND THE NEED TO DETERMINE THEIR UNITY	61
4.1.1	Inability of conceptions of freedom in terms of ‘what one most wants’ to unite discounting considerations	63
4.1.1.1	‘What one most wants’ as what one’s deepest self wants	63
4.1.1.2	‘What one most wants’ as what one’s most real self wants	65
4.1.2	Unity of discounting considerations, preliminarily defined: the power of practical reason is not sufficiently developed and unimpededly operating	67
4.1.3	Susan Wolf’s Reason View	71
4.1.4	Problem regarding wrong doings	72
4.2	OBJECTIONS TO PREMISE W AND REPLIES	76
4.2.1	First objection to premise W (and subversion of Wolf’s Reason View), from the Paradox	76
4.2.2	Disarming the Argument for the Paradox	78
4.2.2.1	Heteronomy	78
4.2.2.2	Source other than the agent	79
4.2.2.3	Alien causal lineage	81
4.2.3	Second objection to premise W, from the explanatory ultimacy of up- bringing	82
4.3	FROM THE <i>UNITY</i> OF DISCOUNTING CONSIDERATIONS TO <i>FREE</i> <i>ACTION</i>	86
4.3.1	Unity of discounting considerations, defined more commensurably with empirically knowable nature	86
4.3.2	Free action in terms of the alternative to discovery	86
4.4	WHAT IT TAKES AND DOESN’T TAKE TO BE FREE IN ACTION	89
4.4.1	Free action as the manifestation of reason	92

4.4.2 Incorrect action occasions a need for merely discoverable explanations	94
4.4.3 Asymmetry between ideal manifestation of practical reason and incorrect action	97
4.4.4 The need of sufficient development of reason and the consequent rarity of free action	99
4.4.5 A free action need not be uniquely correct	101
4.4.6 A passion-governed behavior may also be free	101
4.5 CLARIFICATIONS AND REPLY TO AN OBJECTION	102
5.0 CONCLUSION	105
APPENDIX. MUST INTENTIONAL ACTION MANIFEST AN UNTRIGGERED POWER?	110
BIBLIOGRAPHY	113

PREFACE

Work on this dissertation has taken several years, and I have incurred many debts in the process. First and foremost, I thank my mother, Kam-Ying Flavia Ho, and father, Tung-Hoi Edwin Law for giving me opportunities for an invaluable education and lending me support in the long process of writing. I thank Alvin Ip, my best secondary school friend in Hong Kong, who through our daily conversations initiated me into philosophy in 2000. Although he does not pursue academic philosophy himself, every time I updated him on my project he was engaged and showed great faith in me and my work. I am also grateful for having met my friend Leif Hancox-Li, who introduced me early on in graduate school to the Pomodoro Technique as well as other ways to be disciplined and productive as an academic. Equally irreplaceable are the years of comradeship and counsel I received from my first accountability group, BLED: Marina Baldissera Pacchetti, Josh Eisenthal, and Haixin Dang, and my current group, HELLP: Samantha Hancox-Li, Josh, and Annalisa Paese.

I thank Logi Gunnarsson and Matthias Haase: they kindly met with me on a regularly basis to discuss earlier versions of the second substantial chapter during my visits at Universität Potsdam and Universität Leipzig, in the summer of 2015 and 2016, respectively. This chapter also benefited much from Kevin Busch's generous comments in our correspondence. Besides, my gratitude goes to the commentators at conferences where I presented versions of the first substantial chapter: Brad Stockdale at Florida State University and Yuuki Ohta at the University of Oxford. Their criticisms are nuanced, fair, and constructive. The same part of the dissertation also benefited from the comments given by the philosophers at my presentation at the Wilhelm Society meeting at UC Berkeley in spring 2016. And I thank Robert Howton, who pressed me over several conversations to elaborate and rework my argument to the effect that the agent who acts wrongly is to this extent unfree. This key component

of the third substantial chapter has much improved as a result. And I cannot overstate how much I owe my fellow graduate student colleagues at Pitt, who gave me countless helpful comments at various Work-in-Progress (WiP) talks and dissertation seminars.

I thank the University's Writing Center for providing a quiet space and often feeding me on days of focused writing. I especially thank Ellen Smith and Jean Grace, who gave me great pieces of advice on how to write the dissertation in a productive and sustainable way. I owe much to Frank Battista, then barista at the coffeeshop Voluto, where I wrote a lot of my dissertation (and most of my term papers). Thanks for feeding me, caffeinating me, and keeping Voluto open at 7 a.m. in the most difficult of times of writing! I thank William Hsu, Jess Liu, Leif, and Samantha for flying all the way from California to be at my defense. I am infinitely in debt to Annalisa, housemate and dear friend who kept me going and looked after me in the homestretch of my writing as well as on that momentous day.

I cannot thank enough my committee for their time, help, and support. The alternative perspectives Kieran and Karl showed me time and again are invaluable for keeping my work relevant and accessible to philosophical traditions and communities outside Pittsburgh. Their incisive but constructive observations are a window through which I am able to view my own thoughts critically at a distance. Steve read my writing with a moving level of attention and care; what he posed as requests for clarification often led to my better understanding of my own view. Michael has been an amazingly sympathetic and kind reader, who often tried to better articulate my ideas with so much effort as if they were his own. Also thanks to his constant exhortation, my sentences are much shorter and more readable now. No words will suffice to express my gratitude toward John for his patience and tireless guidance. I owe many key ideas in this thesis to him. In addition, over the recent years he has written me countless marginal comments with detailed stylistic and editorial suggestions. Guidance on how to write prose well in general is beyond what I can expect from my dissertation advisors. I thank Michael and John for every bit of such supererogatory help. The experience of working with this committee has been transformative and will benefit me for decades to come.

1.0 INTRODUCTION

It can seem impossible for one's action to be free in many scenarios. In one of them, every actual state and event in our world turns out to be caused with necessity by a set of predated states and events. One might argue, as an incompatibilist would, that such a world, which is causally deterministic, precludes free action. To argue for incompatibilism, one may come up with a thought experiment in which a supposedly wrong behavior has an intuitively excusing causal lineage with four stipulated characteristics: (1) how the agent behaves is necessitated by some factor other than themselves, (2) the causal lineage features a cause ultimately other than the agent, (3) the lineage is possibly opaque to the agent, and (4) it is ulterior to the action explanation the agent as such is aware of. The incompatibilist then argues: '(i) The agent in such a case is not fully responsible for the behavior. And (ii) if our world is causally deterministic, every behavior in it must have some cause lineage with those four characteristics. Therefore, we could not be fully responsible for our behavior in a causally deterministic world.' I do not dispute premises (i) and (ii). I argue that the conclusion does not follow because, if the behavior is right, the agent can be free in it and fully responsible for it even if it has a causal lineage of the said kind. In fact, if the behavior is wrong, the agent cannot be fully free in it whether or not the behavior has such a lineage. In either case, it is irrelevant to freedom whether the behavior has such a cause. I argue the above as I defend a conception of free action in terms of a feature present in every intentional action, to which G. E. M. Anscombe has drawn our attention. When someone is acting in a certain way intentionally, they¹ thereby understand why they are so acting not as a fact they encounter

¹ For an indefinite individual, I use 'they'/'them' as the third person singular pronoun instead of 'he'/'him' or 'she'/'her', and 'themselves' as the reflexive instead of 'himself' or 'herself'. These uses date back to early 16th century and late 15th century, respectively. Cf. *Oxford English Dictionary*, s.v. "Themselves, n.5" and "They, n.2".

or theoretically arrive at. If what they so understand, in addition, suffices to explain why they are acting in the given way rather than acting not in that way, then they bear free will in this behavior. What the above etiological sufficiency takes is for the behavior to be an unimpeded manifestation of the agent's power of reason and for this power to be fully developed for considerations relevant to the situation. One feature that distinguishes this view from that of many compatibilists and free will optimists of other sorts is the upshot that unqualified free action occurs much less commonly than it is allowed by those views. It also follows from the present view that, since an action is unfree in every way it is wrong, blame and resentment cannot be justified in a way in which recognition and gratitude can be justified.

In chapter one, I will resist free will pessimism in face of a causally deterministic world on the one hand, and, on the other, in face of the success and promise of human action explanations that are merely discoverable, thus possibly opaque, even to the agent themself. Against the first form of pessimism, I will object to two views each of which is commonly invoked as a premise to argue that full responsibility is precluded in a causally deterministic world. One of them is the view that responsibility requires that the agent be able to act otherwise. The other is that responsibility is precluded where the behavior is causally determined by factors beyond the agent's control. Against the second form of pessimism, I will debunk four tempting reasons for thinking that merely discoverable explanations of someone's behavior preclude full responsibility and free will: the agent is heteronomous, the action stems ultimately from a cause other than the agent, the action has a causal lineage alien to the agent, and there is a causal lineage that is ulterior to the behavior's ultimate causal lineage knowable to the agent as such.

In chapter two, I will argue that the etiological understanding the agent invariably has for their intentional action is not knowledge that they are acting γ -ly intentionally, but it is nonetheless constitutive of the intentional action. This is an actual, spontaneous intellectual relation the agent bears to their behavior, and this relation underlies the primary application of the why-question Anscombe famously invokes as the mark of intentional action.

Chapter three is where I motivate and articulate a conception of free action in terms of the spontaneity identified in chapter two. I motivate it via the need to give a unified account

of all the routinely accepted prima facie reasons for not holding someone fully responsible for their behavior. I first show that this is not satisfactorily accommodated by certain existing conceptions of responsible agency and free agency. I propose an alternative account: those routinely accepted considerations are all indications that an explanation of the behavior that is adequate in some sense *needs* an explanation that is merely discoverable even to the agent themselves, in the sense that the agent does not have enough of a spontaneous etiological understanding of their behavior. Hence, I propose that someone is acting freely just insofar as they have spontaneous etiological understanding of the ongoing behavior that is sufficient to explain why they engage in it rather than acting not in this way. What such sufficiency takes, I explain, is that the behavior be an unimpeded manifestation of an operating power of reason, in the agent, that is fully developed for possible considerations in their situation. My evidence for this is that such a power is indicated to be missing by a subclass of those routinely accepted considerations – the ones that at first blush diminish responsibility for the behavior specifically as belonging to a certain phase or to a class of the agent’s voluntary behavior. While Susan [Wolf \(1990\)](#) offers a similar account of responsible agency and free will that also explain how these prima facie reasons work, I argue, *contra* Wolf, that our kind of conception of free action suggests that one cannot bear a fully free will in their conduct if their conduct is wrong – not in a way, at any rate, in which one can bear a free will in their conduct if it is not wrong.

2.0 POTENTIAL RATIONALES FOR FREE WILL PESSIMISM

Is it possible to act so freely in a causally deterministic world as to be morally responsible for the action? While the case for free will optimism may require more than showing that moral responsibility for action in such a world is possible, it suffices for pessimism if that is shown to be impossible, i.e. if *responsibility* incompatibilism is shown to be true. In this chapter, we will first consider two influential ways in which responsibility incompatibilism is thought to be true. After that, we will look at four ways in which responsibility is thought to be incompatible with action explanations that, though not all associated with causal determinism, assume an objective stance toward human agents, are successful in various ways, but do not coincide with the explanations the agents themselves would give for their actions.

To say that a world is causally deterministic is to say that, for each of its temporal states of affairs and events, H, there is a set of predated states of affairs and events, P, such that P causes H with necessity. And here is what I mean by “morally responsible” or – for my purposes here, interchangeably with –, simply, “responsible”, regarding someone’s action: someone is (morally) responsible for acting in a certain way just insofar as they are a suitable target of engaged reactive attitudes – to use P. F. Strawson’s term (1974) – in accordance with how good or bad the action is. Engaged reactive attitudes (ERAs) are attitudes toward an agent for their conduct as if it exhibits their good will, lack thereof, or ill will toward another. ERAs include approval and indignation, gratitude and praise, resentment and guilt. In this sense, responsibility for an action is not ruled out just because the action does not call for any particular reactive attitude e.g. because it is permissible but not remarkable. And it is a subject’s responsibility for *acting in a certain way* (in contrast to a state or an object) that I am primarily concerned with, when I speak of responsibility here.

Acting in a certain way can be choosing something, willing something, doing something, or doing something for some reason.

According to one influential way of thinking, the deterministic aspect of a causally deterministic world, which precludes alternate possibilities in one sense or another, makes moral responsibility impossible. The grip of this thought is largely due to the Principle of Alternate Possibilities (PAP), i.e. that an agent is morally responsible for an action only if they could act otherwise. This has been taken by many to be *why* causal determinism and moral responsibility are incompatible. Although cases have been constructed as counterexamples to PAP, most famously the Frankfurt-style cases (FSCs), they have been resisted forcefully with the Kane-Widerker, dilemmatic objection. According to another influential way of thinking, responsibility incompatibilism is true because what I call the No-CD-beyond-control View is true: if a course of action is causally determined by factors beyond the subject's control, then the agent is not morally responsible for it. One famous contention for this View is Derk Pereboom's Four-case Manipulation Argument.

I will briefly review the PAP-versus-FSCs debate in §2.1 and, in §2.2, explain a trade-off that the responsibility-incompatibilist faces in construing “could...act otherwise” in PAP. There is incentive to construe it so thinly that no Frankfurt-style counterexample can be constructed to refute the resultant version of PAP. But this comes at the cost of the appeal of PAP even as a *prima facie* explanans of responsibility incompatibilism. I will argue in §2.3 that no sweet spot can be found in this trade-off and, in fact, no plausible version of PAP is true. Then, in §2.4, I show that the inference-to-best-explanation strategy in Pereboom's Manipulation Argument does not succeed in proving that the deterministic aspect of an action's deterministic causal lineage is relevant even if the No-CD-beyond-control View turns out true. I will also present an independent reason to think that the deterministic aspect is not sufficient as a ground even to diminish responsibility. This suggests that the only remaining candidate as a truth-maker of the No-CD-beyond-control is a view in terms of causation, apart from the notion of necessity. I will debunk this candidate and three other perceived blockers of responsibility in §2.5.

2.1 A CRITIQUE OF FRANKFURT-STYLE CASES

Frankfurt's case, in essence, is as follows. Someone, Black, stands ready to make sure that someone, Jones, will both decide to do A and have done it, if it becomes clear to Black that Jones is otherwise going to decide to do something else rather than A. But Black will not 'show their hand' unnecessarily. And, as it turns out, Jones ends up deciding to do A and having done it anyway, even though Black did not intervene. The point Frankfurt wants to illustrate is that, although circumstances like the presence of Black make deciding-to-do-A and doing-A unavoidable for Jones, this fact exonerates Jones neither from the responsibility of the decision nor from that of the doing – given that Black's power was not involved in bringing about Jones's decision to do A.

Let us make one modification to the case so that it may better illustrate the point. Frankfurt's original description of the case is silent about the possibility in which Jones merely does not decide to do A (as opposed to deciding to do something else incompatible with deciding to do A): it leaves open whether Black would intervene in that case. But that may well be an alternate possibility relevant to PAP: one in which the subject ends up not deciding to do A and not doing A, even though they have not made any decision that displaces or rules out their decision to do A either. So, let us add to the original case the condition that, should it become clear to Black that Jones is not going to decide to do A, Black would also intervene to reliably make sure that Jones decides to do A and will have done it.

To be a valid counterexample to PAP, a Frankfurt-style case (FSC) must be one in which an individual (a) is morally responsible for doing A and (b) could not act otherwise. A prominent challenge to FSCs is advanced by [Kane \(1985, 51\)](#), [Kane \(1996, 142-145\)](#), [Widerker \(1995a, b\)](#); [Ginet \(1996\)](#); [Wyma \(1997\)](#): no FSC can meet both conditions without some sort of question-begging. An FSC, this challenge goes, cannot meet condition (b) unless the decision and the action are causally determined (i.e., by some laws in conjunction with some set of predated states or events), but if they are supposed to be causally determined, then it would beg the question against at least a strain of responsibility incompatibilism to affirm condition (a), that Jones is morally responsible for those ways in

which he acts. In effect, the objection confronts Frankfurt's case with a dilemma: along the causally deterministic horn Frankfurt would seem at least to some to be presuming, on no independent ground, compatibility between responsibility and the truth of causal determinism; on the non-causally-deterministic horn there seem to be alternate possibilities to the action in question.

Let us take a closer look at the first horn. If Jones's actual way of acting is causally determined, then to claim that Jones is responsible without giving an independent reason why it is so despite this deterministic causal profile would be to presuppose compatibility between the conduct being deterministically caused and the agent's responsibility for it. But some might believe that, if it should turn out that a causally deterministic world does preclude responsibility, then the deterministic causation of an action would *alone* preclude responsibility for it. *For those who believe this*, to presuppose what must obtain in the present horn, i.e. compatibility between the conduct being deterministically caused and the agent's responsibility for it, would also be to presuppose the possibility of responsibility in a causally deterministic world. If they are also pessimists about this latter possibility, then to present a case as an FSC along the present horn would be to claim to have found a counterexample to PAP by presupposing what they are pessimistic about.

Now consider the second horn. Given that Jones's deciding to do A and doing A are not causally determined (i.e., by any laws in conjunction with any set of predated states and events), it seems that Jones could act otherwise, despite Black. If no combination of laws and predated states and events deterministically causes Jones's deciding to do and doing A, then there could have been no sure, prior sign indicating whether Jones would decide to do A as he in fact did. So, no matter how much Jones had seemed to be 'on track' to making that decision and how well Black's power had been monitoring Jones's circumstances, Jones might turn out in the last moment not deciding to do A, absent Black's intervention. Call this the no-warrant challenge. Peculiar to the non-deterministic horn, this is a challenge to show how, even if one's way of acting is not causally determined, they nevertheless could not act otherwise.

Eventually, I will in effect make the first horn viable when I argue that a subject's responsibility for acting in some given way is not precluded by the fact that it is deterministically

caused by some combination of laws and predated states or events. Over §§2.4.2, 2.5, 4.2, I will separate the tightly braided threads in such a causal profile that give rise to the apparent tension with responsibility and, one by one, show that they are in fact compatible with it. Once this is done, the coexistence of responsibility and the deterministic causal lineage of the action in the FSC would not be merely presupposed; an argument will be available to dissuade one from believing that the deterministic causation of an action would alone preclude responsibility for it even if a causally deterministic *world* does.

At the moment, however, we will be considering how some defenders of the Frankfurt-style argument take the second horn. They try to show that there can be a sure sign even if Jones's way of acting is not causally determined. Even though no combination of laws and prior states and events ensures Jones's decision to do A and his doing A, the response goes, it is still possible that Black's power works in such a way that Jones could not have avoided deciding to A and doing A, after all. One can attempt to demonstrate this possibility by adding to the FSC some further specifications, which I gather from cases conceived independently by [Stump \(1996, 1999, 2003\)](#); [Hunt \(2000, 2005\)](#); [Pereboom \(2000, 2001, 2005\)](#).

The first specification is that the actual decision ensued as a non-determined turn of events. Second, each possible course of events that is deviant enough to count as an alternate possibility entails a requisite process (e.g. entertaining a certain alternative, attending to certain considerations, etc.) that is necessary but itself insufficient to rule out the possibility of Jones's decision to do A. Third, a counterfactually-intervening power, like Black's, would have intervened upon detecting any of these initial processes, making Jones decide to do A and do A. Thus, Jones could not act otherwise than deciding to do A and doing A.

Those who advance the no-warrant challenge might be tempted to reply thus. The above, 'enhanced' FSCs themselves imply – in fact rely on – the availability of *some* alternate courses of events other than the actual history in which Jones decided to do A. While the inevitability of deciding to do A and thereby doing A means that Jones had no way around executing this decision of doing A, it does not rule out Jones's having some alternate possibilities to the actual world in which Jones proceeded. Given that someone is acting γ -ly in reality, let us call any alternate possibility in which they do not so act a robust alternative to the actual γ conduct, and one in which they also act γ -ly a non-robust alternative. It

seems that there must be some non-robust alternative to Jones' actual conduct if it is not deterministically caused. For, given that the actual decision to do A is not deterministically caused, at some point prior to its occurrence, what would unfold as the actual course of events must not be the only possible one. Now, the by-standing power, Black, would intervene only in response to some sign that some alternate course would unfold without his intervention. It is a sign that did not occur in reality. So, if Black had ever intervened, then the world *would have* had turned out in some way other than how it had in fact turned out. The world had to have deviated somewhat from its actual history before Black would intervene. To put it in terms of possibilities, Black's power could stop an alternate possibility from unfolding fully into a more substantial alternative only after it had unfolded at least a little bit. Therefore, if Black's power was ready to stop any robust alternative to Jones's non-deterministically-caused action from unfolding, there must be alternate – albeit non-robust – ways in which the world could have had turned out. So, the reply goes, we do not see a counterexample to the view that someone is responsible for their actually acting γ -ly only if there is an alternate possibility – robust or not robust. This view is even more modest than claiming that someone is responsible for their actual acting γ -ly only if there is a *robust* alternative. And the truth of the more modest view will already do the job for the

PaAP variant of the PAP Argument for Responsibility Incompatibilism

1. In a causally deterministic world, there is no alternate possibility whatsoever to any actual course of event.¹
 2. *Principle of any Alternate Possibilities* (PaAP). If there is no alternate possibility whatsoever to one's actually acting in a certain way, they are not morally responsible for it.
- \therefore *Responsibility Incompatibilism*. In a causally deterministic world, no one would be morally responsible for how they act.

¹This premise does not say that a causally deterministic world precludes what Hume calls hypothetical liberty: “if we choose to remain at rest, we may; if we choose to move, we also may” (*Enquiry* 8.23 / 95). It has argued that, for action responsibility, being able to act otherwise is needed in no other sense than something like that: e.g. if one had chosen (or tried, intended, etc.) to act otherwise, one would be acting otherwise. This way of construing PAP to stall incompatibilistic arguments, however, will not be discussed here. But cf. [Moore 1912](#), ch. 6 for instance and [Chisholm 1964](#) for a response.

2.2 A TRADE-OFF BETWEEN MODESTY AND APPEAL

However, in the actual discourse, effort has been made to show that there needs² or does not need³ to be a *robust* alternative (depending on the side the author takes) to the agent's actual course of action in an FSC along the second horn. It makes sense to debate this rather than whether there needs to be any alternative whatsoever, despite the modesty of PaAP. For as regards *why* someone is or is not morally responsible for their actual deciding or doing, not all alternative possibilities are *prima facie* equally relevant. And the less obvious the construal of PAP invoked, the more burden there is upon the defender of the PAP Argument to give further arguments for the connection alleged in the PAP so construed – between the lack of responsibility and the lack of alternate possibilities in the sense in question.⁴ Even if no FSC can be conceived to refute some version of PAP (which takes the place of the second premise), one need not thereby accept it – not unless it has some *prima facie* plausibility or we see some reason why it might be true.

And there is in fact a trade-off when it comes to construing 'could... act otherwise' in PAP: the thinner the sense of 'could... otherwise' and the more modestly PAP is construed, the less obviously the notion 'the agent could not act otherwise' therein could explain the agent's lack of responsibility. Compare the following construals of PAP ('If the agent could not act otherwise, they are not responsible for how they are actually acting'):

PcAP: If the agent cannot but act γ -ly, they are not responsible for their actual acting γ -ly.

PrAP: If there is no alternate possibility in which the agent does not act γ -ly, they are not responsible for their actual acting γ -ly.

PaAP: If there is no alternate possibility whatsoever, the agent is not responsible for how they are actually acting.

Each preceding construal entails the subsequent one. They are in an increasing order of modesty but – independently of any construal that precedes each – in a decreasing order of even *prima facie* relevance of the antecedent to the consequent. In particular, if there is

²E.g. [Palmer 2011](#).

³E.g. [Hunt and Shabo 2013](#).

⁴ Cf. e.g. [Hunt and Shabo 2013](#), 617.

a Frankfurt-style counterexample to both PrAP and PcAP, then even if no Frankfurt-style counterexample is possible against PaAP, the only sort of alternate possibility that makes FSCs along the non-causally-deterministic horn unviable against PaAP would be the non-robust sort. In that case, as I will show in §2.3, the prospect of defending PaAP and making a convincing case out of it for responsibility incompatibilism is very grim, at say the least.

This background concern, whether PAP is true *in a sense that could be apt for explaining responsibility incompatibilism*, is why it is worth settling whether there has to be at least a robust alternate possibility in an FSC along the non-causally-deterministic horn, as opposed just any old alternative to reality. Whether there is a sense of PAP that is immune to Frankfurt-style arguments is one question. Whether PAP is true in any sense that vindicates the apparent grip of responsibility incompatibilism is a different question. And to settle the second question, one must in turn discern whether anything about the fact that someone could not act otherwise, in any sense, might be a source of the lack of responsibility.⁵ To this task I now turn.

2.3 THE LACK OF ALTERNATE POSSIBILITIES IS INSUFFICIENT TO VINDICATE THE APPARENT GRIP OF RESPONSIBILITY INCOMPATIBILISM

In this section, I will present an argument to the effect that, if an individual in causally deterministic world could not act otherwise – in any of the discussed senses of ‘could not act otherwise’ –, that is sufficient to vindicate pessimism about responsibility in such a world. This argument also strongly suggests that the corresponding construals of PAP are false: PaAP, PrAP, and PcAP.

Suppose, on a public bus, Jeppa, seeing that an elderly person, Olanna, is struggling to maintain balance while standing, decides to yield her own seat to Olanna. We will imagine, one at a time, each of the following to be Jeppa’s world, and we will focus on whether her responsibility for her actual decision to yield to Olanna seems relatively diminished as we

⁵ A similar distinction of topics has been made in [Franklin 2011](#), 189-191 and [Pereboom 2001](#), 25.

move down the list:

World 2+: There are two sorts of alternate possibilities to Jeppa's actual decision to yield to the senior: (i) she decides to yield the seat to the senior later or (ii) she never decides to yield throughout the ride. And even in reality she is capable of remaining in her seat and bearing the sight of the senior's struggle throughout the ride.

World 2: There are exactly two sorts of alternate possibilities to Jeppa's actual decision to yield to the senior: (i) she decides to yield the seat to the senior later or (ii) she never decides to yield throughout the ride. In reality, though, she cannot but decide to yield to that senior, insofar as she is aware of the situation and so has a decision to make. Otherwise, this world is the same as World 2+. [Lacks cAPs]

World 1: There is exactly one sort of alternate possibilities to Jeppa's actual decision to yield to the senior: (i) she decides to yield the seat to the senior later. Otherwise, this world is the same as World 2. [Lacks cAPs and rAPs]

World 0: There is no alternate possibility whatsoever to Jeppa's actual decision to yield to the senior. Otherwise, this world is the same as World 1. [Lacks cAPs, rAPs, and aAPs]

Suppose Jeppa is responsible in World 2+. This supposition is not question-begging against any version of PAP, since Jeppa in this world could act otherwise in all senses.

World 2 differs from World 2+ only in that Jeppa2 (Jeppa in World 2) is not capable of acting otherwise than deciding to yield. This does not seem to make Jeppa2 any less suitable than Jeppa2+ as a target of engaged reactive attitudes (ERAs), e.g. approval or gratitude, for her actual decision. If so, then Jeppa2 is also responsible for her actual decision, and PcAP is therefore false.

World 1 differs from World 2 only in that there is no robust alternative to Jeppa's deciding to yield her seat to the senior. But Jeppa1, just like Jeppa2, cannot but decide to yield insofar as she notices the situation and the decision she has to make. So, World 1 lacks the robust alternatives apparently because it lacks alternatives in which Jeppa does not become aware of the situation. Such a difference does not seem to make Jeppa1 any less suitable than Jeppa2 as a target of ERAs for her actual decision to yield. At the very least, this suggests that the lack of robust alternatives itself is insufficient to vindicate the sense of tension between responsibility and the truth of causal determinism. And insofar as Jeppa2 can be responsible, so can Jeppa1. If Jeppa1 can be responsible, then PrAP is false.

World 0 differs from World 1 only in that there is not even a non-robust alternative to Jeppa's actual deciding to yield her seat to the senior, i.e., there is not even another manner in which Jeppa might arrive at the decision to yield. Such a difference between the Worlds does not seem to make Jeppa0 any less suitable than Jeppa1 as a target of ERAs for her actual decision to yield. At the very least, this suggests that the lack of non-robust alternatives itself is insufficient to vindicate the sense of tension between responsibility and the truth of causal determinism. And insofar as Jeppa1 can be responsible, so can Jeppa0. If Jeppa0 can be responsible, then PaAP is false.

The above thought experiment suggests this: even if it should turn out that causal determinism systematically precludes one's moral responsibility for how they act, it would not be because the agent could not act otherwise. So, if responsibility incompatibilism is true, something other than PAP has to account for it.

2.4 THE DETERMINISTIC ASPECT OF A CAUSALLY DETERMINISTIC WORLD IS ALSO INSUFFICIENT TO VINDICATE THE APPARENT GRIP OF RESPONSIBILITY INCOMPATIBILISM

2.4.1 Resisting Pereboom's inference to the best explanation

Even if the lack of alternate possibilities does not itself block moral responsibility, one may argue, causal determinism may nonetheless rule it out for a different reason. After all, in a causally deterministic world, it seems that one's actions and choices would all be deterministically caused by factors *beyond one's control*, and the following reasoning seems to hold:

CD-beyond-control Argument for Responsibility Incompatibilism

1. In a causally deterministic world, any course of action is ensured as the outcome of a process that begins with certain facts about natural laws plus causally relevant facts that predate this subject's existence.
2. If someone's course of action is ensured as the outcome of a process that begins with certain facts about natural laws plus causally relevant facts that predate this subject's

existence, then this course of action is causally determined by factors beyond their control.

3. *No-CD-beyond-control View*. If a course of action is causally determined by factors beyond the subject's control, then the agent is not morally responsible for it.

∴ *Responsibility Incompatibilism*. In a causally deterministic world, no one would be morally responsible for how they act.

For premise 3, Derk Pereboom argues that there is no relevant difference between how someone's choice of action might be ensured by natural laws and facts that predate x's existence and how it might be ensured by some manipulating power, as in the following scenario:⁶

Case 1. A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will [i.e., deterministically] produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. . . (2014, 76-77)

For case 1, the action is causally determined by factors beyond Plum's control and Pereboom thinks that *this* is why Plum's moral responsibility is compromised. And to the extent that we agree with Pereboom's analysis about this sci-fi case (and two other intermediate cases), he argues, for the same reason and to the same extent we should not regard people as morally responsible for their doings in *our* world if it turns out to be causally deterministic – if every historical fact in it is ensured by earlier facts in combination with natural laws.

It has been conceded, even by those sympathetic to Frankfurt,⁷ that the Frankfurt-type examples (thus far seen in the literature) are not enough to meet this form of challenge to moral responsibility. I now contend that this challenge can be met, in a way, if we employ Frankfurt-style reasoning to a scenario in which, contrary to Frankfurt's *scenario*, the agent is being manipulated. We can come up with one by modifying Pereboom's Case 1:

Random Disruptor. Everything is the same as in Pereboom's Case 1, except that (1) a random disruptor is available whose (random) activation would disrupt the radio-like

⁶ As well as in two other, intermediate scenarios, in his four-case manipulation argument.

⁷For example, see Fischer 2006, 200-202.

operation or intervene between Plum's reasoning process and his decision-making, such that the neuroscientists' ability to manipulate is not deterministic, and that (2) this disruptor has not actually intervened.

And the Frankfurt-style reasoning I want to employ is this. Given the disruptor, the natural laws and facts that predate Plum's existence do not ensure his decision to kill White. So, the radio-like technology is no longer a deterministic power over Plum. But this difference does not make Plum more morally responsible for his killing White than if it were absent (cf. Case 1), given that his decision is nonetheless caused by the radio-like technology. Therefore, the subtraction of the indefeasibility, the necessity, from the causal determination does not make Plum more morally responsible for his killing. Thus, it cannot be concluded by inference to the best explanation, from the comparison of a causally deterministic world to case 1, that the fact that an action is ensured, rather than merely caused, by factors beyond the agent's control has to do with why responsibility incompatibilism is true, even if it turns out to be true. In fact, if we grant Pereboom precisely his contention that there is no relevant difference between a causally deterministic world and Case 1 (2014, 76-79), then subtracting the necessitating aspect of the causal determination in the causally deterministic world would not affect how responsible Plum is for his course of action either.

Another straightforward response to Pereboom is, of course, to propose a *better* explanation than what he alleges for why Plum does not seem to be fully responsible in Case 1.⁸ But the persuasiveness of this alternative explanation depends on a successful defense of whatever is invoked therein as at least a necessary condition of free will. My explanation would be this: what is knowable to Plum as the agent is inadequate to explain *in some freedom-relevant way* why he killed White rather than acting without killing White. But I am not in a position to spell out and defend this explanation until I have defended my conception of free action.

2.4.2 Argument from the case of someone who knows it all by observation

Now I will argue that the deterministic aspect, i.e. the aspect of necessity, of the fact that an action is causally determined by factors beyond the agent's control is insufficient

⁸Cf. Haji 1998, 24-5. Cf. Boxer 2013, 31-5 for an evaluation of this and other forms of objections.

as a ground for taking them not to be responsible. I will argue this by arguing that the deterministic aspect of a causally deterministic world is insufficient to vindicate pessimism about the possibility of responsibility in such a world.

Bobo. Suppose some being is present in all of time and, at any point, already has observational knowledge of every historical fact in this world, and suppose this world has no temporal beginning. Call them Bobo, short for ‘being with omniscience by observation’. Every historical state of affairs and every event in this world, then, is predated by Bobo’s observational knowledge of it. Thus, in this world every historical fact is deductively derivable from natural laws and at least one earlier fact, namely, that Bobo knew it. Therefore, in this sense, this world is deterministic. Now suppose at time t_1 Plum decides to kill White. It follows that, before Plum comes into existence, Bobo already receptively knew that at t_1 Plum would decide to kill White. That Bobo had such receptive foreknowledge and the natural laws “ensure” Plum’s formation of his decision at t_1 to kill White. And both whether Bobo had this knowledge before Plum’s existence and the natural laws are beyond Plum’s control. So, the fact that, at t_1 , Plum decides to kill White is determined with necessity by earlier facts and natural laws, and these are factors all beyond Plum’s control. However, since Bobo’s knowledge is observational, it has no part in causing the decision to kill White. Such necessity, with which Plum’s behavior is determined by some set of states of affairs and events that predate him, is no reason to think that he is not responsible.

This suggests that the deterministic aspect, i.e. the aspect of necessity, of the fact that an action is causally determined by factors beyond the agent’s control is insufficient as a ground for taking them not to be responsible.

When someone’s behaving in a certain way is causally determined by factors beyond their control, (1) that the action occurs is deductively derivable from the fact that those factors obtain and (2) something among those factors causes the action. The case of Bobo shows that a feature of form 1 is realizable without a feature of form 2. It also shows that a fact of form 1 alone is insufficient as a ground for taking someone not to be responsible for the way they act.

So, even if responsibility seems precluded or diminished in a case where the action is causally determined by factors beyond the agent’s control, the deterministic aspect, i.e. the aspect of necessity, of this causal profile does not seem sufficient to account for it. The case of Random Disruptor suggests that it is not necessary either.

Pereboom’s No-CD-beyond-control View is, again, that if a course of action is causally determined by factors beyond the subject’s control, then the agent is not morally responsible

for it. What remains of it to be considered, then, seems to be a residual view in terms of a feature of form 2: an agent is not responsible (and therefore has no free will) in their acting γ -ly if this behavior is caused (deterministically or not) by factors beyond their control. That is to say, my thought experiments featuring the Random Disruptor and Bobo suggest that the No-CD-beyond-control View is true only if this residual view is true. And since an action's being ultimately caused by something other than the agent is a case of an action's being caused by factors beyond the agent's control, the residual view entails this weaker view:

No-source-other-than-self View. An agent is not responsible, and therefore has no free will, in their acting γ -ly if this behavior has an ultimate cause that is other than the agent themselves.

Indeed, that an action has an ultimate cause other than the agent might seem already sufficient to preclude the agent from being responsible for the action. But I will argue that even this weaker view (and three other neighboring, seeming rationales for free will pessimism) is false in §2.5 and reply to an objection in §4.2.2.2. If so, i.e. if an action's being ultimately caused by something other than the agent is not a true blocker of the agent's responsibility, then that the behavior is caused by factors beyond the agent's control is not so either. In that case, the residual truth-making candidate of the No-CD-beyond-control View is false.

2.5 THE W ARGUMENT, DEBUNKING FOUR PERCEIVED BLOCKERS OF RESPONSIBILITY

Besides the agent's not being able to act otherwise and its being the case that their action is causally determined by factors beyond their control, each of four other conditions is also widely held to preclude the agent from being responsible for their action, and hence preclude the action from being free: if (1) how the agent behaves is necessitated by some factor other than themselves, if (2) the action has an ultimate cause other than the agent, if (3) it has a causal lineage that is not knowable to the agent otherwise than as something they find out,

and if (4) the action’s ultimate causal lineage that the agent has access to *just by being the agent* has an ulterior causal lineage to which the agent has no such privilege of access. I will call these the four perceived blockers of action responsibility. Since one is not free in doing something unless they are also responsible for doing it, we have in view, accordingly, four supposed potential rationales for free will pessimism:

1. *Aheteronomy View*. An agent is not responsible, and therefore has no free will, in their acting γ -ly if whether they so act or not is determined with necessity by some factor other than themself.⁹

This View motivates (and commits) one to endorse more specific conditions of non-necessitation, e.g. the “volitional control” that Haji (1998) claims (168) to be necessary for action responsibility: “Action, A, performed by agent, S, is under S’s volitional control if and only if, holding constant the motivational precursor of A (that is, the proximal desire or pro-attitude that gives rise to A) and S’s evaluative scheme¹⁰, there is a scenario with the same natural laws as the actual world in which, relying on S’s evaluative scheme, S decides or forms an intention to do something other than A, and successfully executes that intention or decision” (169).

2. *No-source-other-than-self View*. An agent is not responsible, and therefore has no free will, in their acting γ -ly if this behavior has an ultimate cause that is other than the agent themself.
3. *No-alien-causality View*. An agent is not responsible, and therefore has no free will, in their acting γ -ly if it has a causal lineage that is not knowable to the agent otherwise than as something they discover.
4. *Ultimacy View*. An agent is not responsible, and therefore has no free will, in their conduct, if it has a causal lineage that is ulterior to its ultimate causal lineage knowable to the agent just by being the agent of this conduct.

⁹ The Aheteronomy View may sound similar to the Principle of Alternate Possibilities (PAP), that an agent is not responsible for their acting γ -ly if they could not act otherwise. In fact, however, neither the idea that whether someone acts as they actually do is necessitated by some factor other than themself, nor the idea that they could not act otherwise than how they actually acts, entails the other. Therefore, neither of the Aheteronomy View and PAP entails the other.

¹⁰ “Such a scheme has doxastic elements including beliefs about normative standards to assess reasons for action, and beliefs regarding deliberative principles to be used to arrive at practical judgments about what to do. Such a scheme also has motivational elements like the agent’s values and pro-attitudes to engage the deliberative principles the agent believes should be utilized to make practical judgments” (Haji 1998, 169).

These four Views above are each a potential rationale for pessimism about the possibility of free will, in face not only of causal determinism but also of the increasing success and promise of e.g. psychological, neurological, and genetic explanations of human behaviors. I will call them the four supposed rationales for pessimism.

One reason why the perceived blockers of action responsibility appear to preclude an action from being free is this. Each of them seems to be the relevant condition indicated by some commonly accepted reason to not treat the agent as a fully suitable target of ERAs for a behavior, e.g. ignorance, brainwashing, addiction, etc. But if any of these perceived blockers were genuine, then someone's foolproof good upbringing would be a reason not to take them to be responsible for their resultant *good* conduct, as someone's deprived and inevitably corrupting upbringing can be a reason to discount them as a fully suitable target of ERAs for their resultant bad conduct. For even a foolproof upbringing that gives someone practical wisdom can (1) necessitate the agent to act in a certain way, e.g. wisely, (2) be a source other than the agent, (3) be a causal lineage that is not necessarily transparent to them, and (4) be a causal lineage *ulterior* to the ultimate causal lineage knowable to the agent as such for their action. Unlike a deprived and inevitably corrupting upbringing, however, we would not take a foolproof education of practical wisdom (if there is happily ever such a thing!) to detract from its graduates' desert of approval when they, out of their infallible sensitivity and unwavering character, *act justly* in response to their situation despite the adverse circumstances. Even if this person is so educated and constituted that there is no scenario in which they with the same "evaluative scheme" would act otherwise than justly, so that they do not have what Haji takes to be the necessary "volitional control" of their just behavior, neither the education nor the lack of such "control" seems to reduce the level of approval appropriate to them for the just conduct. An increase in the extent to which an upbringing makes someone wise or excellent in action does not incline us to decrease the degree of approval we deem suitable to the agent regarding their consequently excellent conduct. There is an asymmetry, between good conduct and bad conduct, in how the agent's upbringing can affect their suitability as a target of moral responses.¹¹ To summarize my W

¹¹ This echoes Susan Wolf 1980, "Asymmetrical Freedom", though there Wolf targets in effect only what I labeled the Aheteronomy View. And whereas she appeals to our intuition about someone who couldn't help responding to their situation in the way the situation calls for, here I appeal to our intuition about someone's

Argument,

- F. If any of the four perceived blockers of action responsibility is genuine, then someone's foolproof good upbringing would be a reason to discount them as a fully suitable target of ERAs for their conduct insofar as it is a product of the upbringing.
- W. A foolproof education of practical wisdom would not be a reason to discount the educated person as a fully suitable target of ERAs for their resultant wise conduct.
- D. Therefore, none of the four perceived blockers of action responsibility is genuine.

There are objections to premise W which I shall address when it is opportune, in §4.2. In §4.3.1, I will propose an alternative as the relevant factor indicated by a deprived and inevitably corrupting upbringing and other commonly accepted considerations, insofar as they excuse someone for their wrongdoing.

good *upbringing* to debunk Heteronomy as well as three other perceived blockers of action responsibility.

3.0 THE INTELLECTUAL RELATION ONE BEARS TO THEIR INTENTIONAL ACTION

As [Anscombe \(2000\)](#) suggests, one's intentional action is their doing to which the question, why they are doing it, in a special sense, has application.¹ The special sense, I believe, has to do with the fact that the agent intentionally doing something thereby has access to an answer to the 'What?' and 'Why?' about this conduct. To be more specific, there is fact, as to what they are doing and why they are doing it, that is knowable to them not as something they discover. By 'discovering x' I mean *finding x out*, a way of representing x non-accidentally truly. This includes ways of encountering x: observing x, introspecting x, learning x by testimony, etc., but it also includes theoretically arriving at a conclusion that affirms x. These are all discovery-involving ways of representation, intellectual relation, and knowledge. Sometimes I will refer to them as *theoretical*:

Familiar theoretical ways of intellectually relating to a fact, i.e. familiar ways of intellectually relating to a fact as what one discovers, finds out.

- Receptive: relating to the fact as what one encounters
 - by observation
 - by introspection
 - by testimony
 - ...
- relating to the fact insofar as it is affirmed by the conclusion in one's theoretical inference

Thus, it seems that the agent invariably enjoys a non-discovery-based intellectual relation to their intentional doing. If certain further conditions are met, then, this dissertation

¹ She thinks that, more precisely, the question is also applicable, at least in a way, to voluntary actions that are not intentional, e.g. when the answer is 'I don't know why' in such a sense as concedes that the movement is of the type that is typically executed for a reason [2000](#), 25-26, 89.

contends, to bear such an intellectual relation is to be acting freely, to be free in the doing, to bear free will in the conduct. The present chapter is dedicated to determining what this relation consists in.

It may appear that it must be knowledge of some sort, since saying ‘I didn’t know I was doing it!’ is a way to imply that one did not do the thing intentionally. Hence the

Naïve View. The agent who is doing A intentionally, as such, knows that they are doing A.

And if we may call doing A under the condition disavowed by ‘I didn’t know I was doing A!’ ‘doing A knowingly’, we are looking to determine the sense of this ‘knowingly’ in which

Doing-knowingly. To be doing A intentionally is at least in part to be doing A knowingly.

And the Naïve View is a straightforward version of *Doing-knowingly*.

But the Naïve View seems to fly in the face of many possible scenarios. In §3.1 we will see examples in which someone apparently is intentionally doing something without knowing that they are. In some cases, one is intentionally doing (and sometimes also intends to do) either something that is transparently difficult (Davidson 1978, Marušić 2015, 80-96) or something such that though it is not difficult, it is not transparent to one how easy it is to do, or even how probable it is that one will manage to do it (Setiya 2008). In some other cases, the on-going intentional action depends on the fact that the world has turned out one way rather than another, and the agent has virtually neither control over, nor any way of knowing other than by finding out, which way it has turned out.

There appears a way to reconcile these supposed counterexamples with the Naïve View that the agent who is doing something intentionally invariably has knowledge that they are. This strategy, which we will review in §3.2, is made available in Kevin Falvey 2000 and Michael Thompson 2011, though Falvey and Thompson themselves do not maintain the Naïve View. The strategy is to appeal to the fact that whatever knowledge someone may have *in virtue of their intentionally doing something* is necessarily a thought of the doing in the progressive, which holds quite independently of what has and has not happened. I will argue that, since this independence is limited, there are possible cases of intentional action that cannot be accommodated by a defender of the Naïve View using this strategy.

At the end of the day, I am convinced, an agent doing A intentionally may meanwhile be uncertain or even rationally disbelieve they are doing A, and they may really have to find out to know that some specific thing has not happened that would have defeated the point of their effort, e.g. the regime has not executed one's target of rescue. And given that certainty or absence of rational disbelief is necessary for knowledge, one can be intentionally doing something without knowing that they are doing it: the Naïve View is false. So, the non-discovery-based intellectual relation, borne invariably by the agent to their intentional doing, is not knowledge.

Nonetheless, it has been insisted, this intellectual relation is to be understood in terms of knowledge. Some, although they concede the falsehood of the Naïve View, are keen to insist that practical *knowledge* is essential to intentional action. But to say exactly how this is so is difficult work. In §3.3, we will consider a proposal by John McDowell, according to which the agent, doing something intentionally, invariably has, in potentiality, some non-discovery-based knowledge of their doing. Specifically, McDowell argues that someone intentionally doing A is always ready to recognize, without finding out, the conceived effects of doing A as due to their own doing. While at least some variant of this view has to be true, I argue that there is also an intellectual relation, not based on discovery, that the agent invariably *and actually* bears to their intentional doing.

Suppose, then, in some other sense than the Naïve View, to be doing A intentionally is in part to be doing it knowingly, and to be doing A knowingly is to be doing it with at least something *like* actual knowledge that I am doing A. Three questions need to be answered. (i) What intellectual relation does an agent as such invariably bear to their intentional action?² (ii) What is the nature of this intellectual relation *so that* this agent's authorship extends over happenings that take place not only in their head? And, again, (iii) how is practical knowledge essential to intentional action even though the Naïve View is false?

The key can be found as we make sense of how 'I didn't know I was doing A!' can squarely deny that one was doing A intentionally, rejecting the question 'Why?' – *even* when circumstances are such that, although one might have been doing A intentionally, one

² For example, does it include the fact that the intentional agent is themselves doing A, that they are trying to do A, that they are making such and such bodily movements to do A, or only that they intend to do A? And does it include their reasons for doing A, that it is in some sense good to do A, etc.?

would have been so acting only with doubt whether one was doing A and so could not have known that one was doing A. In §3.5.2, I distinguish between two cases that make it true to say ‘I did not know’: having no idea and merely not being certain. Having no idea precludes certain items of knowledge that someone doing A intentionally as such must have. (Now I answer question (i).) One is non-theoretical knowledge that they are trying to do A. Another is non-theoretical knowledge that they are not ready to deny, with rightful certainty, that the description *doing A* describes something they are doing intentionally. The agent who is doing A intentionally has these items of non-theoretical knowledge, for it is part of what it is to be doing what they are intentionally doing that they represent the doing. (Thus, my answer to (ii).) To represent a doing in this constitutive way is to bear a *spontaneous* intellectual relation to it. In §3.6.2, I will suggest that the agent must bear such a relation not only to the what but to a why of their intentional doing. The Anscombian why-question is applicable just when the agent non-theoretically possesses an answer to the question, why they are doing A, that is non-accidentally true. But such non-accidental truth on its own, considered separately from knowledge, is metaphysically dubious. This is the beginning of an answer to question (iii), how is practical knowledge essential to intentional action as a kind. In §3.7, I will discuss this conceptual instability in detail. From this, I will demonstrate that those fringe cases count under intentional action only because they are relevantly similar to normal cases – in which one does something of the same sort in a transparently familiar context. And in these normal cases, the agent’s spontaneous representation is knowledge as much as is our everyday success in exercising our capacity to know by encounter or theoretical inference.

3.1 THE NAÏVE VIEW OF THE AGENT’S KNOWLEDGE OF THEIR ACTION

The idea that to act intentionally is also to act knowingly seems to be reflected in our everyday practice, that to say ‘I didn’t know I was doing A!’ is a way to deny that one did A intentionally (and thus to deny that one is responsible for doing A, to some extent), which

practice occasions some of the most intriguing discussions in *Intention* ([Anscombe 2000](#)). That disavowal of knowledge is a way to reject the question, what is it that the subject is doing or why the subject is doing A, if this presupposes the doing in question to be intentional. This custom seems to suggest that, if someone is doing something intentionally, they *know* that *they are doing it*, i.e., the Naïve View.

Philosophers have objected to this, e.g. Donald Davidson and, more recently, Kieran Setiya and Berislav Marušić. Davidson came up with the following example in which the agent is doing A intentionally and successfully without knowledge, or without even much confidence, that they are doing A:

... in writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally. (([Davidson 1978](#), 92))

It is still fair to say that this copier knows that he is writing heavily on the topmost page in order to make ten carbon copies. So, the apparent moral of this example is that, even if the subject doing A intentionally as such knows that they are doing *something* in order to do A, they do not necessarily know that they are really doing A.³

But one might even question whether the intentional doer must have any knowledge that they are doing anything at all. Consider this cautiously optimistic post-paralysis fist-clencher.

Imagine, for instance, that someone has recently been paralyzed. At a certain point in their recovery, they were cautiously but not irrationally optimistic: they thought that they might be able to clench their fist. Then they tried to hold their fist in a clench for two minutes. Now, suppose they have succeeded, but during these two minutes they had no way to find out whether their fist was tightened. Further, based on what they know about similar patients in the past, they were not confident they were clenching their fist. Nor is it plausible to claim that they believed they were doing something in order to get their fist clenched. For, say, this patient denies that they clench their fist by an inner act of volition, and they have no thoughts about the firing of nerves. There is nothing else that they believed they were doing, as a means to clenching their fist.⁴

So, apparently, even if someone is intentionally doing A, we cannot thereby attribute to them knowledge and certainty that they are doing. It may appear that we must qualify the claim

³ Note that this is not what Davidson tries to show when he originally introduces the example. His point is, rather, that intending to do A does not entail believing that one will do A or is doing A.

⁴ This is a more detailed adaptation of the original example from [Setiya 2008](#), 389-391.

in a way similar to the one proposed in [Setiya 2017](#), 10:

When one is doing A intentionally, they are more confident that they are doing A than they would otherwise be.⁵

In fact, even in the fist-clencher’s scenario above, it can be reasonable to presume the patient, while clenching their fist intentionally, must believe they are doing *something* relevant to clenching their fist intentionally. It is reasonable given that they can *usually* clench their fist at will and have done so many times before, as we commonly can do and have done many times. For even as they are clenching the fist intentionally in such unusual circumstances, they will have strained in a way that they can recognize as how they *usually* strain to clench their fist intentionally – even though this time their straining is not accompanied by any proprioception or perception. If they cannot usually clench their fist at will or have not done so many times before as we commonly can do and have done, it is doubtful whether they could now clench their fist intentionally – even if their kinesthesia and proprioception were not specially disabled. (In fact, there would not yet seem to be a distinction between usual and unusual circumstances for fist clenching for this individual.) Thus, this scenario is not a convincing counterexample to the claim that the agent must believe or know that they are doing something to do A, or trying to do A, insofar as they are doing A intentionally.

Whether there is a need to ‘retreat’ so far as to Setiya’s claim above or not, whether there is a counterexample to Setiya’s claim or not and, if there is, whether some variation of the claim is true, the attribution of high credence or confidence to the agent does not suffice to illuminate the immediacy and non-theoretical way in which the agent relates to their intentional action.⁶ For having a relatively high credence or confidence about something might be entirely based on observation, after-the-fact introspection, the conclusion of a theoretical inference, etc. It would be rather surprising if *Doing-knowingly* consists just in that. Further, even given the truth of a claim like the above, it is still obscure why the disavowal of knowledge that one is doing A can ever be a way of legitimately denying that one is doing A intentionally.

⁵ Cf. [Setiya 2008](#), 391, [Paul 2009](#), 549-550, and [Setiya 2009a](#) for Setiya’s earlier views and part of the discourse that led to his current view.

⁶Setiya, for the questions he wants to answer about the agent’s cognitive standing regarding their intentional action, also does not stop at defending an attribution of confidence to the agent. See [Setiya 2011](#), [2012](#), and [2017](#), 11-13.

3.2 LIMITED INDEPENDENCE FROM MERELY DISCOVERABLE INVOLVEMENTS

Both examples above challenge the Naïve View thus. Often, certain happenings are intended by the agent as part of their intentional action. For example, the appearing of script on some carbon copies is an integral, intended part of the intentional action of making ten carbon copies. Call such involved happenings the intentional action's involvements. Often, these involvements depend on things' having turned out one way rather than another, on states of affairs that are not knowable to the agent otherwise than as something they find out. Call an involvement with such a dependence a worldly involvement. The appearing of script on a sheet, for example, requires that some bit of script is already present on it, which is a fact not knowable to the agent unless they find out. The appearing of script thus is a worldly involvement of the intentional making of carbon copies. At a moment when the agent would not be doing A unless one of this action's involvements is happening, they cannot know that they are doing A unless they know that this involvement is happening. But the above scenarios appear to show that they can be doing A intentionally without knowledge of the latter sort because the involvement is worldly. Thus, they apparently also lack the knowledge that they are doing A. Therefore, the argument goes, the Naïve View is false.

One might object that scenarios like the two above do not force us to give up the Naïve View. There is a distinction, the objection goes, between the perfective and the progressive aspects of descriptions. Drawing on this difference, one can argue that the content of the agent's special knowledge of what they are intentionally doing is always in the present progressive, and merely being engaged in doing A, in the progressive, need not involve any particular involvement entailed by one's having advanced toward having done A, in the perfective. If this is true, then the fact that the subject may not sense or perceive this or that event presents no obstacle to the view that the subject always knows that they *are doing* whatever they are doing intentionally.⁷ While the copier is writing hard on the topmost sheet, for example, even if they do not know whether any script is appearing on all ten sheets, and so do not know if they have produced any mark on them all, it can still be the

⁷Cf. [Thompson 2011](#), 209-210, though, again, Thompson does not defend the Naïve View.

case that this agent, who is really making ten carbon copies intentionally, as such knows that they *are making* ten carbon copies. This defense of the Naïve View does not dispute the idea that, in the case of certain processes that are not actions, something's being in progress requires that some progress has been made. E.g. a billiard ball must have moved if it is moving, some script must have appeared on all ten sheets if some script is appearing on all ten sheets. Apparently, then, in the typical case of a non-action process, V-ing, the following holds: *V-ing* entails *having V-ed to some extent*.

The defense does challenge, however, whether an action in progress requires that this or that non-action process is under way. It contends, e.g., that the mere fact that someone is in the process of intentionally making ten carbon copies does not require that script be already appearing on the sheets, and so does not thereby entail that some bit of script *has* appeared on them. For this subject may give it another try, adjust how they do it, etc. if they, say, later learn that their impression has not been strong enough to produce the desired copies all at once. Thus, that this person does not know that some script *has* appeared on these sheets does not rule out their knowing that they are making ten carbon copies.

Admittedly, being in the process of doing something can be somewhat independent from worldly involvements and from whether one has done something. So accordingly, knowledge that one is in the process of doing something can also be somewhat independent from the knowledge of worldly involvements. In many cases where the agent is doing A, there is hardly any concurrent, individual happening that is *singly, irreplaceably* necessary for the subject to be in the process of doing A, rather than its being necessary that a happening merely *of some sort* be actual. However, such independence may amount to nothing more than that, *usually*, the worldly involvements entailed by an intentional doing A in progress are *less definite than* the happenings entailed by a non-action process in progress that is also called doing A. The independence seems limited in two ways.

First, for all that has been said, it may hold, even in the case of intentional action, that V-ing entails having V-ed to some extent. Grant, for the sake of argument, that intentionally doing A does not simply entail the occurrence of a worldly involvement called doing A, and does not thereby simply entail that one has advanced somewhat toward having done A. Intentionally doing A may nevertheless require a disjunction of worldly involvements. In

Davidson's scenario, were no script already appearing on the sheets, the subject would still count as making ten carbon copies intentionally only if some script is appearing on the first page, or the pen-holding hand is exerting pressure on the stack of sheets, or... each of which is a worldly involvement. Accordingly, the agent knows that they are doing A intentionally only if they know that one or more of its worldly involvements is happening.

Here is another way in which the independence is limited. We can imagine a situation in which, had something not taken place, the subject would already have failed to do A and would no longer be doing A intentionally. For example, we can add to the fist-clenching scenario the supposition that the patient intended to clench their fist for two minutes *from 2:00 to 2:02*. Suppose they succeeded: they were clenching their fist the whole time. But during that time, just as in the original scenario, the patient had no knowledge about the status of their hand. So, e.g. at 2:01, they did not know that their fist had been in a clench. Yet, had it not been in a clench then, this patient would already have failed, and so would not be clenching their fist from 2:00 to 2:02 intentionally.

Our defender of the Naïve View may reply thus. In a case where (a) the subject is rationally skeptical of the chance of success of their attempt to do A and (b) there is no opportunity to check or try again, their success would be no more an example of intentional action than winning a lottery is with a ticket one bought with this aim.⁸

In response, it may be argued that the way an ability is involved in the case of the fist-clencher distinguishes their clenching as intentional action. First, the patient is rationally skeptical about whether it is once again *within their power* to clench one's fist, whether they have regained the ability to reliably squeeze with their hand, potentially many times over. In contrast, someone entering a lottery is rational to be skeptical about winning it because it would depend on the outcome of a *random* process or, at least, a process intended for randomization, which seals off the outcome from one's power. This makes the fist-clenching more of an example of intentional action than lottery-winning.

Yet, one who is keen to discount this scenario as a case of intentional action could point out that, even if fist-clenching is within the clencher's power, the manifestation of such power is not thereby an intentional action. The fist-clencher may be confronting compelling reasons

⁸ This is the response to Davidson's case of the carbon-copier in [Thompson 2011](#), 210.

to be quite confident that they have not recovered the power and so are *not* exercising it to clench their fist, even though, in fact, they are. This patient may know that, say, of the 1,000 people who had had surgery of the exact same kind as theirs and who are not – as far as is known – relevantly different from them, only 50 of them recovered from paralysis in so many days after the surgery; the others took longer to recover. Statistically speaking to the best of everyone’s knowledge including the patient, then, the chance the patient is no longer in paralysis right now is only 5%. So, it may even be *irrational* of them *not* to be *uncertain* that they are clenching their fist, and quite reasonable of them to believe with confidence that they are not doing it. It follows that doing something by exercising a bodily capacity, so long as the possession of this capacity is not known to one, seems insufficient for this doing to count as an intentional action. Nor is there any other know-how that the patient takes themselves to be exercising in order to get their fist clenched. On these counts, we do not seem to have a case of intentional action.

However, some other features about both this and the carbon-copying scenario mark the contextually salient doings as *not* merely voluntary actions, whether in the sense of an inadvertent movement to which the why-question applies in a borderline way⁹ or in the sense of a movement considered by the agent but merely as concomitant to some other action in which they are engaged. I will now rule out these two disjuncts one at a time.

First, each of the two doings under dispute is considered by its agent not merely as concomitant to some other activity in which they are engaged. First, the answer to the question ‘What is the subject doing?’ must be readily accessible to each agent during their engagement. In this way, each of the doings under discussion is considered by the doer while in process. And insofar as the doer considers their doing through the description of making ten carbon copies or clenching one’s fist, they also consider the doing’s primary involvements, such as the appearing of script and the fist’s being clenched. So, the script’s being written on all ten carbon copies by the writer, and the fist’s being brought to and held in a clench by the patient, must be considered by their respective subjects in the meanwhile. Clearly,

⁹ Cf. Anscombe 2000, 89, §49: “Mere physical movements, to whose description our question “Why?” is applicable, are called voluntary rather than intentional when (a) the answer is e.g. “I was fiddling”, “it was a casual movement”, or even “I don’t know why” (b) the movements are not considered by the agent, though he can say what they are if he does consider them. . . .”

then, in each case, the doing under dispute – even along with some worldly involvement – is considered by the agent during their engagement.

Second, I now argue, the question ‘Why?’ that applies primarily to intentional actions applies here in a non-borderline way. Suppose for the sake of argument that e.g. the patient is rationally skeptical whether they are clenching their fist and, they would bet, given the odds, that they are not clenching their fist. Still, the rational way for them to respond sincerely if someone asks them why they are clenching their fist, is not ‘But I am not clenching!’ Nor is it ‘I didn’t know I have been clenching my fist!’ Rather, it would be e.g. ‘I am? Good. I have been trying to clench my fist in order to see whether I have recovered from my paralysis.’ Likewise in the carbon-copying scenario. And this answer is true in the sense that this agent is clenching their fist on consideration of a reason. Likewise in the carbon-copying scenario. Hence, in each contested case the doing is considered by the agent during their engagement not just as concomitant to some other of their activities.

To be more specific, in each scenario, the what- and the why-questions applicable primarily to intentional actions are here applicable, and the what- and some why-answers are readily accessible in the agent, in the sense that both what they are doing and why they are doing it are knowable to the agent, during their engagement, *not as something they discover*. Call this feature of an action the

What/why condition: both that oneself is doing A and why oneself is doing it are knowable to the agent not as something they discover, while they are doing A.

In contrast, the question ‘Why did you win the lottery?’ would find no application in this way.¹⁰

When I argued above that the fist-clenching in our scenario is more than merely voluntary, I relied on the assumption that the patient counts as trying to clench their fist. It may be objected that, since the patient is at least skeptical whether they are clenching their fist,

¹⁰ Suppose someone knows how to play basketball in a very elementary way, and is not good enough to count as having the skill to sink a three-point shot. If they tried, it could be true that they *tried* to sink the shot in order to e.g. win a stuffed animal. And if they turned out to have succeeded, perhaps it can even be said that they *sank* the shot in order to win a stuffed animal. But at no point during the entire process would the question ‘Why *is* this person *sinking* a three-point shot?’ find application in the relevant way: they were never *sinking* the shot – in the progressive – in order to win a stuffed animal.

and does not believe that there is any distinct thing to do (not even summoning something in their head) or distinct know-how to exercise, in order to clench their fist, they cannot count as trying to clench their fist. This objection seems to assume that an attempt to do A must, to the subject themselves, also fall under some description other than what is conceptually contained in *attempt to do A* to be an attempt to do A. This assumption seems to me a prejudice. I was once knocked out in a road accident. I woke up, opened my eyes, still lying on the sidewalk, staring at a group of passers-by who were trying to help. They asked me if I was okay. Then I tried to say yes – but I could not. (Instead, my entire body began to convulse.) There is no other distinct thing that I was aware of doing as my attempt to say yes, but it was very clear to me that I did try. Now, further, imagine instead that I did succeed: ‘yes’ was being uttered by me – but for some reason I could not hear it myself or feel the vibration in my vocal cords, etc. Nevertheless, if so, I would have been uttering the sound on consideration of a reason: to get my message across to those passers-by. I would have been saying what I was meaning to say. I would have been the author of the utterance. Most importantly, the what- and the why-questions applicable primarily to intentional actions would be applicable here, and the what- and some why-answers would be readily accessible in me, in the sense that, during my utterance, both what I was doing and why I was doing it would be knowable to me not as something I discover.

Neither the case for nor the case against classifying the carbon-copying and the fist-clenching as intentional actions is unreasonable. That is because we are looking at fringe cases. On the one hand, they belong together with undisputed cases of intentional action in that the questions ‘What?’ and ‘Why?’ in the sense I described in the two preceding paragraphs apply to them all: they all meet the what/why condition. On the other hand, they are only fringe cases because the what- and why-questions in that particular sense, as we know them, would not be part of our discursive repository unless they had application to cases of doing A where the agent exercises a capacity for doing A in a context familiar to them in every way for exercising it, all transparently so to the agent. When someone is doing A intentionally in normal circumstances, they are voluntarily doing something which they have been able to do and about which they have no reason to be diffident now, all transparently so to this agent. Call this feature of an action the

Familiar exercise condition: that the agent exercises a capacity for doing A in a context familiar to them in every way for exercising it, all transparently so to the agent.

Unlike the what/why condition, the familiar exercise condition is not present in all cases where the what- and why-questions apply: it is optional. But like the what/why condition and unlike some other optional features, the familiar exercise condition is a constituent of what gives the *focal* sense of those what- and why-questions, i.e. an attribute of the paradigm of intentional action, in virtue of which intentional action as the kind of thing we know is possible. I will explain in §3.7 how and why the familiar exercise condition, albeit optional, belongs to the paradigm. The contested cases are fringe cases, then, because what supplies the focal sense, i.e. the paradigmatic set of attributes, is not entirely present in them. As long as the reader agrees with my remarks about the two conditions so far, it is immaterial to me if they reserve the title ‘intentional’ to actions that meet both conditions. However, I will continue to call every action that meets the what/why condition intentional, for the rest of this dissertation. Indeed, as I later (§4.4) expound the further conditions that must be met for the action to be free, the sufficiency of them will make it clear that the familiar exercise condition is not one of them either.

It is time to sum up the results so far. First, the carbon-copying and fist-clenching scenarios show that, although an intentional action may be independent of this or that particular happening as its worldly involvement, this independence is not unlimited. Second, it remains unchallenged that, if there is any truth to *Doing-knowingly* (that to do A intentionally is also to do it somehow knowingly), it would not entail that the agent has no doubt that they are doing A. The Naïve View is false because:

Certainty-optional. The agent intentionally doing A need not be certain that they are doing A.

Third, even though the doings under study do not meet the familiar exercise condition, they meet the what/why condition, which seems to *be* the applicability of the ‘What?’ and ‘Why?’ that applies primarily to intentional actions. They are applicable to a bit of conduct just insofar as it satisfies that condition.

I have been mentioning a way in which something is knowable or is known not as something the knower discovers. In §3.5.2, I will flesh out this alternative, giving positive charac-

terizations of the way a fact is knowable to someone *as* the subject of the knowing. (More specifically, for the domain presently under study, where the knowable fact in question is that someone is doing something, this fact is knowable to someone *as* the doer featured in the knowable fact.)

3.3 IN POSITION TO RECOGNIZE ONE’S DOING NOT SOLELY BY DISCOVERY

In the previous section, we considered one construal of the idea (*Doing-knowingly*) that to be doing A intentionally is also to be doing A knowingly, namely, the Naïve View that the agent doing A intentionally, as such, invariably knows that they are doing A. We ran into difficulties. Here we consider another construal of *Doing-knowingly*. The defender of this construal can agree that the agent doing A intentionally may be rationally skeptical whether they are doing it, and that they cannot know that they are doing it without observational or other discovery-based knowledge. These notwithstanding, they claim that

Involvement-idiognosis. When a subject is intentionally doing A, insofar as they know that a happening they intend to be part of the doing (i.e. an involvement) is happening, they know, not solely by discovery, this happening as due to their own doing.¹¹

This means that, if the agent is doing A intentionally, when they have whatever discovery-based knowledge they need to know that an event is taking place, and if this event is one they intend to be part of their intentional doing A, they recognize this involved event as such and they know this fact of authorship not as something they merely discover. According to *Involvement-idiognosis*, then, so long as I am intentionally raising my arm, for instance, if I know even only by discovery that the arm is rising, I know not only by discovery that its

¹¹ The idea that agential knowledge may be a sort of self-knowledge in potentiality was originated in an attempt to interpret what Anscombe calls practical knowledge, especially in the context of the following passage in *Intention* §28: “. . . in so far as one is observing, inferring etc. that Z is actually taking place, one’s knowledge is not the knowledge that a man has of his intentional actions. By the knowledge that a man has of his intentional actions I mean the knowledge that one denies having if when asked e.g. ‘Why are you ringing that bell?’ one replies ‘Good heavens! I didn’t know *I* was ringing it!’ ” (2000, 50-51). This particular version (*Involvement-idiognosis*) of the general idea, especially as an interpretation of the passage, is due to McDowell (2013).

rising is due to my raising.¹² Note that the very intentional doing of A itself also counts as an involvement. So, it is part of the present view that, if the agent is doing A intentionally, then, when they have whatever theoretical knowledge they need to know that they are doing A intentionally, they recognize this doing as such and they know this fact of authorship, i.e., that *they* themselves are doing A intentionally, not as something they merely discover. This is true, its proponent might say, because doing something intentionally already places the agent in a position, in potentiality, to have this knowledge of action authorship. This potential is realized as soon as they know that whatever sub-event needs to be happening is happening, even if this latter knowledge is purely a matter of discovery. If we think of being in such a potential or position for the special authorship knowledge as how one is doing something ‘knowingly’, then we may treat *Involvement-idiognosis* as a gloss on *Doing-knowingly*: to be doing something intentionally is also to be doing it knowingly.

One may object that, sometimes, there is no other way than discovery to know precisely that one is the doer of what one is doing intentionally, let alone to know that its involvements are taking place. One could be doing something intentionally even if one has to find out to know that one is the doer.

Suppose I, recovering from paralysis, try to lift and hold my arm up while a nurse is also holding it. Suppose I see that arm go up and stay in the elevated position I desired but – because, say, I haven’t recovered all of my proprioception – I am not certain whether I myself moved it or only the nurse did.¹³ Even if I later find out that I myself did it, I may very well say, ‘I *didn’t* know I was doing it’, but this doesn’t seem to deny intentionality.

The defender of *Involvement-idiognosis* may rule this lifting of the arm out as an intentional action by the following argument. Insofar as the subject does not recognize the rising of the arm as due to their own lifting, as an involvement of their lifting, it cannot actually be an involvement of their lifting. Given that they are not lifting their arm intentionally unless their arm is rising *as its involvement*, they are not lifting the arm intentionally.

However, as before, the what- and why-questions that primarily apply to intentional actions apply to this lifting. The what/why condition is met. If my nurse Ned, feeling my

¹² McDowell is committed to this: so long as I am intentionally raising my arm, if I know even only by discovery that the arm is rising, I know not by discovery that its rising is my doing (personal communications), which seems to me an even stronger claim than the view discussed in the main text above.

¹³ I thank Karl Schafer for the original suggestion of this scenario.

arm move of its own accord, asks me, ‘Why are you moving your arm?’, this has application because I am in a position to say ‘I am? Good. I have been trying to see if I have recovered from paralysis.’ During my lifting of the arm, both what I am doing and why I am doing it is knowable to me not as something I discover, and knowable to me as the very one who is lifting the arm.

This shows that the what/why condition does not require that the agent be certain about their authorship over all of the involvements of their action when they discover their occurrence. However, a weaker view implied by *Involvement-idiognosis* has remained unscathed by the thought experiment:

Doing-idiognosis. When a subject is intentionally doing A, if they know that they are doing A intentionally, they know, not solely by discovery, this doing as their own.

Now, the crucial stipulation in the scenario is that, even when one notices that one’s arm is rising, one still needs to *find out* to be certain whether this arm is being raised *by one oneself*. But such need for finding out does not contradict *Doing-idiognosis*. For, so long as the patient still needs to find out whether their arm is being raised by themselves, they are not yet certain that they are raising their arm. Obviously, as is clearly assumed by the way the scenario is used, knowing one’s arm to be rising does not per se amount to knowing that one is *raising the arm*.

As with *Involvement-idiognosis*, what makes *Doing-idiognosis* plausible is that doing something intentionally already places the agent in a position, in potentiality, to have the knowledge of action authorship. This potential is realized as soon as they know that whatever sub-event needs to be happening is happening, even if this latter fact is knowable only as something they discover.

Of course, insofar as I cannot know the fact that I am doing A without finding out some distinguishable fact, and insofar as it is an aspect of the former fact (or the former fact but under another description) that I am the author in this doing A, I also cannot know that I am the one doing A without also finding out some fact. But it does not follow that, when I do know that I am doing A, I know it in every aspect, including the aspect that I am the author in this doing, just by finding out. For even if I cannot know a certain fact without knowing something conceptually distinct in a specific way, it does not follow that, when I

do know the former, I know it in all its aspects just in that way.¹⁴ Non-practical examples that illustrate this point are abundant. E.g. even if I cannot know without observation that something in front of me is either living or not living, it does not follow that, once I know that that thing is in front of me, I know solely by observation that it is either living or not living.

It may appear that, e.g. when this patient finds out from the nurse's inquiry that they themselves raised the arm, they figure this out *entirely* on the basis of reasoning from evidence, which would leave no room for non-discovery-based knowledge. However, if receptive knowledge and theoretical reasoning from it were all that was available to this patient, it would be more reasonable for them to infer e.g. that their arm moved on its own. In contrast, when they infer that they themselves raised the arm, we find this inference more reasonable of them to make than inferring that the arm moved on its own. This suggests that they are inferring at least in part on the basis of something knowable to them not as something they find out.

Doing-idiagnosis gives us one sense in which to be doing A intentionally is also to be doing A knowingly: the subject, just by doing A intentionally, is in a ready position to know that they are the one doing it and to know this not as something they merely discover. It also sheds some light on the putative idea that, even when an intentional action has worldly involvements, the way it is knowable to the agent is not exhausted by discovery.

3.4 TWO DESIDERATA ABOUT THE NON-THEORETICAL INTELLECTUAL RELATION THAT ISN'T KNOWLEDGE

In §3.2, we saw that the Naïve View is false: an agent doing A intentionally is not thereby certain that they are doing A or that its involvements are happening. Yet, the what/why condition seems to obtain insofar as a what- or a why-question can be asked of an agent for knowledge of their intention or justification. So, this much seems true:

¹⁴ Cf. this famed statement from Kant's *Critique of Pure Reason*: "But even if all our knowledge begins with experience, it does not thereby just all arise out of experience" (B1, my translation). Even if a particular piece of knowledge must be occasioned by *Erfahrung* – experience, finding out –, that does not mean that, when acquired, what is known is given entirely in the *Erfahrung*.

What/Why-Knowable. While a subject is doing A intentionally, that this is what they are doing and why they are doing it are knowable to them not as something they discover.

I also suggested that the familiar exercise condition, though seemingly optional for intentional action, is part of the paradigm of intentional action. In §3.3, we saw a counterexample to *Involvement-idiagnosis*, but *Doing-idiagnosis* seems true.

Recall the first two of the leading questions in this chapter: (i) What intellectual relation does an agent as such invariably bear to their intentionally doing A? (ii) What is the nature of this intellectual relation *so that* this agent's authorship extends over happenings that take place not only in their head? *What/Why-Knowable* says: that A is what they are doing and why they are doing it are knowable to them not as something they discover. *Doing-idiagnosis* says: this agent's being uniquely in a position to have knowledge of their intentional doing is such that the knowable thing here, when known, is not something they merely discover, namely, that they are the one doing A. Together, they say that the agent is in a position to know what that doing is, why it is happening, and whose doing it is, all not as something they discover. Two desiderata remain. First, given that it is optional to know that one is doing what one is doing intentionally (*Certainty-optional*), why can 'I didn't (actually) know I was doing A!' *ever* squarely deny that one was doing A intentionally? This is not explained by the view that the subject doing A intentionally knows *potentially* that they are the one doing it. Second, it still remains obscure what *actual* relation the subject invariably bears to their intentional doing, so that it is correct for *What/Why-Knowable* and *Doing-idiagnosis* to ascribe to the agent a potential for non-discovery-based knowledge of the what, why, and whose of the intentional doing. With an eye towards these desiderata I will identify the actual intellectual relation in which the subject stands to their doing A intentionally.

3.5 DESIDERATUM 1: HOW DISAVOWAL OF KNOWLEDGE REJECTS 'WHY?'

3.5.1 The payoff of finding desideratum 1

Sometimes, we intend to do and intentionally do what is transparently difficult (Davidson 1978, Marušić 2015, 80-96). Sometimes, it is not hard, but its easiness or even probability is not transparent to us (Setiya 2008). In sum, we do things intentionally not always in a context that is transparently familiar to ourselves. Therefore, it is hard to deny that the agent does not always know that they are doing what they are doing intentionally. Nonetheless, many recent philosophers (e.g. Anscombe (2000), Falvey (2000), Thompson (2011), Ford (2011), Marcus (2012)) have been driven to the idea that practical knowledge is essential to intentional action. Among them, those who concede that the agent does not always know that they are doing what they are doing intentionally construe such cases as derivative from paradigmatic cases, in which the agent does know, non-receptively, that they are executing their intention. These secondary cases, it is said, owe their membership in the class of intentional action to the paradigm. But how this is has not been sufficiently explained, given that practical knowledge is absent in those allegedly fringe cases.

Given this lacuna, it is not unreasonable for one to remain skeptical whether practical knowledge is really central or essential to intentional action. Why should the absence of knowledge make an intentional action second-class? Even if practical knowledge sometimes obtains, why isn't it just icing on the cake of intentional action?

In a recent, actual example, Berislav Marušić expresses such discontent:

On Anscombe's view, we act intentionally under the description under which we have practical knowledge of our action. The role of practical knowledge is thus to single out the descriptions under which we act intentionally. But if intentional action is possible in the absence of practical knowledge, then something else must be capable of playing this role. And if something else is capable of playing this role, then why should we think that practical knowledge, rather than this other thing, is the crucial notion for understanding intentional action? (2015, 95)

My aim is to fill this lacuna about the role of practical knowledge. One thing that is actually present in *all* and only cases of intentional doing is the what/why condition,

which implies that the agent bears a non-discovery-based intellectual relation to their doing – even when this relation falls short of knowledge. They must represent the action fact, that they are engaged in the intentional action, in such a way as to constitute this very fact. In the literature, this practical spontaneity in the absence of knowledge has gone largely unobserved.

An indication that such spontaneity is present even in the fringe cases can be found in our practice. ‘I didn’t know I was doing A!’ can squarely deny that one was doing A intentionally, rejecting the question ‘Why?’ *even* when circumstances are such that, although one might have been doing A intentionally, one could have been so acting only with doubt that one was doing A and so would not know that one was doing A. In §3.5.2, I distinguish between two circumstances that make it true to say ‘I did not know’: having no idea and merely not being certain. Having no idea precludes a certain knowing-*of* that is non-discovery-based and grounded in a spontaneous intellectual relation the agent bears to their intentional doing. In §3.6.2, I will argue that they must bear this relation to some why, not just the what, of the doing. For such an intellectual relation to obtain, the agent must know of an answer to the why-question that is *non-accidentally and non-theoretically true*, whether or not they know that it is true. But such non-accidental truth on its own, considered separately from knowledge, is metaphysically dubious. In §3.7, we will see this conceptual instability in detail. From this, I will demonstrate that those fringe cases count under intentional action only because they are relevantly similar to normal cases – in which one does something of the same sort in a transparently familiar context. And in these normal cases, the agent’s spontaneous representation is knowledge as much as is our everyday success in exercising our capacity to know by discovery.

3.5.2 Desideratum 1, found: the agent’s non-knowledge relation is a practical thought of doing A

Let us pick up the first desideratum. Suppose what we concluded in §3.2 is true, that (*Certainty-optional*) someone doing A intentionally may not be certain that they are doing it. Grant that knowledge that one is doing A is optional for intentionally doing A. Well,

then, why may disavowing knowledge that one was doing A ever be a way of denying it as one's intentional action? Does this practice just reflect sloppy, inaccurate reasoning?

I think the answer is no: there are cases in which the disavowal, 'I didn't know I was doing that!', is a way of denying that one is doing A intentionally on the basis of sound reasoning. For there are at least two different levels at which one may not know something:

One is *merely* doubtful or not certain about a given topic, T.

vs.

T is not a distinct topic for one at all.

Not every reason that makes it true to say e.g. 'I did not know one was poisoning the party chief' precludes the poisoning from being my intentional action. This possibility is not automatically precluded just because e.g. 'I could have been wrong that the person I was poisoning was the party chief / the ingredient was poisonous for the chief / the toxin could pass through the system / it was my toxin, rather than someone else's, that poisoned the chief'. For, even if one could have been wrong about all these, but if one was in fact right, one's poisoning of the party chief *could* still be intentional.

However, the disavowal of knowledge would be a denial of an intentional action if the reason is 'I had no idea that the person I was poisoning was the party chief', or '... no idea that the water content that turned out to poison the chief was poisonous', or '... no idea that the toxin could go through the purifying filter system I used', or '... no idea I was the one doing it', etc.

I can fail to know that *p* because of mere uncertainty whether *p* or because of my wholesale unawareness of the topic *whether p*. And the difference lies in this. If the subject is doing A intentionally, even if they are themselves not positive that they are doing it, the *topic* of their uncertainty or disbelief must not stop at 'What on earth might be happening now?' or 'Am I doing anything now?'; they must, *inter alia*, be uncertain, incredulous, etc. about, specifically, whether they are *doing A*. In other words, *doing A* must stand out to the agent who is doing A intentionally – if only as what is very unlikely to happen.¹⁵ Insofar as

¹⁵ Brian O'Shaughnessy invokes virtually the same necessary condition for precluding an event from being an intentional action (1980, vol. 2, 67): 'not, under any conceptual or descriptive heading, in any part of the mind, superficial or deep, Freudian unconscious or pre-conscious, an object either of knowledge, belief, certainty, doubt, suspicion, etc.'

the subject has no idea about their doing A, so that they have no idea that *doing A* may describe themselves, lack of knowledge in this way always means that the subject cannot be doing A intentionally. In contrast, *doing A* may still stand out, as a topic, to someone who does not know that they are doing A for the reason that they are not sure whether they are doing A. Lack of knowledge, insofar as it is merely due to uncertainty, is not enough to rule out that one is doing A intentionally.

So, our first desideratum finds a solution. The puzzle was that ‘I didn’t know I was doing A’ seems to aptly deny that one was doing A intentionally even though, to do it intentionally, one need not know that they are doing it. We now solve this by distinguishing between two reasons for the lack of knowledge. Saying ‘I didn’t know I was doing A’ squarely denies that one is doing A intentionally *when* ‘I didn’t know’ is true due to cluelessness, to the fact that the speaker did not have any idea of their doing A as a distinct possibility, in the sense that *doing A* was not on their radar as something at stake.¹⁶

This does not entail that the subject must know that they are doing what they are doing intentionally. It merely follows that the subject intentionally doing A would *not have no idea* that *doing A intentionally* might describe a thing they are doing. They must not have no idea for the following reasons. First, the subject intentionally doing A must know that they are trying to do A. The main objection to this is that sometimes one can do something intentionally without taking oneself to be doing anything (that contributes toward the intentional action), as Setiya concludes from his fist-clencher case. But I have resisted this conclusion toward the end of §3.1. Second, for a subject who is capable of

¹⁶ Note that this cluelessness, about the possibility of an action-description’s applying to oneself, is a sufficient condition but not a necessary condition to rule out that one is acting intentionally under the description. Dmitri Gallow provided the following example. Suppose two electricians are operating with a certain contraption. Suppose it is part of J’s job to press a certain button. For all that is known to Agnes, with pressing this button comes a risk, but a very low risk, of electrocuting the other operator, Siu-Kwan. But suppose Agnes pressed the button just to do her job, but it did cause the other to be electrocuted. Now, Agnes did consider this as a distinct possibility. But, if asked, ‘Why did you electrocute Siu-Kwan?’, Agnes could honestly say ‘I did not know I was doing that’ and deny that she did it intentionally.

However, I do not think in this case Agnes could say ‘I didn’t know I was doing that!’ to *squarely* deny that she did it intentionally. For even if Agnes pressed the button with the intention to kill Siu-Kwan, she could not have done it *knowing* that she was electrocuting him, given the low probability of success. It would make more sense to say directly, ‘It wasn’t intentional!’ I do not know how commonly people feel inclined to disavow knowledge to deny that they were doing something intentionally in this sort of circumstance – that is an empirical question –, but even if many feel compelled, it would not follow that it makes logical sense.

intentional action, at any given time, there is easily a whole lot of descriptions about each of which, if they are to think of it, they *cannot be wrong* that they are not acting intentionally under it. Insofar as the action-description or practical predicate, *doing A*, belongs to this lot, the subject is not doing A intentionally. But insofar as they are doing A intentionally, they also know that *doing A* is not one of those predicates, and they know this independently of modes of knowledge that involve discovering what is known. This holds, even though (*Certainty-optional*) they may not be certain that they are doing A.¹⁷

Description-aware. When a subject is doing A intentionally, *doing A* is not a description they are in rightful readiness and certainty to judge inapplicable to themselves as acting intentionally, and they non-theoretically know this.

In other words, the description, *doing A*, stands out to the agent with such distinction. Call this agential knowing-of-a-description-*doing-A* the agent's practical thought *of* doing A.¹⁸

3.5.2.1 An objection to *Description-aware* For the sake of argument, let us suppose that the following scenario is possible.¹⁹ Ab, even though he is certain that a certain boulder is way too heavy for him to move, can nonetheless try to lift it, since he knows what it means to try in this case: grasping the boulder and straining. Even as he is trying, he is certain that *lifting the boulder* does not describe anything that he is doing, and so he is judging it inapplicable to himself as acting intentionally. Thus, this action-description does not stand out to Ab as not a description which he is in rightful readiness and certainty to judge inapplicable to anything he is doing intentionally. But if Ab manages to lift the boulder – e.g. because it turns out to be hollow – he has moved it intentionally.

This appears to be a counterexample to my claim that, as long as someone is doing A

¹⁷ Why do I not say instead, at this juncture, what is more straightforward to say, namely, that the agent doing A intentionally invariably knows non-receptively that they intend to be doing A? First, I prefer the formulation in which the agent's intellectual attitude is more clearly about the doing (as opposed to intending) even in cases wherein it is short of knowledge. Second, I am wary of Michael Bratman's objection (1984; 1987) that someone doing A intentionally need not intend to do A (though they must intend to do something). As Kevin Falvey points out (2000, 41, fn.5), the video game example Bratman uses does not as clearly rule out that the agent intends *to be doing* (stressing the imperfect, the progressive) what they are doing intentionally. Still, this requires a detailed discussion that would be a digression for the point I am driving at here.

¹⁸ I mean to distinguish someone's thinking *of doing A* from someone's thinking *that they are doing A*.

¹⁹ I devise this scenario by modifying an example in Albritton 1985, 244. The original point there is that one can try to do something even though they know that it is impossible to do.

intentionally, they have a practical thought of doing A, i.e. *doing A* stands out to them as not a description which they are in rightful readiness and certainty to judge inapplicable to their intentional activities. My response is that, even if we grant everything in the alleged possibility, it does not follow that Ab, in the process of straining, counts as lifting the boulder intentionally. His straining does not count as lifting the boulder intentionally *unless and until* he comes to have the practical thought of lifting the boulder – which can happen e.g. if he feels that the boulder has actually budged and goes on straining to lift it. Until then, the question ‘Why are you lifting the boulder?’ in the relevant sense does not apply. Ab can rightly reject it by saying, ‘What do you mean? I don’t think it is possible.’²⁰ Thus, *Description-aware* withstands the objection.

Description-aware picks out an aspect of intentional action the absence of which would amount to cluelessness. This *explains* the solution we found for the desideratum; it explains why the lack of knowledge due to cluelessness squarely precludes the possibility that one is doing A intentionally.

3.6 DESIDERATUM 2: THOUGHT THAT REMAINS NON-THEORETICAL EVEN WHEN IT AMOUNTS TO KNOWLEDGE

3.6.1 Desideratum 2, found: the agent must have *constitutive* practical thought of their intentional doing

The second desideratum remains: when a subject is doing A intentionally, what *actual* relation do they invariably bear to their intentional doing, so that they have a potential for non-theoretical knowledge of the intentional doing, so that, even when they have found out what all they need to find out to have knowledge of their doing, they know facts not as something they find out? So far, we have only observed that a certain description must stand out to the agent with a certain distinction and that this standing out is well founded on some non-theoretical basis. The relation that serves as this basis will have to be a certain

²⁰ In this response, I am implying that one can have done A intentionally without intentionally doing A *throughout* the process of doing A. Cf. fn. 9.

non-accidental agreement between the description and the fact. In some cases, a description agrees non-accidentally with a fact as something discovered: the employer of the description thinks of it because it matches the fact they *find out*. What is the supposed alternative? The employer of the description thinks of it in employing it in such a way as to constitute its instantiation. Their intellectual relation to what is predicated is not other than what it is for the predicate to be true of the subject. And I shall use the modifier ‘spontaneous’ to signify (1) any way for a subject, x , to have a thought of a description, D , so that having this thought constitutes what it is for it to be the case that $D(x)$, (2) any such intellectual relation x bears to the fact that $D(x)$, or (3) any such way for a description to be true of a subject.

The agent’s spontaneous practical thought of doing A is their intellectual relation to their intentional doing with respect to its identity, the *what*, and to its plain factivity, the *that*. But this relation turns out to be inseparable from another, namely, the agent’s intellectual relation to their doing with respect to its own non-accidentality, the *why*. This latter relation has the character of spontaneity as well. We will see this in what follows.

3.6.2 The agent must bear their constitutive intellectual relation to some why

In the preceding discussion, we observed that, insofar as one is doing something, A , intentionally, they must have a spontaneous practical thought of doing A . It is not just that the description, *doing A*, must stand out to them as not among the host of descriptions that they are in rightful readiness to judge inapplicable to themselves as an intentional agent; it must stand out to the subject in such a way as to constitute what it is for the description to be true of them. Should this thought amount to knowledge that they are doing A intentionally, they will know not as a fact they discover that A is what they are doing, that they are the one doing it, and that it is being done. Thus, it is relatively clear that this non-theoretically true practical thought consists of the agent’s intellectual relation to their doing with respect to its identity, the *what*, to its propriety,²¹ the *whose*, and to its factivity, the *that*. I now argue that this relation is inseparable from the agent’s intellectual relation to their doing

²¹ I am here reviving the obsolete but attested use of the word ‘propriety’ to mean *fact of belonging specially to a particular thing or person*. See *Oxford English Dictionary*, s.v. ”Propriety, n.1b”.

with respect to its non-accidentality, the *why*.

It is true that we take people to do things intentionally but for no particular reason. That is, at least sometimes, when the agent replies, ‘For no particular reason’, we do not take them to be reporting something they found out, not even by introspection. Thus, we take it to be a spontaneous answer to the question ‘Why?’ It follows that the doing is no accident. ‘Nothing’ does not reject ‘What is in your pocket?’ as ‘What you are looking at is not a pocket’ would.²² Similarly, ‘For no particular reason’, if it is a spontaneously possessed why-answer, does not reject ‘Why are you doing A?’ as ‘What you are observing is not my intentional action’ would. Possessing ‘For no particular reason’ spontaneously counts as having an explanation²³ in a way, since it *does* specify a way in which my doing is no accident: for no particular reason. But it *barely* counts as having an explanation, because the answer affirms that there is no *explanans*. And apparently such actions, with no reason-giving why-answer to be sought, involve no exercise of practical reason.²⁴ So, why should we think that the agent’s intellectual relation to the what of their intentional action is inseparable from one they bear to some why?

First, it seems that someone who is capable of no answer other than ‘For no particular reason’ is incapable of intentional action. If they are to be capable of intentional action, they must understand a *why*-question, and they do so at all only as far as they understand (perhaps without articulate grasp of a rule) what kind of thing a legitimate answer can cite *to provide a reason*. And even if they are intentionally doing something for no reason, they must acknowledge that the doing is the sort of thing that exists because of its role in certain possible stories that contain it. In particular, they must acknowledge that their acting, in its queried aspect, is the sort of doing that can figure in a story that reveals the point of the doing by reference to something else. In particular, they understand the why-question of the sort under investigation – that which marks an intentional action – only as far as they understand reason-giving answers of a specific class, namely, answers that reveal what Anscombe might call motives. All reason-giving answers to our why-question reveal motives:

²²A modification of Anscombe’s analogy in [Anscombe 2000](#), 25, §17.

²³ In this dissertation, I use ‘explanation’ to mean an idea supposed to specify how some given fact, the explanandum, is not an accident. I mean the entirety of the idea expressible in the form ‘[explanandum] because *p*’, not just what is statable in the ‘because *p*’ clause.

²⁴ I thank Stephen Engstrom for pressing me on this.

they all interpret the queried doing, putting it in a certain light. A motive-revealing answer may cite something in the future relative to the doing (e.g. ‘in order to win their favor’, ‘because I won’t see them again’), something in the past relative to the doing (‘because they did me a huge favor’), or something that is neither (‘out of love’). Further, the agent must understand certain motive-revealing answers as apt in and only in certain circumstances for a given course of action. This understanding includes the ability to abort what one is doing or revise one’s why-answer in view of new evidence. Only from such a repository of answers would the agent’s possession of a specific one distinguish their doing from doings with etiological facts they merely encounter. I call something etiological insofar as it is about whether some given fact is explicably non-accidental and, if so, how. Thus, an explanation is the etiological representation that some given fact *is* explicably non-accidental in a specified way.

Second, more specifically, if someone is engaged in some behavior for no particular reason *and* this indeed involves no activity of their sensitivity to reasons, this thing that they are doing could be very unreasonable for an intentional action, and it would be just luck if it is not so unreasonable. If it appears very unreasonable, but the agent maintains that they are doing it for no reason – especially if this is a recurring phenomenon with the individual –, one begins to suspect that either they are lying or this behavior is ‘for no reason’ in the sense that, like such merely mentally caused acts as “starts and gasps”, it could not have been performed for a reason. In the latter case, the agent does not possess the answer spontaneously – i.e., it affirms something as a fact the agent just discovers. Deployed thus, ‘For no particular reason’ would just be saying that the act is random, an accident, with respect to whatever practical reasons there are for acting one way or another. This would deny the why-question application as ‘Because I was startled by the clap of thunder’ would. This suggests that it makes sense to accept ‘For no particular reason’ for an intentional action, typically, only on the assumption that the agent perceives that there is no strong reason against their present doing or, at least, is somewhat receptive to reasons that bear on their present situation, if there are any. In short, ‘For no particular reason’ seems to fit something as an intentional action only if the speaker is also ready to say ‘Why not?’ of it. Even an intentional action for no particular reason, then, seems to issue from the power of

reason as one of the optional things that are not patently unreasonable to do.

Third, if the action does not involve activity of reason even minimally, nothing seems to distinguish it from other acts knowable to the subject without observation, as it is briskly argued in [Anscombe 2000](#), 32, §20. This suggests that a doing for no particular reason is an intentional action in the sense we are familiar with *only* if it is still a manifestation of the agent's power of reason. If there is to be an intentional action for no particular reason, 'for no particular reason' here has to be an answer, albeit a marginal answer, to the *why*-question that elicits motives. Suppose we bracket the practical-rational power, and suppose the why is removed from the what/why condition as a necessary condition for intentional action, so that the condition is shrunken to *that A is what one is doing must be knowable to the agent not as something they discover*. There would be nothing left to determine whether a given act of doing A that is knowable to the subject without observation meets or does not meet this condition. So, if the agent bears a spontaneous relation to the what of their doing, they must bear it also to some why at least in two senses: they constitutively represent the doing's etiological fact (even if it turns out to be bare), and the doing issues from their power of reason as one of the optional things that are not unreasonable to do.

Hence, to bear a spontaneous intellectual relation to the what of one's intentional action is inseparable from bearing such a relation to some why. That is, to have a spontaneously true practical thought of doing-A is inseparable from having a spontaneously true practical thought of doing-A-because-*p*,²⁵ even when '*p*' is 'I just thought I would'. That is, *doing A because p* must stand out to the agent in such a way as to constitute what it is for the answer to be true. Thus, the why-question in the relevant sense has application to someone's doing insofar as this agent's possession of a practical thought of doing-A-because-*p* is spontaneous. Insofar as a representation (e.g. an explanation) is true, and thus what it represents is a fact, in virtue of the way in which the subject possesses this very representation, it is a spontaneous representation.

So, we have identified, with a positive characterization, an *actual* intellectual relation in which the agent invariably stands to their concurrent intentional action, the actual relation

²⁵With 'doing A because *p*' I mean to leave room for cases in which the agent is doing A intentionally but for no particular reason: 'I am doing A because I thought I would.'

that explains the agent's potential for non-theoretical knowledge, the relation that underlies the way in which what they are doing and why they are doing it are knowable to them not as something they discover.

3.7 SPONTANEOUS PRACTICAL THOUGHT POINTS IN THE DIRECTION OF *KNOWLEDGE*

I said that the intellectual relation that gives application to the characteristic why-question consists in a practical predicate's standing out to the agent as not a description they are in rightful certainty and readiness to judge not applicable to their intentional behavior. Negatively characterized, seemingly minimal, and short of knowledge, how could such a 'standing out' constitute the fact that the predicate is true of the subject? The spontaneous practical thought of doing-A-because-*p* is conceptually in suspense and dubious unless it is anchored in spontaneous knowledge somehow, and there is more than one determinate way to think of this conceptual 'anchoring'.

3.7.1 The 'split view'

On the one hand, the questions what *x* intends or is trying to do and why *x* is trying to do A, in a special sense, always apply to *x*'s intentional doing-A: the agent invariably has a true what-answer and a true why-answer in some immediate, first-personal way, without finding out. On the other hand, as we observed above, sometimes the agent who is doing something intentionally does not even believe they know that they are doing it. These two considerations might motivate the following, split conception of the agent's intellectual relation to their intentional action. The agent is doing A (because *p*) intentionally only if they are or have been in a mental state of intending or trying to do A because *p*. Some philosophers believe that intending to do A (because *p*) or trying to do A (because *p*) are such that one can be in such a state with definite content or meaning even if they are not doing A and never begin to. According to the split view, the agent doing something

intentionally can and typically does know, not by discovery, what they intend or are trying to do and why – but this is only knowledge that they are or have been in some psychological state that does not entail any actual doing A. When it comes to whether they are actually doing or not doing A, if they ever know, they know this by observation or some other way of discovery.²⁶ This split view resolves the unsettling suspense – about the curious concept of a spontaneous practical thought – by allowing some non-theoretical knowledge to the agent of intentional action, but limiting the possibility of such knowledge to something less than the intentional action, perhaps a supposed element in it, that is independent of the doing. A spontaneous practical thought of doing-A-because-*p*, on this view, would be construed as just part of the agent’s knowledge that they are themselves in a purely mental state with the content, *doing A because p*.

3.7.2 Why must the agent’s spontaneous practical thought of doing-A-because-*p* be conceptually anchored in spontaneous practical knowledge that embraces the doing

Kevin [Falvey \(2000\)](#) resists this sort of split conception by appealing to the continuity of intending to be doing A, which the intender always knows without observation, and doing A intentionally. It belongs to the concept of *intending*, he argues, that in its central instances intending to be doing something and doing that thing are just one reality in two aspects, so that knowledge of the former amounts to knowledge also of the latter. His project is to show, first, how the intention to be doing something is conceptually dependent on intentionally doing something and, second, how first-personal knowledge of the former is conceptually dependent on first-personal knowledge of the latter. What I am going to show here is different, namely, how the spontaneous practical thought of doing-A-because-*p* depends conceptually on first-personal knowledge that one is intentionally doing something for some reason. But a consequence of what I am going to show coincides with Falvey’s view: the split conception

²⁶ This split conception is in essence the same as what is named and resisted in [Falvey 2000](#) as the Two-factor View, which is defended e.g. in [Adams and Mele 1989](#) (the View is stated in 516-7). Cf. also [Grice 1971](#) and [Langton 2004](#). However, this view in [Adams and Mele](#) is formulated to concern what knowledge the agent can have only of the that and the what: the fact that they are doing something intentionally, whereas my target view is formulated to concern also what knowledge the agent can have of some why: the *etiological* fact that they are doing something intentionally *because ...*

is not viable because the spontaneous knowledge that the intentional agent as such has, in some cases of intentional action, must embrace the actual intentional doing.²⁷

On the basis of three premises below, I demonstrate how exactly the more generic notion of spontaneous practical thought is conceptually anchored in spontaneous knowledge that embraces the actual intentional doing.

1. My spontaneous practical thought of doing-A-because- p is also a constitutive intellectual relation I bear to the fact that I am doing A.
 2. A constitutive intellectual relation I bear to the fact that I am doing A, in a paradigm case of intentional action, involves what is transparently the activity of my capacity to do A in a context familiar in every way relevant to doing A.²⁸
 3. What is transparently the activity of my capacity to do A in a context familiar in every way relevant to doing A is spontaneous knowledge that I am doing A.
- ∴ The agent's spontaneous practical thought of doing-A-because- p , in its paradigm case, involves spontaneous knowledge that they are doing A.

And I now elaborate on each premise in turn.

3.7.2.1 Premise one: My spontaneous practical thought of doing-A-because- p is also a constitutive intellectual relation to the fact that I am doing A. So far, it has remained an observation that, insofar as someone is doing A intentionally, they must thereby have practical thought of doing A. We may now see why it is so.

²⁷ But the way my demonstration entails this is different from how Falvey argues for it. For example, my reasoning below will neither suppose or involve the claim that someone intentionally doing A always or paradigmatically intends to be doing A.

²⁸My appeal to the notion of the agent's capacity to do A here is similar to Setiya's appeal to the notion of the agent's knowing how to do A to make sense of how the agent can have more than accidentally true belief of the success of executing their intention. However, the practical capacity to do A seems to play for Setiya a more central role in distinguishing the agent's cognitive status from a theoretical one than the role I conceive it to have. For him, whether someone's cognitive status or knowledge is theoretical depends on the manner in which some belief involved is not merely accidentally true, and the exercise of one's knowing how, which for him is a disposition to act on the belief involved in one's intention, is what makes the true belief not theoretical (2017, 14). On my view, one must also be using a capacity to do A if they are to count as doing A intentionally. But what makes the agent's representing of their doing A non-theoretical is its spontaneity, i.e. its constituting what it is for them to be doing A. This spontaneity can be present in activities that do not involve the exercise of a know-how in Setiya's sense. It is present, e.g. in one's knowledge that one is making an inference (practical or theoretical). Even when one is drawing a theoretical conclusion, one can have knowledge that one is doing so that is not theoretical, but constitutive.

As we have also observed, insofar as someone is doing A intentionally, this doing must be suitable for a certain why-question. That is, the agent must spontaneously possess a why-answer, i.e. they must have a practical thought of the etiological fact, as regards why they are doing A. In this way, intentional action must involve the operation of reason, the power that copes with the *why*. I mentioned and briefly argued above that this intellectual relation, in which I stand to what I am doing when I am doing A (identity) and to the fact that I am doing it (factivity), is inseparable from the one in which I stand to how this doing is explicable (non-accidentality). Now we see another aspect of this inseparability: the former must be spontaneous because the latter is. That is, insofar as one spontaneously possesses an answer as to why they are doing A, they also spontaneously have the practical thought of their doing A.

The truth of the preceding proposition cannot be taken for granted on the ground that any proposition as to why p entails the proposition that p , though this entailment holds. Not even if we grant in addition that I think, such and such is why p , only if I think that p . For, despite all that, it is not true across the board that *the way* I intellectually relate to the fact that p must be the same as how I relate to the etiological fact why p . For example, Vis may have a thought as to why he dropped his plate – because he was startled by thunder – non-observationally, even though he has the idea *that* he dropped his plate only by observation. However, insofar as I relate *spontaneously* to at least one etiological fact about the explanandum that p , it seems that I must enjoy *some* spontaneous relation to this fact that p . For, in this case, the intellectual relation I bear to the etiological fact, say, p because c , constitutes its being the case that c is (at least part of) why p . Therefore, by constituting the etiological connection, such an intellectual relation to why p contributes also to its being the case that p . If I did not bear such an intellectual attitude, the etiological nexus in reality would not hold, and the explanandum end of this nexus, i.e. the actual token of fact, that p , would not obtain either, even though another fact token might obtain that is describable also by ' p '. Thus, I must spontaneously be in some intellectual contact with the fact that p , so long as I in fact relate spontaneously to its etiology. This applies to our topic of action in the following, specific way. Insofar as I spontaneously stand in an intellectual relation to the etiological fact that I am doing A because p , I am also spontaneously in some

intellectual relation to the explanandum fact, that I am doing A. The intellectual relation I bear, in this case, is a spontaneous practical thought of doing-A-because-*p*. That is to say, the description, *doing A because p*, stands out to me as not among the host of descriptions that I am in rightful readiness to judge untrue of myself as an intentional agent, and it stands out in a way that is part of the fact that I am doing A because *p*. Therefore, this standing out is also part of the fact that I am doing A.

Here we see a connection between, on the one hand, the activity of the power to spontaneously apprehend one's doing with respect to its *why* and, on the other, the spontaneous relation one bears to it with respect to its *what* and *that*.

3.7.2.2 Premise two: Constitutive intellectual relation to the fact that I am doing A, in a paradigm case of intentional action, involves what is transparently (to me) the activity of my capacity to do A in a context familiar in every way relevant to doing A. The preceding consideration leaves us at a thought with the following implication. Suppose Donald is making ten carbon copies at once intentionally in order, say, to amuse his children.²⁹ And by 'making ten carbon copies at once' we mean here a narrow sense: a bit of script in fact *has appeared* on all ten underlying sheets. Even if Donald is rightly skeptical about the reach of the pressure he is exerting with his pen, the predicate, *making ten carbon copies to amuse my children*, stands out to him as not a description that he is in rightful readiness and certainty to judge as not describing an intentional action he is engaged in, and it stands out in such a way as to constitute what it is for Donald to be making the ten carbon copies. The possibility of such a case ought to strain our comprehension, unless there are cases in which Donald is rightly confident in succeeding in what he is trying to do. It is hard to see how the intellectual relation Donald bears can be *constitutive* of his doing if in all possible cases of doing he is either skeptical or wrongly confident.

Suppose, for Basti, to sing a named pitch solely 'from his head' is simply an uncertain business – perhaps because e.g. he was not born with this ability, he has not learned it, and perhaps there is simply no skill of doing so learnable by Basti. He can *try* to sing the pitch I name since, after all, he can at least sing. But any attempt he makes at singing a named

²⁹ A modified version of an example in [Davidson 1978](#).

pitch, *even before any further detail is taken into account*, has no better prospect than an attempt, a trial. Insofar as this is so, then even supposing Basti is singing intentionally and the pitch sung is the pitch I had named, he is not singing the named pitch intentionally. He is not capable of singing a named pitch ‘from his head’ as an intentional doing as we know it.

A successful, on-going attempt at doing A can be an intentional doing-A even if the agent themselves is uncertain about the success. But such a case is possible only because it is relevantly similar to normal cases of doing A – cases in which the agent succeeds in effecting relevant movements and there is no reason for them to be doubtful of their efficacy. There is no cause for such concern because the agent is self-transparently utilizing their capacity to do A in a familiar context. Only with resemblance to normal cases like this can their other attempts to do A pass as intentionally doing A. For example, Donald’s uncertain attempt passes as making ten copies at once intentionally – even if ten copies are indeed getting made all at once – only if it can be said that, normally or generally, he knows how to make a carbon copy with a thin layer of stuff between it and his writing utensil. It is just that, given certain further relevant details of a case – e.g. that the number of copies is specifically ten, which is a few more than the maximum number of copies Donald has ever made before – it would be reasonable for him to doubt if his exercise of the *same* capacity is successful this time (namely, on each of the ten underlying sheets). Similarly, in the case of Setiya’s recently paralyzed patient (2008), their attempt passes as clenching their fist intentionally only if it can be said that, normally or generally, it is transparent to themselves that they have the capacity to clench their fist. It is just that, given what the doctor has told them, it is reasonable for them to doubt if their exercise of the same capacity is successful now. In both scenarios, it is indispensable that, generally or normally, the agent has the ability to do something of a sort that is instantiated by what they are intentionally doing now – it is just that now the context in which they exercise the ability is not entirely familiar and transparent. Furthermore, in both scenarios, the doing is intentional only given that the relevant *success* has no etiology beyond the range of etiology allowed for the parallel success in the normal cases. Davidson’s making ten carbon copies stays within this confine insofar as it is just an easily conceived extension of the same mechanism he would normally deploy to

make e.g. five copies. The fist-clenching by Setiya's patient stays within the confine insofar as it is the same old operation of their motor skill (but, now, with proprioception and all other means of feedback subtracted).

I have just presented what I take to be a necessary condition for a course of action to rate as a paradigmatic case of intentionally doing A, namely, that it consist in what is transparently the agent's exercising their capacity to do A in a context familiar in every way relevant to doing A. One may object that this necessary condition is too strong, for the paradigm must allow cases in which the agent has to and does deliberate overtly on how to do something, decides on a plan, and then takes the chosen means according to the plan. For example, a baker, Kail, who wants to make vegan scones but has never made them before has to devise a recipe by deliberation. But in this process, the objection goes, this agent would not count as having a capacity to make vegan scones and there is not yet any such thing as a context that is familiar to him for exercising that capacity.

My reply is as follows. What motivates this objection is the true idea that doing A for some reason and, specifically, doing A in order to do something else are both part of the paradigm of intentionally doing A, for each feature either is shared by all cases of intentionally doing A or explains why some cases lacking the feature are nonetheless cases of intentionally doing A. But *insofar as* Kail is making vegan scones by executing instrumental actions he has decided on by necessary planning, this aspect does not belong to paradigmatic intentional action. For a case is a paradigm of intentional action only *insofar as* it has attributes shared by all cases of intentional action and attributes that explain why some cases lacking those attributes are nonetheless cases of intentional action. But doing A by executing instrumental actions decided upon by necessary planning neither is shared by all cases of intentionally doing A nor explains why some cases lacking this attribute are nonetheless cases of intentionally doing A. That said, if Kail is applying his skill of e.g. kneading in a context familiar to him in every relevant way and he is doing it according to his plan to make vegan scones, my necessary condition does not exclude his kneading and kneading-in-order-to-make-vegan-scones from being a paradigm of intentional action insofar as he is exercising his capacity to do just that – knead – in a transparently familiar context. Thus, my account does not exclude from the paradigm the fact that the agent is doing one thing with the

intention to do another thing or that they are doing something for a reason. Further still, my necessary condition does not preclude the agent's needful and overt deliberation from being part of the paradigm.

3.7.2.3 Premise three: What is transparently the activity of my capacity to do A in a context familiar in every way relevant to doing A is spontaneous knowledge that I am doing A. So far, I have argued that, if an agent stands in an intellectual relation to the fact that they are doing A intentionally that is spontaneous but short of knowledge, that is possible only as derivative from the case in which the agent is self-transparently exercising their capacity to do A in a context familiar to them in every way relevant to doing A. In these normal cases, I now argue, it is at least no less reasonable to take the agent to know that they are doing A than to take someone to know that p when they are self-transparently exercising their capacity for theoretical knowledge that p in a relevantly familiar context. By that I mean, for example, the case in which I make the correct judgement that the coffee shop *Voluto* is going to close at 7 p.m. today, based on my knowledge of their weekly schedule, my (background) awareness that no special holiday is being observed today, etc. Unless one is prepared to renounce all such everyday cases of discovery-based knowledge, it is unclear on what ground one could withhold, selectively, the title of knowledge from the everyday success achieved with a capacity to do things. And insofar as I do not just find myself exercising the capacity to do something, the knowledge has to be spontaneous.

3.7.2.4 Conclusion Given the three premises above, it follows that my spontaneous practical thought of doing-A-because- p , in its paradigm case, involves spontaneous knowledge that I am doing A. This conclusion contradicts the split view, according to which the agent's spontaneous knowledge, invariably, embraces only some of their mental states *as opposed to* what they are actually intentionally doing.

3.7.3 How the fringe cases are derivatives of the paradigm

I now further specify how the fringe cases hang together with the paradigm as the latter's derivatives. Recall that, whenever an agent is doing A intentionally, they invariably have a spontaneous practical thought of doing-A-because-*p*, which is their constitutive representation of the etiological fact. As my explanation of premise one shows, the spontaneity of this practical thought accounts also for the spontaneity of the practical thought of doing A. Furthermore, whether the agent knows they are doing A or not, they pass as doing A intentionally only if they are exercising some capacity that is the same as one that they would exercise in normal cases to do something of relevantly the same sort (e.g. in the case of Donald's making ten carbon copies intentionally, the capacity to make one carbon copy) – as I explained while elaborating on premise two. And they must be self-transparently exercising it in a context familiar enough to them for them to count as *trying* to do A. All these apply in paradigm and fringe cases of doing A intentionally alike. In a paradigm case of *doing A* intentionally, in which the agent possesses the capacity to do A and they are self-transparently exercising it in a context familiar for exercising it, their spontaneous thought of doing A is spontaneous knowledge that they are doing A, as I argued in defending premise three. When the context is not transparently familiar to the agent as one of a sort in which they exercise the capacity to do A, the agent still possesses a spontaneous practical thought of doing-A(-because-*p*), but its factivity, the that, is opaque to them, and so their possession of the what and why falls short of knowledge.

3.7.4 The split view is wrong even in a fringe case of intentional action

For all that we have observed, the split conception may still appear tempting for the fringe, secondary cases. For, even in these cases, the question 'Why are you doing A?' still applies, since the agent intentionally doing A invariably has a why-answer that is true spontaneously. The agent doing A intentionally thereby has spontaneous knowledge why they are doing it – but this is knowledge only of some mental state(s) they are in *only*. When it comes to whether they are actually doing or not doing A, if they ever know, this fact would be something they discover.

I argue that the split conception of the agent's intellectual relation is not true of such secondary cases either: in particular, I think the last statement in the previous paragraph is false. This can become clear if we attend to the fact that, if the agent qua agent ever knows that they, at the very present, are doing A intentionally, they cannot know this fact as what is affirmed in the conclusion of an inference.

When I am doing A intentionally, I must possess a why-answer, *I am doing A because...*, in a way that is immediate, i.e. not on a derivative basis. Even when this possession amounts to my knowledge that I am doing A because..., this knowledge, too, is not derivative. Now, this knowledge the agent enjoys is contemporaneous to their doing A. Further, insofar as the fact that they are doing A means that a process, their doing A, is taking place, there must be at least a bit more of their doing A in the future.³⁰ For any process, V-ing, is taking place at time t only if there is some more V-ing after t as well. The instant a drone gets stuck upon a sticky trap, the drone is no longer in the process of flying. But a fact someone encounters up to time t can never include something's taking place after time t. The closest approximation to a present process fact one can encounter by a given point in time is, *there has been V-ing up to this very moment (exclusive)*.³¹ So, if such an encountered fact, that p, is an actual basis or justification for my opinion at time t, *such and such a process is taking place now*, I must be taking a step, I must be inferring, from the fact that p. Encounters can inform me on only what I have done or, at best, what I have been doing up to now. They never inform me on what I am doing intentionally, insofar as this activity is a process, which always reaches into the future. Therefore, encounter with facts can at most enable, but never be a justificatory part of, the agential, non-derivative knowledge that they are doing A intentionally.

3.7.5 Two suspicions concerning *spontaneous possession of why-answers*

The picture that I spontaneously possess a why-answer to my doing A may suggest that I fabricate what counts as a reason for me to do A, I fabricate what explains my doing it, or

³⁰ I thank Michael Thompson for this thought.

³¹ The running argument does not appeal to the idea that our knowledge by encounter, because our sensual operation takes time, generally lags behind the fact known (with perhaps the exception of non-observational receptive knowledge, e.g. that I am being startled by a clap of thunder).

both. That is, the possibility of spontaneous explanation seems to presuppose, respectively, radical subjectivism about reasons, some power to make up etiology by fiat, or both.

Spontaneous possession of a why-answer for one's doing does not really presuppose radical subjectivism about what counts as a reason to do something. It may be true that, insofar as I spontaneously possess the etiological account that I am doing A because *c*, that *c* must be in some sense a reason to do A.³² Nonetheless, that it is a reason to do A and *how good or bad* a reason it is to do A need not depend *entirely* on things that happen to be true of myself as an individual. The extent to which something counts as a reason for someone in some situation to do something can at least partly depend, for instance, on intersubjective norms, a life form or a practice or an institution,³³ our second nature, a historically accumulated tradition about what count as reasons for what under what circumstances,³⁴ etc.

The other suspicion about the idea of spontaneous possession of a why-answer is that it seems to presuppose a spooky power to fabricate etiology. But this power appears spooky, I think, only if we assume that the *explanandum* – I am doing A – is something I merely encounter. *If* that were the case, then this explanandum fact would be settled independently of any conscious relation I may stand to it. If so, then to think that I could nonetheless have a spontaneous why-answer for it would be to think that I can make up a bit of etiology even though the fact is already settled in some other way. However, recall that, on the current view, the explanandum that I am doing A is not something I merely encounter. I must stand in a spontaneous intellectual relation to it.

³²And the extent to which this is true has been controverted e.g. by [Setiya \(2009b\)](#).

³³ See e.g. [Thompson 2004](#), [Thompson 2008](#).

³⁴ See e.g. [McDowell 1994](#), 126.

4.0 AN IMAGE OF FREEDOM

In this chapter, I develop a conception of free action through diagnosing what is unsatisfying about some existing attempts to defend free will optimism in the face of a causally deterministic or of an indeterministic world. Many recent moves made to resist the corresponding pessimisms in effect involve arguing that causal determinism (or indeterminism) is relevantly unlike any of the prima facie reasons for not treating someone as a fully appropriate target of engaged reactive attitudes in general for their conduct, i.e., for not treating them as fully responsible for. Some other non-pessimists, such as the Strawsonians, would agree, ‘well, the question of whether it makes sense to treat anyone as fully responsible does not itself make sense unless it is raised within the space of engaged reactive attitudes’. But to argue for the possibility of free will despite the truth of causal determinism (or indeterminism), one needs to identify a free-will-enabling condition in the agent that is not only compatible with the truth of causal determinism (or indeterminism) but also in tension systematically with all of the routine discounting considerations. And the Strawsonian line does not imply that it is impossible or not worthwhile to identify such a unifying condition. Until the unity is identified, it is unsurprising if the pessimists remain unconvinced.

My review of many existing conceptions of free will will show that they cannot explain how the routinely accepted reasons for not treating someone as fully responsible for their conduct function as such. I will propose an alternative construal. I will also argue that a subclass of those considerations, namely, prima facie reasons for not treating an agent as fully responsible for their conduct *either as an instance of a class of their voluntary actions or as occurring within a certain phase*, specifically suggests that the individual is lacking a sufficiently developed and unimpededly operating power of reason. From these conceptions about unfree actions, I arrive at a conception of free action: one is free in acting in a certain

way just insofar as what is knowable to them as the agent of the conduct is adequate – according to a certain notion of explanatory sufficiency – for explaining why they are acting in that way rather than acting not in that way. In this sense, one is free in relation to their action just insofar as a fact as to why they are engaged in it rather than acting without engaging in it is *constituted* by their representation of why they are engaged in it. Such a constitutive representation suffices to explain the explanandum just insofar as the behavior manifests the unimpeded operation of a relevantly developed power of reason. Free will optimism in the face of a causally deterministic world and in the face of an indeterministic world is therefore true, since whether an intentional action is causally determined or not deductively derivable from any set of predated events and states is orthogonal to whether the action unimpededly manifests the operation of a relevantly developed power of reason. However, because to have one’s power of reason fully developed for a given situation seems to be an ideal actual humans can only approximate, fully free actions turn out to be less common than they are allowed to occur on the accounts offered by many free will optimists. Furthermore, I will argue that, for an action to be an unimpeded manifestation of reason, it must not be incorrect. And an asymmetry follows: because the individual and their action are *unfree* in every way they are incorrect, blame and resentment can never be as justified as recognition and gratitude can be.

4.1 DISCOUNTING CONSIDERATIONS AND THE NEED TO DETERMINE THEIR UNITY

As P. F. [Strawson \(1974\)](#) points out, we share a canon of considerations each of which is routinely accepted as a prima facie reason for not taking an individual as a fully appropriate target for engaged reactive attitudes. Engaged reactive attitudes (ERAs) are attitudes to an agent for their conduct as if it exhibits their good will, lack thereof, or ill will toward another. ERAs include approval and indignation, gratitude and praise, resentment and guilt. And here are some examples of at-first-blush reasons for discounting the individual, *x*, as a fully appropriate target of ERAs, at least for anything they do in a certain period or for a class

of things they do voluntarily:

1. 'Someone secretly put drugs in what x drank.'
2. 'x is sleepwalking.'
3. 'x is acting under post-hypnotic suggestion.'
4. 'x is (literally) addicted to it.'
5. 'This is a symptom of x's obsessive compulsive disorder.'
6. 'x was brainwashed to believe that this is God's purpose for them.'
7. 'x has been manipulated into thinking and behaving this way.'
8. 'This is a symptom of x's Alzheimer's / brain damage.'
9. 'This is because x's childhood was deprived of love; they grew up not knowing any affection or care.'
10. 'x is a child.'

I will call such *prima facie* reasons discounting considerations. All discounting considerations belong to a broader class: *prima facie* reasons for mollifying or not having some ERA(s) about an individual's behavior.

A commonly formulated free will problem is whether free will is possible in a world for which (mere) determinism is true: every event is necessitated by events predating it according to some rule. At some point philosophers also began to worry whether free will is possible in a world for which (mere) indeterminism is true: no event is so necessitated by any events predating it. In the debate about the problem thus formulated, many recent moves in effect involve either of these two:

'(In)determinism is relevantly like some of the discounting considerations' (as one might argue for the incompatibility of (in)determinism with free will or with responsibility).

'(In)determinism is relevantly unlike any of the discounting considerations' (as one might argue for compatibility of (in)determinism with free will or with responsibility).

But the existing arguments seem not to satisfy someone of a different persuasion, in some cases because the optimist argument does not present a satisfying account – explicitly of free will or not – that reveals what unites all the discounting considerations, in other cases because the pessimist argument assimilating a perceived blocker of free will to a particular discounting consideration does not present a convincing account of how that discounting consideration functions to discount the agent from being a fully suitable target of ERAs for a class of their behavior.

4.1.1 Inability of conceptions of freedom in terms of ‘what one most wants’ to unite discounting considerations

A free will optimist, in the face of causal determinism or some other perceived blocker of free will, needs to identify a free-will-enabling feature of the agent that is not only sufficient for free will despite the perceived blocker, but also necessary for free will in such a way as to be in tension systematically with all of the discounting considerations.

Until the systematic unity is found, it is unsurprising if an optimist finds themselves in a passive position, being able to respond to pessimistic arguments only in a piecemeal, case-by-case fashion, e.g. every time an incompatibilist argues that determinism is just like this or that discounting consideration.

One class of compatibilist-friendly conception construes the relevant freedom as the agent’s being able to do what they most want – they are doing something freely if they engage in this doing precisely as what they most want. And the idea that acting γ -ly is one of the things in which x most wants to engage admits of more than one specific construal.

Since freedom of action, *as distinct from* freedom of the will, seems to be the freedom to do whatever one wants, one may model freedom of the will after that. Freedom of the will, in such a sense as distinct from freedom of action, is the freedom to want what one wants to want.

4.1.1.1 ‘What one most wants’ as what one’s deepest self wants One may, more specifically, construe ‘Doing A is what one wants to want’ so as to mean that doing A is one of the things in which x wants to engage come what may, no matter how thoroughly they reflect on their wants. Accordingly, one may find appealing the

Deeper-self condition: that the agent’s acting γ -ly is the expression of their volition to be acting γ -ly and this volition must be within the control of the agent’s self in some deeper sense than having this particular volition

Harry G. Frankfurt, for example, proposes a specific version of the above relation between the volition to act in a certain way and the deeper self:

[Frankfurt \(1971\)](#): the agent’s first-order, effective desire to do A (which Frankfurt calls this agent’s will to do A) conforms appropriately to their second-order desires for which effective desires they are to have

But this Deeper-self condition suffers a shortcoming. A controlling capacity to endorse or renounce a first-order self presupposes a higher-order point of view from which one endorses or renounces. As (Wolf 1990, 30-1) points out, if an agent's first-order volition can be alien to them, so can their second-order attitude toward this volition. And if the authenticity of the first-order volition requires control from a higher point of view, it seems that the higher-order point of view itself would need to be within the control of a yet-higher-order point of view *as much as* the first-order point of the view does.

Now, if (freedom of the will is human-psychologically possible and) it is not humanly possible for our action to be within the control of an evaluative standpoint of an infinitely high order, then the Deeper-self condition cannot be necessary for freedom of the human will. And even if such infinitely deep level of reflective control is somehow humanly possible, it seems that – for all that is mentioned in the deeper-self condition – *the basis* on which one takes any of the higher-order attitudes is not *thereby* different from the basis of their first-order attitude so decisively as to make the latter free or authentic.

Frankfurt's response is that a second-order desire of their effective first-order desire can be authentic, making the desired first-order desire authentic and free, even if the subject has not formed a higher-order desire reflectively endorsing the second-order one, for “[w]hen a person identifies himself *decisively* with one of his first-order desires, this commitment “resounds” throughout the potentially endless array of higher orders. . . The decisiveness of the commitment he has made means that he has decided that no further question about his second-order volition, at any higher order, remains to be asked.” (16).

But if this should render any higher-order desire unnecessary, it seems that it is the decisiveness rather than the reflectiveness that is doing all the work, so that a decisive first-order desire will also do. A first-order desire to do A may, in a happy case, already be based on all the considerations the agent would care to consider. Sheer reflectiveness, or the mere fact as to how many times the agent has stepped back to form a reflective stance, seems irrelevant to whether their first-order volition is free or whether it is really their own will. Perhaps it is necessary that the subject in some sense have the *ability* to form a higher-order desire about their will, but to actually form one seems unnecessary.

4.1.1.2 ‘What one most wants’ as what one’s most real self wants Sensing that a volition’s relation to a deeper point of view – a controlling or higher-order perspective – in the agent does not make any decisive difference, one may try a different way of construing freedom to consist in the conduct’s being fundamentally what the agent most wants. They may give a substantial characterization of what point of view or aspect of an agent counts as their real self, distinguishing this from the rest of their psychological makeup. On this alternative view, the agent’s freedom in their conduct is attributed to the conduct’s relation to a point of view not as deeper but as constituting who the agent really is. Wolf (1990), for the sake of argument, outlines a refined version of the condition:

Real-self condition: that the agent is “at liberty (or able) both to govern her behavior on the basis of her [volition] and to govern her [volition] on the basis of her valuational system” (33).

“What one most wants” in “doing what one most wants”, on this alternative, is construed in terms of the *quality* of the volition to do what one is doing rather than *how many times* one has exercised their power of willing, reflectively, over its own act. David Hume, Gary Watson, Martin Fischer and Mark Ravizza, and Alfred Mele maintain some of the most promising versions of the real-self condition in their conceptions of free agency:

Hume:

- “hypothetical liberty”: “if we choose to remain at rest, we may; if we choose to move, we also may” (*Enquiry* 8.23 / 95)
- the action proceeds from some cause in “the characters and disposition of the person” (*Treatise* 2.3.2.6 / 411)
- “liberty of spontaneity”: the agent performs the action not due to force, violence, or constraint (*Treatise* 2.3.2.1 / 407-8)¹

There may be no conclusive textual evidence for determining what exactly Hume means by “force, violence, or constraint” or how exactly he takes an action’s freedom from it – the “liberty of spontaneity” – to be related to the cause-in-character condition. It is probable, however, that by “force, violence, or constraint” he just means whatever hin-

¹ Whereas it is clear from the *Treatise* and the *Enquiry* that Hume thinks of the first two conditions, cause-in-character and hypothetical liberty, as necessary conditions for taking a doing as ground for ERA-like attitudes toward the agent, those works leave open whether he thinks this of liberty of spontaneity. As I go to suggest below, however, by his formulation of the liberty of spontaneity he may just be giving a negative characterization of an implication of what his “cause-in-character” formulation positive describes.

ders liberties or abilities similar to the ones mentioned in the real-self condition above.²

[Watson \(1975\)](#): the action is governable by one's system of values, i.e. system of desires for things as judged good by one's faculty of reason, i.e. as what one thinks there is some reason to want.

[Mele \(1995, 187, 193; 2006, 164\)](#): the individual is doing A with autonomy, in the sense that

1. they are “an ideally self-controlled and mentally healthy agent who regularly exercises [their] powers of self control” and
2. “a reliable deliberator”,
3. have “no compelled or coercively produced attitudes”,
4. their “beliefs are conducive to informed deliberation about all matters that concern [them], and they
5. “nondeviantly [do A] on the basis of a rationally formed deliberative judgment that it would be best to [do A].

[Fischer and Ravizza \(1998\)](#): the agent exercises “guidance control” of their behavior, in the sense that their present action issues from a mechanism that is

1. moderately reasons-responsive (that the agent exhibit “a pattern of reasons-recognition” that is at least moderately intelligible and react to at least one possible incentive to behave otherwise) (69-76) and
2. the agent's own (that the agent have acquired the belief, on the basis of relevant evidence, i.e., they are “an agent and sometimes appropriately subject to the reactive attitudes” for behaviors issuing from the said mechanism) (210-4).

Although the real-self condition may be necessary, it seems insufficient for freedom of the will, since it does not seem to be in tension systematically with all of the discounting considerations. We can have a prima facie reason not to treat someone as a fully appropriate target of any ERAs, in particular, with respect to an entire class of their voluntary behavior, even if the behavior is an unmistakable expression of their stable character, valuational system, authentic evaluative scheme, etc. In fact, we have such a discounting consideration if (1) the agent's voluntary behavior is bad, (2) it expresses their real self, and (3) the expressed aspect of this self can be attributed to their deprived childhood, in which they knew no affection or care and were not taught to have concern for others.³ Suppose Ilana had such a childhood. Although she has not become a sociopath or acquired any mental illness, through that childhood she did arrive, in a deep-seated and even reflective way, at the judgement

² I thank Karl Schafer for this reading. A similar interpretation is suggested in [Wolf 1990](#), 27-34.

³ It was Martha [Klein \(1990\)](#) who called my attention to such deprivation and its significance.

that it is of supreme value to outdo others. On one occasion, this value manifested itself in her deliberately misinforming her competitor Anali to outdo the latter in a competition. This behavior issues from a mechanism that is, in Fischer and Ravizza's sense, moderately reasons-responsive and Ilana's own. The featured childhood *prima facie* weakens the general appropriateness of ERAs toward Ilana for this lying as an instance of her voluntary self-promoting behavior. Thus, it is a discounting consideration. We have here, then, a case in which a discounting consideration is compatible with the most promising real-self conditions and, in fact, also with Frankfurt's deeper-self condition. This strongly suggests that the freedom of the will we care about is not exhausted by the freedom to do what one most wants to do – even in the most promising senses of “what one most wants to do”.

One may object that, intuitively, this upbringing does not discount Ilana as an appropriate target for ERAs for her self-promoting lying. Someone with no particular philosophical supposition may learn of this upbringing and still disdain Ilana or hold her responsible for this lying. Nonetheless, I reply, her deprived childhood is a basis on which it is *less* appropriate to e.g. disdain or resent her for any instance of resultant voluntary self-promoting behavior than if she had not been so deprived, *ceteris paribus*. If so, then the way in which the objection can be true does not contradict my claim that the real-self conditions and the deeper-self condition here coexist with a discounting consideration.

4.1.2 Unity of discounting considerations, preliminarily defined: the power of practical reason is not sufficiently developed and unimpededly operating

I will propose what I take to be the relevant factor indicated by a deprived and inevitably corrupting childhood and other discounting considerations, insofar as they excuse someone for their wrongdoing. I will also suggest what is indicated in common by discounting as well as non-discounting reducers of general ERA-aptness.

A Strawsonian may find my quest to determine the unity in all the discounting considerations futile. For it seems to presuppose the existence of an answer to the question, “What would rationally justify full-fledged ERAs in general despite determinism (or indeterminism)?” To this question, a Strawsonian would say, “[t]he existence of the general framework

of [ERAs] itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, a ‘rational’ justification.” For “questions of justification are internal to the structure or relate to modifications internal to [the general framework of ERAs]” (Strawson 1974, 23).

But even if we grant that the expectation of a rational justification for the very framework of ERAs is a confusion, there may still be a unity to be found in all the *discounting considerations*. The unity may turn out to allow rational justification for only a subclass of ERAs and not another. Such an asymmetry may call for a modification internal to the framework, which is exactly what I eventually argue: many behaviors that commonly invite a negative ERA actually reflect a condition in the agent that renders this very ERA less than fully appropriate to the agent for those very behaviors.

Two key features distinguish discounting considerations from other considerations in this broader class.

First, when a discounting consideration discredits a behavior as ground for ERAs to the agent, it discredits this behavior as ground for *any* possible ERAs. This separates discounting considerations from considerations that *prima facie* discredit a behavior specifically as ground for certain negative ERAs toward the agent without discrediting it as ground for some other ERAs. I have in mind considerations insofar as they suggest the behavior to be reasonable and closer to being justified. For example, if Mila has given away a watch she promised to give to Agnes, this may turn out to be reasonable and closer to being justified if and insofar as she used it to bail out her best friend Phat, or insofar as she was robbed with a gun. Either consideration is an at-first-blush reason for Agnes not to resent Mila, but neither consideration altogether weakens – not even *prima facie* – the appropriateness of every possible ERA toward Mila for this bit of conduct. For instance, it can be reasonable *of Phat* to be grateful toward Mila precisely if he learned that Mila used what she promised to give someone else to bail him out. These rationalizing considerations undermine the appropriateness of negative ERAs by showing the soundness of the power of reason manifested in the will being expressed in the act. Discounting considerations, by contrast, do not work by showing the soundness of the agent’s reason.

The second key feature of discounting considerations is that, when a discounting consideration discredits a behavior as ground for ERAs to the agent, it does so through discrediting it *as occurring in a certain phase of the agent, or as belonging to a certain class of their voluntary behavior*. For example, someone's addiction or OCD prima facie undermines the appropriateness of *any* ERA toward this agent *whenever* it comes to e.g. dealing with alcohol or dealing with the tidiness of their environment. This distinguishes discounting considerations from Aristotle's sufficient conditions for non-voluntariness: that someone is doing A out of ignorance or due to an external force. For example, if Tripp is trampling the geraniums I planted along the sidewalk, but either because he could not see his way due to a blinding sandstorm or because the storm caused him to stumble, maybe I should not be angry or hold any ERA toward him on the ground of this trampling. But Tripp does not thereby appear to be less than a fully appropriate target for ERAs for *everything* he is doing during the sandstorm or everything of a certain type he voluntarily does. Invocations of ignorance or external force undermine the appropriateness of ERAs by suggesting the behavior not to be an expression of will of someone may meanwhile have a perfectly developed power of reason that is manifesting itself in other ways without impediments..

Every discounting consideration, to repeat, suggests that the agent is less than fully appropriate target for any ERAs with respect to any of their behavior within a certain period (e.g. sleepwalking) or with respect to an entire class of their voluntary behavior (e.g. smoking, if they are addicted). This is a sign that there is a deeper unity behind discounting considerations. They all seem to suggest the *absence* of something necessary *in* x if anything x does is to be a potential ground for any ERA at all toward x. They suggest that a certain capacity is missing, underdeveloped, dormant, suppressed, or disabled. Conversely, if any one of the ERAs is to be fully appropriate toward an agent, the said capacity must be operational in them without being compromised in any relevant way. In particular, this capacity seems to be precisely that the soundness of which a rationalizing consideration defends and that by which an invocation of ignorance or external force suggests a behavior to be undirected. If so, then what a discounting consideration suggests to be absent is a sufficiently developed and unimpededly operating power to act:

Unity. A discounting consideration is a prima facie reason to think, with respect to someone's acting in a certain way, that this subject does not have an unimpededly operating power that is sufficiently developed for the kind of choice and situation involved.

I will argue that neither determinism nor indeterminism makes it impossible for one's behavior to be a manifestation of a sufficiently developed and unimpededly operating power of reason. To show clearly why it is so, why determinism is not relevantly similar to any discounting consideration, I will spell out *Unity* in terms more commensurable with empirically knowable nature and offer a positive account of free action in those terms. Before I get to that, however, I shall first discuss Susan [Wolf 1990](#), *Freedom within Reason*, whose approach to the free will problem is very similar to mine in vision and in spirit. More pointedly, its conception of the free will necessary for responsible agency is also similar to what I interpret the discounting considerations to concern and, by my lights, agrees with our actual practice of discounting considerations. Every time a discounting consideration applies to a given agent in their circumstances, her conception also implies that this agent is not fully responsible, and when it implies that a given agent *is* fully responsible, no discounting consideration applies.

However, [Wolf 1990](#)'s defense of the conception does not show – to the satisfaction especially of someone who is a pessimist in face of certain perceived blockers of action responsibility – how, when the conception implies that someone bears such a free will that they are responsible for something they are doing, some of the perceived blockers can also obtain. In other words, it is unclear why some of those conditions are compatible with free will as conceived by Wolf. In particular, if the agent's behavior is determined with necessity by some factor other than themselves, something other than the agent is an ultimate cause of their behavior, and the doing has some causal lineage that may be opaque to the agent, then it can only be an illusion if the agent appears to be doing the right thing for the right reason. Appreciation of the difficulty involved will motivate my understanding of the discounting considerations and non-discounting indications of non-voluntariness, as well as my conception of free action.

4.1.3 Susan Wolf's Reason View

Wolf says, “[A] responsible agent is one who has the ability to act in accordance with Reason” (1990, 68). To have this ability, according to her, one must have “the power to exercise right Reason”, which is also the power to recognize the True (and the Good, if there are evaluative facts):

According to the Reason View, having responsible status depends rather on a distinctive intellectual power, the power to exercise right Reason and to govern one's actions accordingly. Since, again, right Reason refers to those faculties which will, in most circumstances, lead one to form true beliefs and good values, the power to exercise right Reason may be redescribed as the power to recognize the True and the Good. The ability to act in accordance with Reason might then be redescribed as the ability to act in accordance with, and on the basis of, the True and the Good. (71)

And by “to have the ability to act in accordance with Reason”, Wolf means not only to have the power described above for recognition and governance, but also the absence of hindrances. Thus,

(Wolf's Reason View) A responsible agent is one who (1) has the capacity (1.a) to see and appreciate what reasons there are for doing the right thing and (1.b) to do it for these reasons, insofar as (2) nothing prevents or interferes with the exercise of this capacity. (1990, 87-8)

To be clear and fair, the Reason View was not presented as an account for determining when a doing is a free action, but when an individual is a responsible agent. The book, [Wolf 1990](#), is primarily concerned to answer “How, if at all, is responsibility possible?” – more specifically, to determine *how an individual must be constituted*, in relation to their will, for them ever to be responsible for the results of their will (3-4). The Reason View is Wolf's answer primarily to *this* question.

However, the Reason View does contain a conception of free will. For Wolf takes the above problem about responsibility also to be a “philosophical problem of free will”, and she understands free will to be “that relation to one's will which is necessary for one's actions (as well as one's character and life insofar as they are governable by one's will) to be “up to oneself” in the way that is necessary for responsibility” (3-4). So, insofar as the Reason View affirms something to be such a responsibility-according relation that the agent bears to

their wills, this is Wolf's conception of free will. In effect, the View identifies the responsible-status-according free will with the unimpeded capacity to act in accordance with Reason, or unimpeded receptiveness to the True (and the Good, the Good turns out to be part of the True). A salient implication of this view, which Wolf explicitly draws, is the following asymmetry. One can be responsible and praiseworthy for doing the right thing for the right reason even if they are deterministically caused to do so, but to be responsible for doing something wrong one must be psychologically undetermined to do that. This particular implication is correct in its 'predictions' – as we will see as an upshot of the conception of free action I defend.

4.1.4 Problem regarding wrong doings

There are, however, two problems. One of them is the hitherto unsatisfied need to show that the conditions (1) and (2) for responsibility in the Reason View are in fact compatible with widely-perceived blockers of responsibility. This problem will be taken up in §4.2.1.

The problem to be discussed here is that conditions (1) and (2) in the Reason View are not, *contra* Wolf, jointly sufficient for attribution of responsibility to an agent for acting wrongly. For if they were sufficient, they would attribute responsibility to the agent also for things for which they clearly need not be responsible. Let's start by considering the more general question: How does the Reason View thus far formulated tell us whether an agent is responsible for any *particular, given thing* that they are doing? Surely, Wolf could not mean that an agent who meets conditions (1) and (2) is thereby responsible for *everything* they can be said to be doing. For even someone who is doing the right thing on the basis of the right reasons can at the same time be doing something non-voluntarily, e.g. performing a reputable piece of music during encore that causes some stranger in the audience to faint for whom this piece is peculiarly associated with a previous trauma. To be charitable, we shall start with the happy case, in which the agent is acting in the right way for the right reason, and supplement conditions (1) and (2) to formulate the following sufficient condition:

The Reason View regarding a right doing: Given that A is the right thing for an agent to be doing and they are doing A, this agent is responsible in doing it if (1) they have the capacity (1.a) to see and appreciate what reasons there are for doing the right thing and

(1.b) to do it for these reasons and (2) nothing prevents or interferes with the exercise of this capacity, and (3) their present doing A is a manifestation of the capacity described under (1).

But what if A is *contrary* to what is the right thing for the agent to do? Wolf goes to lengths to argue that people can have the abilities described in criterion (1) and in criterion (2) without exercising them and, in this sense, could be acting otherwise than how they are acting in reality (88-9). So, according to Wolf, one can be responsible for not recognizing the right reason or not doing the right thing in accordance with it. . To defend this possibility, Wolf specifically argues that, even if determinism is true, one can have unexercised abilities to do, and so could be doing, what one is not in fact doing (101-16). On her view, then, the way one is responsible for doing something wrong does not require that this doing be the manifestation of their ability to act in accordance with Reason:

The Reason View regarding a wrong doing: Given that A is contrary to the right thing for an agent to be doing and they are doing A, this agent is responsible in doing it if (1) they have the capacity (1.a) to see and appreciate what reasons there are for doing the right thing and (1.b) to do it for these reasons and (2) nothing prevents or interferes with the exercise of this capacity.

Such a conception of responsibility prompts the following question. If someone can be responsible for doing something wrong even though it does not issue from the ability that makes them a responsible being, why are they not responsible for everything they are doing, including ones that are neither the right thing to do nor a wrong thing to do? On Wolf's behalf, one might reply that one is responsible for doing something only if it is either due to their exercising *or due to their omission to exercise* their ability to do the right thing for the right reason. This excludes doings that are due to neither (e.g. twitches, starts, etc.), while every wrong doing is presumably due to the agent's omission to use that ability (or the absence of the ability).

Still, one who omits to exercise the ability they have to do the right thing for the right reason may thereby bring about consequences that they do not anticipate and cannot be reasonably expected to anticipate. Thus, this class of doing – doing something due to one's omission to exercise the ability to act in accordance with Reason – is more extensive than wrong doings, and the additional necessary condition adduced in the above projected

response on behalf of Wolf above offers nothing to resist the claim that the agent is a suitable candidate for praise or blame for each of their doings in this class. For example, Ilana, who is wrong in intentionally misinforming a competitor, Anali, about the time of the competition only in order to outdo her, presumably is not to be credited for e.g. saving Anali's life even if Anali would otherwise have shown up at the venue and been killed during a mass shooting that surprised Ilana herself.

Now, one might say on Wolf's behalf that an agent can be responsible for doing something wrong without being responsible for all the consequences they bring about in omitting to exercise the ability to act in accordance with Reason, for the wrong doing is precisely something they are able to grasp *the reasons why not* to engage in and something they are able to refrain from for these very reasons. (The agent may even actually grasp the reasons why not to engage in their actual behavior, one might say.) But it is not obvious that this is possible. For

Comprehensive. If someone has a power to act rightly for the right reason, then whenever they are doing anything intentionally at all, this power, *unless* something impedes its activity, manifests itself in all their concurrent conduct, so that the latter is in every aspect compatible with acting rightly for the right reason.

So, if such an individual is doing something wrong while doing this or something else intentionally, either they do not even have the power to do the right thing for the right reason in this situation (hence failing to meet even condition (1) in the Reason View), or they have it but, in this case, something must have prevented or interfered with the exercise of their rational power in some respect (hence condition (2) fails to obtain). In either case, something is missing that is necessary for action responsibility (i.e. full appropriateness of ERAs in general toward the doer for a specific doing) – by Wolf's own lights, at any rate.

Here, it is of interest to consider R. Jay Wallace's view. A condition very similar to condition (1) in the Reason View is proposed in [Wallace 1994](#) as necessary and sufficient for an agent to be an accountable agent in general (1, 84), for it to be "fair to demand that [the agent] comply with moral obligations we accept". The condition proposed there is that the agent possesses what Wallace calls "the powers of reflective self-control": "the power to grasp and apply moral reasons" and "the power to control or regulate [one's] behavior by the light of such reasons" (157-8). Here, moral reasons are understood to be reasons that

explain why the moral obligations to which the agent is fairly held are moral obligations. It is of interest that something I take Wolf to imply, namely, that one can be responsible or accountable for a behavior *even if* they do not therein exercise the power supposed to suffice for the status of a responsible or accountable agent, Wallace explicitly endorses: “What matters is not the exercise of the general powers of reflective self-control, but the possession of such powers” (190). His view, however, does include a necessary condition in addition to the possession of powers for the agent to be responsible *for a specific wrong doing*. He argues that a morally wrong doing typically is one that expresses a particular choice on the agent’s part, and so typically is an intentional action (132-134).⁴

Let us consider, then, the case where the behavior in question is an intentional action, in addition to the agent’s possessing the power to act rightly for the right reason, for both Wolf’s idea of the ability to act in accordance with Reason and Wallace’s idea of the powers of reflective self-control are conceptions of the power to act rightly for the right reason, in some sense of rightness and reasons. Recall *Comprehensive*: if someone has a power to act rightly for the right reason, then whenever they are doing anything intentionally at all, this power, unless something impedes its activity, manifests itself in all their concurrent conduct, so that the latter is in every aspect compatible with acting rightly for the right reason. If this is true, then something must have impeded the activity of the agent’s power to act rightly for the right reason, given that their wrong doing is an intentional action. And this impediment, by definition, must be alien to the power whose activity it impedes. It is alien, then, to the very sort of power that Wolf and Wallace take – and rightly take – to be the source of responsibility in the first place. Therefore, the impediment is at least a *prima facie* reason not to hold the agent responsible for the wrong doing, not to treat them as an appropriate target of ERAs for it. It is unclear why, despite this *prima facie* reason, it would be fair to hold the agent accountable.

In conclusion, the asymmetry between the happy cases and the cases in which the agent fails to act in the right way for the right reason is far deeper than Wolf thinks. Recall that her Reason View purports to determine the relation the individual must bear to their will

⁴ Presumably, on Wallace’s view, if it is fair to hold the agent responsible for a specific wrong doing, the agent must have the power to grasp, apply, and regulate their behavior by the light of the specific moral reasons that explain why the wrong doing is wrong.

for this individual to be a responsible being, and this required relation is for Wolf free will. So, since she argues that it is possible for an individual to be responsible for doing something wrong, this commits her to the view that one's will may be free even in acting in a wrong way. It is just that, according to Wolf, if one's will in so acting is to be free (and one is to be responsible for so acting), they must be psychologically undetermined to so act. On my view, this conditional is true, but it is trivially so, and it misleadingly understates the asymmetry – one's will in acting wrongly thereby *cannot* be (completely) free. We will see this in §§4.4.2, 4.4.3.

4.2 OBJECTIONS TO PREMISE W AND REPLIES

4.2.1 First objection to premise W (and subversion of Wolf's Reason View), from the Paradox

Notwithstanding Wolf's Reason View, one may worry that a foolproof correct education would make it impossible for the agent to count as acting for the correct reason(s) so as to deserve credit or a positive ERA. Wolf has not shown – to the satisfaction especially of someone who is a free will pessimist of a certain sort – how, when the conception implies that someone bears such a free will that they are responsible for something they are doing, some of the perceived blockers of free will can also obtain. In other words, it is unclear why some of those conditions are compatible with free will as conceived by Wolf. In particular, if (1) the agent's behavior is determined with necessity by some factor other than themselves, (2) something other than the agent is an ultimate cause of their behavior, and (3) the doing has some causal lineage that may be opaque to the agent, then it can only be an illusion if the agent appears to be doing the right thing for the right reason. The defender of a supposed rationale for pessimism can grant me and Wolf, for the sake of argument, that the perceived blockers of action responsibility do not deny the agent credit *if* their behavior in question amounts to doing the right thing in appreciation of the right reason; it's just that the education or upbringing itself precludes the agent's behavior from amounting to

an action genuinely out of their appreciation of the right reason. The truth of this last, paradoxical claim would mean that the Reason View turns out to be a reason for the variety of pessimism that corresponds to the blocker in question.

Also, with respect to the second premise in my W Argument, the paradoxical claim can ground an objection as follows:

1. (CR) One is a fully suitable target of full-fledged ERAs for acting correctly only if it is genuinely out of their appreciation of correct reasons.
 2. (Paradox) A foolproof correct upbringing itself precludes the agent's behavior from amounting to an action genuinely out of their appreciation of correct reasons.
- ∴ (FS) A foolproof correct upbringing is a reason to discount the agent as a fully suitable target of ERAs for their conduct insofar as it is a result of this upbringing.

Given that an education of practical wisdom is a correct upbringing, FS contradicts premise W in the W argument, namely, that a foolproof education of practical wisdom would not be a reason to discount the educated as a fully suitable target of ERAs for their resultant wise conduct. While CR seems plausible to me, I find the Paradox, well, implausibly paradoxical.

But one may argue for the Paradox by the following train of thought. If how one will act is however one has been alienly caused, e.g. taught, to act, then one will so act whether or not it is correct. Then it would just be lucky if some agent has been taught, e.g. by a teacher who had the correct opinion, so that their doing A now coincides with what it is right for them to be doing; the agent would deserve no credit for this coincidence as their achievement. Likewise, it would just be bad luck if what they are doing coincides with what it is wrong for them to be doing; they would deserve no blame for this as their fault. It seems to follow that a subject cannot be doing something both because they have been taught or alienly caused in any other way to so act and, at once, for the correct reasons. To be able to act for the correct reasons, then, it seems that the agent must not be necessitated by anything other than themselves to take action. If being necessitated by something other than oneself implies that the agent does the right things they do only by chance, it seems hard to explain how one is really doing them out of appreciation of the right reason, let alone being responsible,

while one cannot be responsible if the conduct is ill informed with psychological necessity. This asymmetry seems hard to defend when both conducts are products of necessitating factors not controlled by the agent and neither agent could act otherwise. I will call this the Argument for the Paradox.

Wolf's response seems inadequate. She denies that if how one will act is however one has been taught to act, then one will so act whether or not it is correct. Using just action as an example, her objection is that "[if it] weren't right to act justly, the person might well not have been taught that it was" (1990, 73). But the person might well have been taught that it was! Suppose it was not right to act justly, at least in a given situation, but the mentor of our agent had such a hang-up about injustice as to believe that what is just is always right. Thus, this mentor could very well have taught our agent to act justly always, even if it were not right to so act in the present situation.

4.2.2 Disarming the Argument for the Paradox

In effect, the argument argues that certain features of a foolproof correct education preclude the educated person's acting γ -ly from being their action for right reasons. Call these features seeming blockers of action for right reasons.

4.2.2.1 Heteronomy One feature alleged to be such a blocker is what I call heteronomy, that the agent's behavior is determined with necessity by some factor other than themselves. The apparent support it gets from the Argument for the Paradox depends on some confusions. §2.4.2 has prepared us to point out one, so that we may focus on what remains. There, I suggested that if the necessitation of something is ever in tension with someone's responsibility for that thing, it would only be when the necessitation is an aspect of a certain causation, and the tension would only be due to the causation, but necessitation is conceptually distinguishable from causation. Apart from causation, the fact e.g. that x necessitates y to act γ -ly means only that y 's γ behavior is deductively derivable from x' and some rule. In other words, x may necessitate y to act γ -ly without causing y to act γ -ly, even if x is other than y and temporally precedes y . So, even if a subject cannot be acting γ -ly both because

they have been alienly caused to so act and out of genuine appreciation of the correct reason, it does not follow that someone's acting γ -ly cannot be out of their genuine appreciation of the correct reason, if this behavior is necessitated by something other than themselves.

What remains of the Argument for the Paradox still appears to support the paradox in other ways. For one, this still seems to follow:

If the fact that someone is acting γ -ly has a causal profile that is somehow alien to the agent, then it cannot be out of appreciation of the right reason that they are acting γ -ly.

In particular, for instance, one may find it hard to resist this:

If someone's acting γ -ly has an ultimate cause that is other than themselves, then it cannot be out of appreciation of the right reason that they are acting γ -ly,

What remains of the Argument for the Paradox to support this is as follows. Suppose one has been so alienly caused to change, e.g. by education, that they, characteristically of themselves, act γ -ly in some given situation, and they would not be acting γ -ly now had they not been taught the way they were in fact taught. When their education began, they did not prefer one source to another; given all their system of attitudes and capacity to reflectively evaluate the attitudes, it was chance that they received the education they received. Therefore, it seems also by chance that the way they act coincides with the right way for them to be acting. This agent would deserve no credit for this coincidence as their achievement. Likewise, it would just be bad luck if what they are doing coincides with what it is wrong for them to be doing; they would deserve no blame for this as their fault. It seems to follow that a subject cannot be doing something both because they have been taught or alienly caused in any other way to so act and, at once, for the correct reasons.

4.2.2.2 Source other than the agent I shall begin undermining the standing Argument for the Paradox by distinguishing between causality with a cause other than myself and alien causality. To confound these two ideas is also a second confusion that lends appeal to heteronomy as a blocker of action for right reasons. Having an ultimate cause other than the agent seems to exclude the behavior from being an action for right reasons because, if my way of conduct is alienly caused, this renders accidental any coincidence between my way of conduct and that it is right (or wrong) of me to so act. This seems to entail that, if the way

I am doing what is right (or wrong) makes me an appropriate target of ERAs, this way of behaving is not caused by something other than myself. This apparent entailment suggests that to be free and responsible in acting ϕ -ly the agent must not be caused by anything other than themselves to be acting ϕ -ly. But even if a (necessitating or non-necessitating) cause of my doing A is something other than myself, it does not follow that the causal connection between the cause and my doing A is alien to me. That is to say, one alternative to alien causality is non-alien causality by something other than me. Thus, even given that a subject cannot be acting in a certain way both because they have been alienly caused to so act and for correct reasons, it does not follow that, to be able to act for correct reasons, the agent's conduct must have no cause other than themselves.

If causality alien to me is not conceptually the same as causality by something other than me – so that something other than me can cause me to act non-alienly –, then what *is* it? For a bit of causality to be alien to me is for a causal fact, a *why*, to be knowable to me only by my finding it out, i.e. discovering it. Finding out or discovering a fact, e.g. finding out why an individual is behaving in a certain way, is one way of knowing a fact – or one major way of standing in a non-accidental intellectual agreement with a fact (which is more generic than knowing a fact if knowledge is the ideal but not the only possible case of such agreement). Discovery in turn can be divided by how the discoverable fact is discovered: as what one encounters or as what is affirmed by the conclusion in one's theoretical inference. Knowledge by encounter includes knowledge by observation, introspection, and testimony, to name some garden varieties. Such ways of knowing a fact are what I signify by 'finding out'.

Just given that my conduct has a cause that is not me or not merely me, e.g. that it would be unfair to so-and-so if I should misinform them, or that I owe so-and-so \$50, it does not follow that the causal connection between my just behavior and the cause is something knowable to me only by discovery. It does not follow that what is indeed a cause other than me alienly causes me to act justly. That a behavior has an ultimate cause other than the agent is not genuinely shown by the foregoing considerations to preclude an action from being for right reasons.

4.2.2.3 Alien causal lineage I have not yet disputed that, if (a) one has been alienly caused, e.g. taught, to act γ -ly, (b) they might be acting not γ -ly had they been taught somewhat differently from how they were in fact taught, and (c) it was chance that they received the education they received, then it seems also accidental if the way they act coincides with the right way for them to be acting. This conditional suggests that a given course of action cannot be both for correct reasons and caused in a way alien to the agent:

If someone's acting γ -ly has an alien causal lineage, i.e. is connected to something as its cause in a way that is not knowable to the agent otherwise than as something they discover, then it cannot be out of appreciation of the right reason that they are acting γ -ly.

Such processes as psychological conditioning, habituation, upbringing, and education do change how people behave, and they bring about changes in a way that is knowable to each subject as something they find out. In other words, these processes involve causality alien to the subject in whom changes take place. Therefore, to the extent that our behaviors are products of these processes and 'I was trained/taught/conditioned/habituated to act justly' is usually not a correct reason for acting justly, these facts are in tension with acting genuinely for correct reasons – so long as an alien causal lineage seems to be a blocker.

To better appreciate the apparent threat of an alien causal lineage, let us consider a more fleshed-out example. Suppose Anali has been taught by Tung to act justly and today she is acting honestly in a competition, thanks to this education. She would not be acting honestly had she not been taught the way she was in fact taught. Many of her peers in fact received a different education and as a result are acting dishonestly in the same competition. When her education began, she did not prefer Tung or Tung's teaching to any other; given all her system of attitudes and capacity to reflectively evaluate the attitudes, it was chance that she received the education she received. More importantly, she subsequently *learned* how this education had changed her and that she might be of a different character had the education been different in a certain way. Therefore, it seems also by chance that she ended up having a just character and that the way she acts is the right way for her to be acting. Even granted that it would otherwise be unfair of her – and she believes this as her reason to be honest –, this does not seem to be the real cause of her behavior. For, had her education been different, she might believe something else as her reason to behave honestly or she might

believe something as her reason to behave dishonestly. In sum, given that Anali is acting honestly because she was taught to be just, in the sense that (a) she has been alienly caused to be just and (b) she might be acting unjustly had she not been taught the way she was in fact taught – and since (c) she was not free in or responsible for receiving the education she received rather than a different one –, it seems that the just thing she is doing, that is, being honest, she is doing not due to the fact that it would otherwise be unfair or for whatever correct reason there is.

Even given that Anali became just through an alien causal process and by chance she ended up having a just character, and even given that her belief that it is unfair to be dishonest is part of this character, the fact that it is unfair to be dishonest can nonetheless be a real cause of her honest conduct. It may be intrinsic to precisely a just character for the agent to *know* that it is unfair to be dishonest and know it as why one should be honest. And Anali's honest behavior can be, in turn, intrinsic to her knowledge that it would otherwise be unfair, reflecting what she knows to be the reason for behaving honestly. If this is the case, then the coincidence between her behavior and what is the right way for her to behave is non-accidental and the fact that it would be unfair to be dishonest is a real cause for her honest behavior. Then Anali is behaving honestly both because she was taught to be just and because it would otherwise be unfair.

4.2.3 Second objection to premise W, from the explanatory ultimacy of up-bringing

Even having disposed of the Paradox, one may still feel threatened by the applicability of the explanations we *learn* from experimental psychology, neuroscience, or genetics, of our own behavior, or alien causality by way of teaching, social conditioning, habituation, etc. The worry is that some causality of this sort will always constitute an explanation of an action *ulterior* to any correct reasoning on the basis of which the agent engages in that action. Even if Anali is behaving honestly both because she was taught to be just and because it would otherwise be unfair, the worry goes, the first 'because' is of a sort that is more basic. After all, it is thanks to Tung's teaching that Anali now knows that it would be unfair to

be dishonest and, therefore, is able to act for this as her reason, as she does now. This would be in tension with Anali's being free or responsible in her honest behavior under the additional assumption of the Ultimacy View, according to which an agent is not responsible, and therefore has no free will, in their conduct, if its ultimate causal lineage knowable to the agent themselves just in virtue of their being the agent of the conduct has still an ulterior causal lineage, which is merely discoverable to them.

This View seems compelling because it is not inherent to causality of the alien sort that the agent be in touch with it; opacity is perfectly possible: Anali might very well be oblivious of why she came to be just or be mistaken about it (if, say, Tung was subtle in his justice pedagogy). So, if a causality of this sort always underlies whatever the agent thinks is the explanation for their action, the agent seems to lack the authority, authorship, or control required for them to be acting freely.

This also amounts to a second objection to premise W in the W Argument, namely, that a foolproof education of practical wisdom would not be a reason to discount the educated person as a fully suitable target of ERAs for their resultant wise conduct.

1. (UU) Insofar as someone's conduct is a result of their upbringing, this causal connection is (1) merely discoverable to themselves and (2) ulterior to the ultimate causal lineage of the conduct that this agent can know just by being the agent.
 2. *Ultimacy View*. An agent is not responsible, and therefore has no free will, in their conduct, if it has a causal lineage that is ulterior to its ultimate causal lineage knowable to the agent just by being the agent of this conduct.
- ∴ (FS) A foolproof correct upbringing would be a reason to discount the recipient as a fully suitable target of ERAs for their conduct insofar as it is a result of this upbringing.

While UU seems plausible to me, I object to the Ultimacy View as follows. No etiological connection or explanation, I reply, is ulterior to or more basic than another unless they both address the same sort of explanatory question. That is to say, neither causalities nor explanations are ordered as more and less ultimate independently of the erotetic aspect in which something is explained. Further, just because e.g. a behavior's attributability to the agent's upbringing is ulterior to the causal connection they can agentially know with respect

to one sort of explanatory question, it does not follow that it is also ulterior with respect to the sort of explanatory question that is relevant for the agent's freedom or spontaneity in the behavior. In particular, the sort of why-question with respect to which the agent freely relates to their action is peculiar in that a proper answer to it should be revelatory of what the action is, more exactly. Although alien causality, such as the connection between one's conduct and one's former upbringing, is no doubt also the quarry of some why-questions about the agent's behavior, it cannot contribute toward revealing what that conduct is exactly. We can see the same distinction in aspect even in a case of theoretical thought. It is true that, unlike the case of practical thought, why the theoretically thinking subject knows a fact is clearly not *thereby* part of the explanation of this fact. For example, an astronomer, Tara, knows e.g. that a comet when traveling near the Sun forms a tail pointing away from it. Tara's training, certain trained observations, and certain informed judgements she made can contribute to explaining why she knows that, but they do not contribute toward explaining the phenomenon. Something that would contribute to it is e.g. the history of the Milky Way that explains why there are objects of certain compositions revolving around the Sun on certain orbits. Despite this difference from practical thought, there nonetheless is a distinction between something that contributes to explaining an explanandum without contributing toward explaining *what it is* and something that also contributes to that. There are bits of Milky Way history, possibly ones that far predate any comet, that contribute to explaining why some comets form tails without contributing to explaining what a comet tail is. In comparison, the solar radiation that vaporizes volatile solids on the comet and the radiation pressure that pushes the vapor away from the Sun contribute not only to explaining why some comets form tails but also to explaining what a comet tail is. Whatever ultimacy certain Galaxical-historical or even Universal-historical facts enjoy over the more comet-tail-specific facts is irrelevant to the latter's sufficiency for explaining both what a comet tail is and why a comet orbiting near the Sun as such forms a tail.

Of course, practical thought – of e.g. behaving honestly – is, again, different from theoretical thought in that its thinking subject is non-accidentally both one to whom the behavior is knowable in its factivity and one who issues the behavior. (By contrast, for a theoretical thought, its thinking subject as such cannot be the one who issues, in factivity,

what they represent except coincidentally.) Therefore, a pedigree that explains how someone came to have the power for practical thought also contributes toward explaining in a way the occurrence of what is being thought, i.e., explaining in a way why the subject is behaving honestly now. But just as there can be a distinction, in erotetic aspect, between what contributes toward the comet tail's occurrence in such a way as to explain also its essence and what contributes toward the occurrence without explaining the essence, so there can be a distinction between a cause contributing toward the occurrence of someone's honest behavior in such a way as to explain also the behavior's essence and a cause merely contributing toward its occurrence. And just as Tara's causal understanding of the comet tail can be adequate with respect to what it is without knowing all of the Solar System's history leading to comet tail formations, so can Anali's causal understanding of her own honest behavior be adequate with respect to what it is without knowing all of the causal history leading up to her honest behavior. For the explanandum that Anali is behaving honestly, whatever ultimacy the fact that she was taught by Tung in a certain way enjoys over the fact that it would be unfair to be dishonest, is irrelevant to the latter's adequacy for explaining what more exactly her behavior consists in via explaining why she is behaving honestly. And it is in relation to this *what* and the *why constitutive of this what*, I think, that the spontaneity of the agent's will matters to us. Therefore, I reject the Ultimacy View.

In effect, I have suggested that one can be alienly caused, e.g. taught, to become a being to be which is for a different causality to take place, a being that is home to a different kind of 'because'. But, of course, it is also possible to be *merely* alienly caused to behave in a certain way and have merely accompanying, at-best-accidentally-true beliefs about why one is behaving that way. What, then, is the difference between someone knowing that *p* as the correct reason for which they are acting and someone being merely alienly caused to believe so? If someone knows that *p* as the correct reason for which they are acting γ -ly, the etiological connection between the fact that *p* as the explanans and that they are acting γ -ly as the explanandum must be known to themselves not as something they merely discovered. The existence of an alternative kind of causality to alien causality and their difference are key to understanding discounting considerations as a whole as well as to a satisfying account of free will. To these two topics we now return.

4.3 FROM THE *UNITY* OF DISCOUNTING CONSIDERATIONS TO *FREE ACTION*

4.3.1 Unity of discounting considerations, defined more commensurably with empirically knowable nature

In fact, both the invocation of ignorance or external force and the invocation of a discounting consideration suggest that any explanation for the behavior in connection with which they are invoked is not adequate unless it includes a type of explanation that the agent cannot know otherwise than by finding out. But they suggest this via different grounds. Whereas the invocation of ignorance or external force specifically suggests that the behavior in question does not express the will of someone who may well be a perfectly developed and functional agent, a discounting consideration specifically suggests that the agent has not got everything it takes to act with a good will.

Unity, in higher definition. A discounting consideration is, with respect to someone's acting in a certain way, a prima facie reason to think that, *because* the practical power manifested in this behavior is not sufficiently developed and unimpededly operating, an adequate explanation⁵ for their acting in the way in question rather than acting not in this way must include a type of explanation this agent cannot know otherwise than by finding out.

4.3.2 Free action in terms of the alternative to discovery

In the above reflection on discounting considerations together with attribution to ignorance or external force, we saw that a behavior is unfree when there is a need for an explanation that the agent cannot know otherwise than by finding out, if it is to be explained why they are acting in the way in question rather than not in that way. By *modus tollens*, one is free

⁵ Perhaps something can be an explanans or explanation but not adequately so. This happens, I imagine, e.g. when something explicates why a certain course of behavior, *C*, rather than *a* particular alternative takes place but fails to explicate why the subject's course of behavior is *C* rather than *not C*. For example, laziness may explicate why someone is lying to outdo others in a competition rather than practicing the relevant skill more to outdo others, but it does not explicate why they are lying to outdo others rather than e.g. quitting the competition. Thus, laziness fails to explicate why this agent's course of behavior includes lying-to-outdo-others rather than not. In this sense, it may be said that laziness is a partial but inadequate explanans for why this agent is lying to outdo their competitors rather than acting not in this way.

in acting γ -ly only if an adequate explanation for their acting γ -ly rather than not γ -ly need not include any explanation this agent cannot know otherwise than by finding out.

The only alternative to discovery, available to human beings, as a way to know a fact – or to stand in any intellectual non-accidental agreement with it –, is to represent the fact constitutively. That is to represent the fact in such a way as to constitute what this fact is. Here is a summary table of the distinctions:

Ways for a subject to know a fact and, more generally, to stand in intellectual non-accidental agreement with the fact.

- **Theoretical:** relating to the fact as what one discovers, i.e. finds out
 - Receptive: relating to the fact as what one encounters
 - * by observation
 - * by introspection
 - * by testimony
 - * ...
 - relating to the fact insofar as it is affirmed by the conclusion in one’s theoretical inference
 - ...
- **Spontaneous:** relating to the fact in such a way as to constitute it (my representing of the fact is not other than what this fact is). I enjoy such a relation to the fact...
 - that I am (theoretically or practically) inferring
 - that I am intentionally doing A because p
 - that I am doing A insofar as it is affirmed by the conclusion in my practical inference
 - that I am doing A insofar as it is the conclusion in my practical inference
 - that I am intentionally doing A
 - ...

Spontaneous representation is key to the possibility of free action, for which the agent would be an appropriate target of engaged reactive attitudes.

So, we can reformulate the necessary condition for an action to be free, now positively in terms of the agent’s self-sufficient understanding of their behavior: one is free in acting γ -ly only if an explanation they *spontaneously possess* is adequate for their acting γ -ly rather than not γ -ly.

Having debunked the Aheteronomy, No-source-other-than-self, No-alien-causality, and Ultimacy Views as potential rationales for pessimism about the general possibility of free

will, I see no reason why the present necessary condition for freeness of action is not also a sufficient condition. Therefore, I submit the following thesis:

Free Action. One is free in acting γ -ly just insofar as a fact why they are acting γ -ly is represented by themselves in such a way as to constitute why they are acting γ -ly rather than not γ -ly.

In other words, x is doing something, A , freely just insofar as (a) an explanation for why they are doing A rather than acting without doing A applies and (b) this explanation applies because it is knowable to x in the very way they represent why they are doing A . I say ‘knowable’ rather than ‘known’ to accommodate the caveat that my representing why I am doing A does not always amount to my knowledge why I am doing A . My representing myself as doing A because p does not amount to knowledge when I do not know that I am doing A or when I do not know that p . On the one hand, it is possible to know why oneself is trying to do A and to be doing it without knowing that one is doing it, as we observed in §3.2. Even when one is doing A intentionally, the context may not be transparently familiar to themselves in every way relevant to exercising their capacity to do A . On the other hand, it is possible that it is the case that p , one also thinks that p , but one does not know that p – e.g., Gettier cases and cases in which the subject is unaware of a compelling reason for them to be skeptical of their own discerning ability. Because of this possibility, it is possible to constitutively – and, thus, knowably – represent oneself as doing- A -because- p without knowing that p .⁶ Due to these two possibilities, and since knowing why requires knowing the that on both sides of the ‘because’, the agent’s spontaneous representation of their doing A and of why they are doing A does not always amount to their knowing that why. But it is not important to the freeness of my action that why I am doing it be known to me rather than merely spontaneously represented by me; it is only important that it be *knowable* to me in the spontaneous way. To put the thesis in yet other words, one is free in acting γ -ly just insofar as one of the causal lineages of acting γ -ly that is spontaneously knowable to this agent is sufficient to explain why they are acting γ -ly rather than not γ -ly.

It follows that neither a deterministic nor an indeterministic world precludes the possibility of free action. For what truly detracts from a doing’s status as free action is a particular

⁶ I thank Haixin Dang for suggesting this bearing of Gettier cases on what an intentional doer thereby knows.

explanatory need – occasioned by the doing – that cannot be satisfied by any etiological connection knowable to the doer spontaneously. This need may be filled by some merely discoverable, i.e. alien causal lineage, or not filled at all (in this latter case the doing is in this aspect accidental). But whether there is such an explanatory need is orthogonal to whether the occurrence of the action is determined or undetermined. And since such an explanatory need can be absent even when an explanans or cause is a merely discoverable fact to the agent, the No-source-other-than-self View is also false according to *Free Action*.

4.4 WHAT IT TAKES AND DOESN'T TAKE TO BE FREE IN ACTION

With what we have got so far, here are the important implications of the falsehood of the Ultimacy View. Its falsehood itself means that, for an agent who is doing-A-because- p , the fact that p can be a cause for their doing-A-because- p in whatever way is required for the agent to rate as free in this behavior, even if this doing-A-because- p is caused in a way alien to the agent. When the negation of the Ultimacy View is combined with the truth of *Free Action*, it follows that one is unfree in acting γ -ly not because some merely discoverable explanation applies to their acting-in-this-way-rather-than-not-in-this-way, but because no etiological connection spontaneously knowable to the agent suffices to explain it. But doing-A-because- p requires that the agent constitutively represent themselves as doing A because p . So, the falsehood of the Ultimacy View also implies that the fact that someone's behavior is alienly caused does not in general preclude why the subject is doing A from being constituted by this subject's representation of this very why – in whatever way is required for this subject to rate as free in this behavior. Further, given *Unity*, according to which the adequacy of the subject's spontaneously possessed explanation for their behavior depends on their having a sufficiently sound and unimpededly operating power of reason, the falsehood of the Ultimacy View also means that a behavior's being alienly caused does not in general preclude the agent from having a sufficiently sound and unimpededly operating power of reason in that behavior.

These implications call for a deeper investigation into cases like Ilana's. According to *Unity*, Ilana's upbringing is a prima facie reason for not taking her to be a fully appropriate target of ERAs every time she lies-to-outdo-others, because this upbringing suggests that she does not have a sufficiently sound and unimpededly operating power of reason in lying-to-outdo-others. However, the reasoning in the preceding paragraph has a negative implication on why Ilana does not have such a power in her lying-to-outdo-others: it is *not* simply because her being taught that it is always good to outdo others was a bit of causality alien to her. So, what (else) about Ilana's case suggests that she does not have a sufficiently sound and unimpededly operating power behind her behavior?

What the agent can know spontaneously, I argue in §4.4.1, is not adequate for explaining why they are acting in a particular way rather than acting not in that way insofar as this behavior is incorrect. This argument will bear on what sense it can make to ask whether one has a free will or whether one's will is free.

In particular, one might suppose that, to have a free will, one need not be willing anything – just as freedom to do what one wants (in the sense not implied by freedom of the will) requires not that one be doing anything but that one be unconstrained. In §4.4.2, I argue for the contrary: to have a free will is always to have a free will *in some willing*, i.e. *in some way of intentional acting*, and this willing needs to be in a certain orientation. In Ilana's case, it will emerge, *because* her will in lying-to-outdo-others is not oriented in a certain direction, she does not have an sufficiently sound and unimpededly operating reason in this behavior and, so, does not have a fully free will in this behavior.

One might also suppose that, should one have a free will, *every* action they seemingly deliberately choose must be free and that whether one has free will is a single fate fixed for all (healthy adult) human beings (in this world) for the same reason. These suppositions about how free will is attributed bear on what can be concluded from certain findings in experimental psychology. These findings claim that a certain stimulus had influenced what the observed human subjects consciously chose, even though these subjects failed to take this influence into account in their verbal reports of how they arrived at their choices. Grant for the sake of argument that, first, this claim is true and, second, it means that some historical

human intentional actions have at least one etiology that the agent can only find out, i.e. one alien etiology. One may be tempted to conclude that the choices in those cases were not free. From such cases, where merely discoverable explanations are indispensable, one may be tempted to draw one or both of the following further conclusions: the agents in those cases do not have free will at all and *we humans* do not have free will.

I do not accept the first inference, the inference made about those cases under study: even given the fact that a choice is traceable to some causal lineage merely discoverable to the agent, this does not *follow* that the choice was unfree – as I already explained in my refutation of the No-alien-causality View above. That being said, I will go on to draw out an implication of *Free Action* that amounts to somewhat of a concession. If the agent's choice is in fact incorrect according to some standard of reason, what is spontaneously knowable to the agent themselves is insufficient for explaining why they make that choice in their conduct rather than not. In this case, an adequate explanation would call for a causality or etiology that is alien to the agent. With the additional supposition that the choice is incorrect according to reason, the causal profiles identified by e.g. empirical psychology may be necessary to explain the behavior. So, I agree with Wolf insofar as she thinks that, if an action is contrary to reason, that the agent not be psychologically necessitated to engage in it is a necessary condition for it to be free. However, on my view this necessary condition *trivially* holds because an action contrary to reason is unfree anyway.

One may seize on my concession and extrapolate the indispensability of alien explanations of all actions from some, which turn out to be contrary to reason. For one may believe that merely discoverable causalities are equally indispensable for explaining actions contrary to reason and actions in conformity with reason alike. From this supposed symmetry, one may draw one or both of the following further conclusions: alien explanations are applicable to anything an individual does so as to preclude freedom, and *we humans* do not have a free will in anything we do. The first further conclusion, however, seems to presuppose that having free will is an all-or-nothing concept in the sense that, if one's will in a certain action they seemingly deliberately choose is unfree, their will (in *every* action they seemingly deliberately choose) is generally unfree. The other further conclusion seems to presuppose, in addition, that whether one has free will is a single fate fixed for all healthy

adult human beings (in this world) for the same reason. I resist these further conclusions: both suppositions that facilitate them are false. Against the first supposition, I contend the following asymmetry. Merely discoverable causalities are not needed to explain an action that is an ideal manifestation of reason in a way they *are* needed to explain an action that is not such a manifestation. Accordingly, whereas an action contrary to reason cannot satisfy the criterion of *Free Action*, an unimpeded manifestation of a (relevantly) fully developed power of reason does satisfy it. I argue for this asymmetry by drawing on a distinction I already established above in the case of Anali, between two kinds of explanatory questions, why something occurs and what it is, and between why-answers that only address the first and those that also address the second. Against the second supposition, I argue in §4.4.4 that how free an action is in part depends on how developed the agent's discerning power of practical reason is, which certainly can vary from one healthy adult human being to another.

If this asymmetry holds – and if actions that are ideal manifestations of reason do not need merely discoverable explanations in a way that conflicts with freeness –, then one cannot draw from an occasion on which merely discoverable causalities are necessary to explain someone's action the conclusion that such causalities are likewise indispensable for this agent's actions.

According to this conception, there is an ideal or ideals of freedom which we humans can approximate more or less, toward which we can progress. Yet, contra Wolf, having a free will will never stop short of expressing itself in one's behavior; having free will is never a feature that merely places us in the moral realm; it always displays itself in correct actions.

4.4.1 Free action as the manifestation of reason

Free Action constitutes the first and most important premise for the arguments advanced in the rest of this chapter. For starters, the connection it identifies between free action and spontaneous possession of explanation enables us to see free action's dependence on an order of practical reasons:

1. (*Free Action*) One is free in acting γ -ly just insofar as a fact why they are acting γ -ly is represented by themselves in such a way as to constitute why they are acting γ -ly rather

than not γ -ly.

2. Only the power of reason is capable of representing some why or possessing why-answers constitutively.
3. An agent is free in acting γ -ly only if this is a manifestation of their power to act for a reason. [1,2]
4. One does not count as having such a power unless it is a discerning power by whose operation they can be correct in doing what they are doing for the reason for which they are doing it.

Since it has been suggested by some, e.g. [Setiya \(2009b\)](#), that the capacity to act for reasons does not entail the capacity to act well, my premise four can use some defense. To have any power of reason is in part to understand a why-question, and one understands a why-question at all only as far as they understand (perhaps without articulate grasp of a rule) what kind of thing a legitimate answer can cite to provide a reason. In particular, a subject understands a why-question of the sort under investigation – that which marks an intentional action – only as far as they understand reason-giving answers of a specific class, namely, answers that reveal what Anscombe might call motives. All reason-giving answers to our why-question reveal motives: they all interpret the queried doing, putting it in a certain light.⁷ Further and more pointedly, the agent must be able to discern certain motive-revealing answers as apt for a given course of action *in and only in* certain circumstances.⁸ This presupposes a distinction between correct and incorrect pairings of an action and something as its reason, given certain circumstances. Only from such a repository of answers – motive-revealing answers to which certain constraints of aptness apply – would the agent’s possession of a specific answer distinguish their doing from doings with merely discoverable etiologies. But to have mastered some such repository of answers is to be discerning and have some ability to make the right call. Therefore, the power to act for a reason is a discerning power by whose operation one can be correct in doing what they are doing for the reason for which

⁷ A motive-revealing answer may cite something in the future relative to the doing (e.g. ‘in order to win their favor’, ‘because I won’t see them again’), something in the past relative to the doing (‘because they did me a huge favor’), or neither (‘out of love’).

⁸Such discerning understanding may include the ability to abort what one is doing or revise one’s why-answer in view of new evidence.

they are doing it.

That said, premise four itself leaves open where there is a distinction between correct and incorrect operations of reason and where there is none. This premise can be true, for example, even if instrumental rationality is the only norm that distinguishes correct from incorrect practical reasoning; it does not entail that immoral action is contrary to reason. What it does assert is that practical reason is a power some possible operations of which are subject to a standard of correctness, that the distinction between correct and incorrect must apply to *some* occasions of its use.

5. That one is *correct* in doing something for some particular thing as a reason means that there is an order of practical reasons; for a power to discern such correctness from incorrectness is for the power to be tuned into this order of practical reasons.
6. What one has is a power to act for a reason only if there is an order of practical reasons into which this power is somewhat tuned. [4,5]
7. An agent is free in acting γ -ly only if this is a manifestation of their power to act for a reason, which power is tuned into an order of practical reasons. [3,6]

In the next several subsections, I will determine further criteria for one to be free in one's action: the action must not be incorrect and the power of reason must be fully developed for the relevant domains of possible considerations in the situation.

4.4.2 Incorrect action occasions a need for merely discoverable explanations

One might think that, if what is spontaneously knowable to the agent can be adequate for explaining some action if it is a manifestation of their power of reason, this can be the case even if the action is somehow incorrect, e.g. if the agent acts as they do only because the agent is misinformed on or misremembering something relevant. In short, if any action can be free at all, it seems that, according to my account, some incorrect actions can be. This is an illusion, however. For if it is incorrect of the agent to be acting γ -ly, then no why-answer they (1) spontaneously possess (2.a) explains why they are acting γ -ly rather than not γ -ly and (2.b) explains that in such a way that it also suffices to explain what they are doing.

My argument is this. Suppose, for the sake of argument, that one is free in an incorrect behavior. According to *Free Action*, then, this behavior would have a causal lineage that is spontaneously knowable to the agent. And (by my premise 2 in §4.4.1) any such constitutively represented etiological connection must be the work of their reason. Therefore, something is an adequate explanation of what this supposedly free action is only if it shows that there is some such rational etiological connection. *In spite of* this supposedly free action's incorrectness, then, an adequate explanation of what this action is still must show how it is a manifestation of reason (in however contorted a way it might be). But precisely in order to show this, the explanation must take into account some etiological lineage that is merely discoverable to the agent. For then the behavior would be understandable as a manifestation of reason only when the operation of reason is shown to be compromised or limited in some definite way. But this compromise or limitation, by definition, cannot be the work of the power itself whose working is being limited, and nothing other than the work of the subject's reason is spontaneously knowable to this subject.

For example, suppose someone is chanting, and they are doing this because they are trying to bring about more rain, as they would sincerely tell us if we ask them. If this chanting-to-bring-about-more-rain is to be free, it must have an etiological lineage that is spontaneously knowable to the agent. If what they would sincerely say – “I am chanting because I am trying to bring about more rain” – is an expression of a genuine spontaneously knowable etiological connection, then an adequate explanation of what this supposedly free behavior is must show how the expressed thought, despite the action's incorrectness, is nonetheless a limited manifestation of this agent's power of reason. The adequate explanation would have to include (1) the agent's false belief e.g. that how much they chant affects how much it rains, (2) their oblivion to a certain relevant fact (e.g. that how much one chants does not affect how much it rains), or (3) the fact that, even though the agent has all and only the true thoughts they should have and makes all the correct judgments they should make concerning their situation, something has so distracted them as to cause them to act not in accordance with their correct judgment. In each case, the adequate explanation is not knowable to the agent otherwise than as something they discover, for the following reasons. In the first case, the action, under its complete description, is doing-A-because-[a

conjunction where one conjunct is *How much one chants affects how much it rains*].⁹ In this case, the agent's adequate explanation for their behavior would have to affirm something false, which is impossible if the explanation is to be spontaneously possessed by the agent, since everything spontaneously possessed is a constitutive representation, and there is no such thing as a false constitutive representation. In the second case, the explanation with the oblivion as part of the explanans could not be spontaneously possessed by the agent, since the fact that one is oblivious to a certain thing is, as such, bound not to be represented or thought by oneself. In the allegedly possible third case, the erroneous behavior, insofar as it is due to the distraction, would not be an expression of thought since the agent is supposed not to have any erroneous thought pertinent to the situation. Thus, the etiological connection between the behavior and the distraction would not be the work of reason and, therefore, would not be spontaneously knowable to the agent.

So, if it is incorrect of the agent to be doing-A-because- p , nothing can (1) explain why the agent are engaged in this course of action rather than not, *and* (2) explain it in such a way as to suffice to explain what this action is, *unless* the explanation cites an etiological lineage is not constitutively represented by the agent. This would mean that whatever explanation they can spontaneously know does not adequately explain their doing A, and therefore does not fully constitute its etiological profile. Given *Free Action*, then,

Not-incorrect. An agent is free in acting γ -ly only insofar as it is not incorrect of them to be so acting.

This is in contrast to Wolf's view. Recall that, according to her view, given that it is wrong of an agent to be acting γ -ly, they may nonetheless be responsible for this conduct, because they may nonetheless have the required ability to act in accordance with Reason, the manifestation of which would be a right alternative for which they would be responsible. And recall that, for [Wolf \(1990\)](#), to have this ability, allegedly necessary for responsibility "for the results of one's wills", is to bear free will, i.e. to relate freely to one's wills (3-4). So, given her position that it is possible for an agent to be responsible for wrong conduct and that having free will is necessary for such a responsibility, she is committed to the possibility

⁹ It is possible (1) for one to be doing-A-because- $\neg q$, even though (2) it is the case that q and, so, (3) no one is doing A because q . This possibility is no more peculiar than the possibility for one to believe that p -because- $\neg q$, even though it is not the case that q and, so, no one believes that p because $\neg q$.

of having the said ability with respect to one's wrong conduct so as to be responsible for it, and this would be to bear free will with respect to one's wrong conduct.

On my view, however, it is impossible to bear free will with respect to one's wrong conduct because, insofar as it is incorrect of them to be acting γ -ly, whatever the doer as such understands about this conduct is not sufficient to explain it (in contrast to acting non- γ -ly). Recall the implication of Wolf's view that one can have a free-will-according ability with respect to one's wrong conduct so as to be responsible for it. My objection above (§4.1.4) against this implication is that there is no way to construe the 'with-respect-to' relation therein without thereby attributing free will and responsibility to someone with respect also to things for which they are clearly not responsible. If this objection is vindicated, then, on this count, my conception of free will – to be precise, my conception of when one bears free will with respect to their conduct –, is preferable.¹⁰

4.4.3 Asymmetry between ideal manifestation of practical reason and incorrect action

If someone's doing-A-because- p is an ideal manifestation of their power of reason, then something spontaneously knowable to them is an adequate explanation for why they are doing-A-because- p rather than acting without it. It is adequate in a way in which what is spontaneously knowable to someone whose conduct is not such an ideal manifestation cannot be adequate.

As I contended when refuting the Ultimacy View, the sort of why-question with respect to which the agent freely relates to their action is peculiar in that a proper answer to it should contribute toward revealing what the action is more exactly. In this explanatory context, something spontaneously knowable to the agent is an adequate explanation for why

¹⁰ My view is open to the possibility that someone's action is correct and reasonable in some but not all respects, so that the agent is thereby free in those respects but not free without qualification with respect to the action. In such a case, although what is spontaneously knowable to them is not sufficient to explain why they are acting γ -ly rather than acting non- γ -ly – i.e. in contrast to *the* alternative –, it may suffice to explain why they are acting γ -ly in contrast to *some* one or a couple of the more specific alternatives. Accordingly, a behavior that meets the criterion for *Free Action* would turn out to be a special, paradigm case of what meets the criterion for *Free Action-in-contrast*: One is free in acting γ -ly-rather-than- δ -ly just insofar as a fact why they are acting γ -ly is represented by themselves in such a way as to constitute why they are acting γ -ly rather than δ -ly.

they are acting γ -ly rather than acting otherwise, if this conduct is an ideal manifestation of practical reason. In such a happy case, even though it can also be the case that the agent is doing-A-because- p rather than acting otherwise due to some causality they can merely find out, e.g. they have a certain upbringing, this merely discoverable causality is only an enabling condition for the fact that their discerning power of practical reason is manifesting itself.

Now suppose Anali's upbringing has fostered in her, among other things, a power of practical reason that is fully developed for considerations in domains relevant to her present situation. Suppose, in the present situation, the order of practical reasons requires that Anali behave honestly on the ground that it would otherwise be unfair of her, and her power manifests itself: Anali is acting with complete honesty because it would otherwise be unfair of her, and this behavior is an unimpeded manifestation of her power of reason. In a sense it is true that Anali is behaving in the said way *because* she was brought up thus and so. Yet, what Anali spontaneously knows – ... because it would otherwise be unfair – is adequate as the explanation for her being-honest-out-of-fairness rather than acting otherwise. To answer the why-question in such a way as to reveal what the action is exactly, 'For it would otherwise be unfair' is adequate without any need to cite a merely discoverable explanation. The spontaneously knowable explanation, which invokes only causalities facilitated by a relevantly developed power, is enough to explain why the power issues in an activity proper to it. The fact that Anali is behaving honestly because it would otherwise be unfair, rather than acting otherwise, occasions no explanatory need that cannot be met by etiological connections spontaneously knowable to Anali.

To be clear, the asymmetry in agential explanatory self-sufficiency I contend, between an ideal manifestation of practical reason and an incorrect course of action, pertains to the agent's behavior, acting γ -ly as contrasted with its practical alternative: the agent's *acting non- γ -ly*. As to why Anali is doing-A-because- p *rather than being vaporized by a nuclear explosion*, for instance, it is no part of my view that such a contrastive fact can be explained by the sort of explanation spontaneously knowable to a human agent. It is my view that a human being can be acting on their free will, or in this sense acting freely, with respect to *alternative courses of actions*. It is not my contention that it is entirely up to someone

with the power of practical reason to *subsist* as such, or that it is entirely up to a human individual to come to have the power of practical reason.

4.4.4 The need of sufficient development of reason and the consequent rarity of free action

I have thus far argued an asymmetry between an incorrect action and an action that is an ideal manifestation of reason, which includes acting correctly for the right reason. But for an action, acting γ -ly, to be such an ideal manifestation that what the agent's spontaneously understands suffices to explain their behavior, so that they are free in it, it takes more than acting correctly for the right reason: the power of reason that manifests itself must also be fully developed for considerations relevant to whether the agent should be acting γ -ly or not γ -ly in the given situation. I will explain why a fully free action requires that. An interesting consequence is that, because relatively few individuals who behave correctly are in fact fully cultivated for acting in the situation, ideal manifestations of practical reason may turn out to be very rare, among all the things that people can be said to do intentionally. And this, it follows from my account, will also be how rare free action is.

Even if the agent's activity of practical reason issues in a correct course of action, doing-A-because- p , the explanation(s) they can spontaneously know may still not be adequate for explaining it in contrast to some specific alternative course of action, thus inadequate for explaining it in contrast to *the* alternative, namely, acting without doing-A-because- p . For insofar as the agent's discerning power of reason is not fully developed for the relevant domains, they might err if the circumstances were somehow different, so that it is lucky for them that they are acting in the way they do in contrast to some incorrect alternatives.

For example, suppose Ana's upbringing has cultivated in her enough of a constitution so that she cares, among other things, to be fair toward others. Today, this disposition manifests itself: Ana is acting honestly in her interaction with competitors preparing for the same qualifying exam because it would otherwise be unfair of her. Suppose there is no other way to behave correctly, in Ana's situation, than to be-honest-because-it-would-otherwise-be-unfair. However, suppose Ana's discerning power of practical reason is not fully developed

for considerations relevant to her present situation; it is not perfectly tuned into the order of practical reasons in the relevant domains. This means that her power would err in the present situation if it were different in some way. If this test were, say, the final selection process for a specific job opening rather than a mere proof of aptitude and Ana knew this, she would misinform or mislead her competitors about the test in order to outdo them.

In general, then, it seems that the underdevelopment of someone's practical reason means that the presence of some factors can make them act otherwise than how they ought to be acting, doing what they ought not to be doing or acting for a reason they ought not to be acting for. I will call such a factor a distracting factor for the underdeveloped agent and signify the presence of some such factor as the fact that d . Even when the agent is in reality doing the right thing for the right reason, doing A because p , there are factors the absence of which is indispensable for explaining why the agent is doing-A-because- p rather than doing something incorrect, A', or acting for an incorrect reason, that p' .¹¹ Someone whose rational power is well tuned into the order of practical reasons would still do-A-because- p even if d . So, the fact that the underdeveloped agent is also doing-A-because- p , *as opposed to a certain incorrect alternative course of action*, is contingent on an etiological connection that is not the work of their power of reason. Therefore, the etiological fact that is spontaneously knowable to this agent to be why they are doing A, which fact obtains in virtue of the very operation of their rational power, is inadequate for explaining why they are doing-A-because- p instead of doing something incorrect or acting for some wrong reason. Thus, given *Free Action*,

Developed. To the extent that it is either correct or incorrect of the agent to be doing-A-because- p (in their present situation), they are doing A freely only insofar as their power of reason is developed for considerations relevant to whether they are to do A or not (in the given situation).

Given that one human can be more developed in their practical reason than another, *Developed* refutes the other supposition mentioned in §4.4, namely, that whether one has free will is a single fate fixed for all healthy adult human beings (in this world) for the same reason.

¹¹ Note that the wrong reason p' for which the underdeveloped agent acts need not be the same as the presence of the factor, that d , that distracts them from acting for the right reason, as our example above illustrates.

4.4.5 A free action need not be uniquely correct

For an action to be free, according to my view, it need not be the *uniquely* correct course of action. So long as it is one of the alternatives that are not incorrect, a causal lineage that is spontaneously knowable to the agent may be adequate for explaining why they engage in it in rather than acting without engaging in it. Suppose I am at the grocer, and it is not incorrect of me to pick up a watermelon because it is sweet, nor is it incorrect of me to act without picking-up-a-watermelon-because-it-is-sweet. Let's say, given my circumstances, the fact that the watermelon is sweet does not require me to pick it up but suffices as a reason for me to do so, and the fact that the watermelon is heavy does not require me not to pick-it-up-because-it-is-sweet but suffices as a reason for it. So, if I am (intentionally) picking-up-a-watermelon-because-it-is-sweet, then the causal lineage, that I am engaged in this action because the watermelon is sweet, which is spontaneously knowable to me, suffices to explain why I am engaged in this action rather than proceeding without it.

4.4.6 A passion-governed behavior may also be free

If my analysis of the above example is sound, then it is evidence that even an action that is in some sense governed by the agent's passions, i.e. in some sense pathonomous, may also be free. By 'passion' here I mean a state in which one is inclined to act in a certain way without being committed to so act, a state that may be introspectively accessible but only as something one encounters or, at any rate, learns of oneself.¹² By this definition, appetites and sexual desires are often, if not always, passions. Now, what the grocery example above shows is that, if neither alternative course of action is incorrect of the agent to pursue, then even if they engage in one of them in order to satisfy or tend their passion, e.g. one is picking up a watermelon because it is sweet or because they are tending their appetite for something sweet, they may still represent the etiological lineage of the action so as to constitute why they are engaged in it rather than acting without it. In other words, they may still be free in the behavior under that very description. But this can happen only when the agent is tracking the fact that the option they pursue is not incorrect, only as long as they tend

¹² In defining 'passion' here I am helping myself to some useful distinctions made in [Schafer 2013](#), 271.

their passions *as* within the prerequisite confine of what is not incorrect. In this way, one's passions may be a source for an adequate spontaneous explanation for why they are acting γ -ly rather than acting not γ -ly.

4.5 CLARIFICATIONS AND REPLY TO AN OBJECTION

It is my view that free action is the ideal manifestation of practical reason – action expressing the unimpeded operation of one's power of reason that is fully developed for the relevant considerations –, and that such an action is possible even if (1) the agent is so necessitated that they could not act otherwise, (2) the action has an ultimate cause other than the agent, (3) it has a causal lineage that is not knowable to the agent otherwise than as something they find out, and (4) its ultimate spontaneously knowable causal lineage has itself a causal lineage that is merely discoverable. Accordingly, I reject incompatibilism and several other influential rationales for pessimism concerning the general possibility of free will. In other words, I have defended compatibilism and, more broadly speaking, made conceptual room for free will *in the face of* the four conditions above, which are more or less influentially held to preclude free will. The general spirit is to resist such reasoning against the possibility of free action: if we are committed to so much as that we engage in intentional actions – a commitment embedded in our practice of requesting and accepting reasons *from* each other *for* each other's behaviors –, then nothing we subsequently *learn* about human beings can entitle us to draw the sweeping conclusion that free action is impossible. However, unlike many compatibilistic positions – including the one represented in [Wolf 1990](#) –, on my view, one is not fully free in incorrect actions and in the operation of their reason insofar as it is not fully developed for possible considerations relevant to their situation.

I take incorrectness of action to preclude full freedom in the action because, again, in that case, nothing spontaneously knowable to the agent would be adequate for explaining why they are acting in the given way rather than not in that way. This is so, as long as the incorrectness is a fact and the manifestation of the power of reason does not accord with this fact. This is so, even if it is not *irrational* of one to be mistaken about that fact or oblivious

of it. Thus, so long as there is such a fact as that it is ethically wrong to be acting γ -ly, one could not be fully free in such an action, whether or not ethical rationalism is true.

One may protest that my compatibilistic view has little to offer by way of justifying or making sense of blame, resentment, and punishment – which is perhaps the main motivation behind the majority of efforts to defend compatibilistic positions. I will address this in the conclusion of this dissertation.

Here, I will address instead a concern whose grip on my view is compounded by the very way I argued against the potential rationales for free will pessimism. According to my view, if it is incorrect of someone to be acting γ -ly, they are not free in so acting, because nothing spontaneously knowable to them suffices to explain why they are acting thus rather than not in this way. Some merely discoverable causality must be taken into account, if there is to be an explanation. But whatever this alien causality is, the agent is not fully free in the action. For example, suppose it is a fact that one should have concern for others. Now, imagine a driver who is wrong in going on their way, without even stopping, after running over a cyclist. Whether this driver so acts because they were spoiled and never taught to have concern for others, or because they had a deprived childhood in which they knew no affection or care and was not taught to have concern for others, (or because of something else,) they are not fully free in this hit-and-run, according to my view.

This seems to conflict with a “sense of difference” that Martha Klein (1990) points out, namely, that the deprived offender “merits some some sort of special consideration, that he should be judged differently from the non-deprived offender” (76): the offender should be treated with more sympathy and less harshly insofar as their offense is attributable rather than not attributable to the deprivation. This sense of difference, Klein argues, is best explained by the thought that the wrongdoer’s history of deprivation cancels out some amount or severity of “blame (condemnation or punishment)” that the wrongdoer deserves for their wrong doing that is attributable to the deprivation, since in being ‘shortchanged’ in the past they already ‘paid’ some for their present behavior. (77-84) Therefore, any view that implies that the offender is not blameworthy in the first place, apparently, cannot accommodate the sense of difference.

Given that I, when evaluating different conceptions about free action, appeal to our practice of discounting considerations and the asymmetry in how we take someone's upbringing to bear on their desert, I cannot simply brush aside the present "sense of difference" as misguided common sense. In fact, I think there is something to it and we can make sense of it within my framework. My response is twofold. Firstly, my view does not preclude the possibility of a wrongdoer's blameworthiness, although it is my view that a wrongdoer is never free insofar as their doing is wrong. For people are blamed for results of behaviors in which they are not assumed to freely engage, e.g. harms or losses due to a careless or negligent behavior, and my account implies nothing critical of this kind of blaming.¹³ Secondly, even setting aside my view's tolerance toward blame, it can accommodate the sense of difference Klein invokes. Since to treat someone more kindly and with more love is a way to treat them less harshly, we can treat a wrongdoer less harshly (insofar as their wrong doing is attributable to the deprivation) without judging anyone to be blameworthy or dealing out different amounts of blame. There are likely to be many cases in which the deprived offender "should be judged differently from the non-deprived offender", but this does not mean that the judgment has to be, or imply, that the offender of either history deserves any blame or more blame than the other.

¹³ My view does suggest, by eliminating freedom-in-action as a form of justification, that if blame is ever reasonable, the severity of reasonable blame may never be as high as we pre-philosophically think it can be.

5.0 CONCLUSION

Prima facie reasons for not treating someone as a fully appropriate target of ERAs in general for a behavior are all accounts of the behavior that are merely discoverable, even to the agent themselves. One might infer that, as soon as we identify for the action a causal lineage merely discoverable to the agent, the agent cannot be fully responsible or free in relation to this conduct, i.e. the No-alien-causality View. But this oddly implies that a foolproof education of practical wisdom would discredit the educated person. So, I explore an alternative construal: those prima facie considerations all indicate that the agent is not in spontaneous possession of an adequate explanation for their engagement in the behavior, and this inadequacy occasions the need for a merely discoverable explanation. Thus, I defend *Free Action*: one is free in acting γ -ly just insofar as a fact why they are acting γ -ly is represented by themselves in such a way as to constitute why they are acting γ -ly rather than not γ -ly. Furthermore, a subclass of those prima facie reasons, namely, the discounting considerations, indicate specifically that the agent lacks a power of reason that is operating, unimpeded, and sufficiently developed with respect to the behavior in question. To manifest unimpededly such a power of reason, then, seems to me to be just what it takes for a behavior to fulfill the condition in *Free Action*.

Although the asymmetry between behaviors that are incorrect and not incorrect inspires me, it also poses what I perceive to be the greatest challenge in this project. Some discounting considerations that work in a case where the conduct is (supposed by the one who accepts the discounting consideration to be) wrong seem to have no counterpart that works in the case of wise conduct: *to act wisely out of addiction or due to brainwashing* is an oxymoron. I take this asymmetry as something to be accounted for. I am still not certain how to go about this accounting for the ‘unhappy’ half of the asymmetry, but my current account is that

wrong conduct itself is something that cannot be adequately explained by causal lineages that are constitutively represented by the agent. A conduct that is wrong (to someone, x) is not fully spontaneously explicable (to x), in turn, because whatever explanation for the conduct the agent may spontaneously possess is not an explanation that (to x) clearly can be spontaneously possessed. That is because an explanation clearly can be spontaneously possessed only if it clearly displays the causal lineage as the work of a power of reason. But when a conduct is (to x) wrong, it follows that either the conduct does not make complete sense (to x) or its agent fails to reflect acknowledgement of (what is to x) a fact in their conduct. Either way, it calls for something outside the functioning of the agent's reason to show how exactly, despite the problem, the conduct nonetheless manifests the work of their reason.

When a subject is acting γ -ly intentionally, they constitutively represent the what and a why of this doing, i.e. they constitutively possess the what-answer and a why-answer for this doing. What is affirmed in this consequent, the what/why constitution condition, is especially important if the agent is to bear free will in the doing. It is especially important since, on my view, to bear free will in a course of conduct is for one's spontaneous why-answer to be adequate for explaining it in contrast to acting not in that way, i.e. for a fact as to why the agent is acting γ -ly rather than not γ -ly to be constitutively represented by the agent.

Moreover, one can be doing A intentionally even if this doing does not involve what is transparently their exercise of a capacity to do A in a context familiar to them in every way relevant to its exercise, even though this, the familiar exercise condition, is an attribute of the paradigm of intentional action. It is also my view that the familiar exercise condition is not needed for one to be fully free in the doing (although there is no need to commit to this to show the possibility of free action in the face of conventional concerns about the possibility of free will), for uncertainty about one's success in execution does not seem to preclude one's actual doing from being expressive of their will or the expressed will from being a manifestation of a sufficiently developed and unimpededly operating power of reason. When this power thus manifests itself in the subject's way of acting, this agent spontaneously has an explanation for why they are acting in this way rather than not in this way. This

last condition is not met whenever the behavior manifesting the agent's power of reason is incorrect or involves an incorrect belief, or whenever their power of reason manifested therein is itself not sufficiently developed for the behavior's situation, according to my view.

Despite that, my conception of free action itself does not preclude ethical nihilism. Nor does it entail ethical rationalism. It does not preclude nihilism because it leaves open whether there are moral facts, whether e.g. any supposedly immoral act is genuinely incorrect. And if there is such a moral fact, my conception of free action leaves open whether the explanation for why a certain doing is immoral is that it is irrational or its agent reasons badly somehow. Thus, my conception of free action does not presuppose that ethical norms are derivable from the nature of reason. Nonetheless, *if* there is so much as an immoral way of acting, so that it is wrong to act that way, then it follows from my view that the agent cannot spontaneously have an adequate explanation for why they are acting in this way rather than not in this way – whether or not the immorality of the action is imputable to a failure of the agent's reason. From this it follows that, in a world with moral facts of the form we saw, i.e., that it is immoral and thus wrong to act in such-and-such a way, moral knowledge will be important at least in two ways.

First, it will be important for being free in action. For example, if someone is mistaken in thinking that abortion is murder, then if on this belief they blockade a clinic, their action cannot be completely free. And if one is even just not sure whether it is permissible to act in a certain way, one will be at best somewhat lucky in acting that way. This seems to preclude the agent from being fully free in so acting.

Second, moral knowledge will be important for assessing whether someone is free in their action. To know whether someone is acting freely, I will have to know whether their reason for their action is good. If there is such a thing as an immoral way of acting, so that there is a fact as to whether a given course of action is immoral, then in order to know whether someone is fully free in their engagement in an intentional action, one must know whether that action is morally permissible. So, when it comes to a controversial issue or morally difficult situation, e.g. whether physician-assisted suicide should be made legal in a society in certain conditions, one cannot know whether e.g. a single lawmaker is fully free in e.g. legalizing it unless one knows whether it is permissible to legalize the practice.

On my view, action in which one is fully free is possible in the very world we live in, even should it turn out to be causally deterministic, and even given the success and promise of disciplines that represent certain causal lineages of our behaviors as lineages we discover. This is my view because my conception of free action entails compatibilism and opposes free will pessimism in face of several other perceived blockers of free will. It follows from this same view, however, that action in which one is fully free might be a lot rarer than it is thought to be by many compatibilists, free will optimists of other sorts, and libertarians. For full freedom requires the agent's power of reason to be so developed for all domains of considerations relevant to their situation that the agent would not have acted wrongly even if all sorts of distracting considerations were present. Free will optimists might complain that my view, after all, has little to offer by way of justifying or making sense of primary cases of blame, resentment, and punishment. First of all, it should be clear that, since my view is not at all in tension with the possibility of immoral doings, there can still be straightforwardly, rationally justified disapproval and indignation toward someone for their immoral behavior, on my view. Things are more complicated with blame, resentment, and punishment. For if we blame or resent someone for their wrong behavior, we impute the behavior to them as if they need to pay for it. Getting punished, and even getting resented, might be thought to be a way for the wrongdoer to pay for their wrong doing. It is true that my view denies one straightforward way blame, resentment, and punishment can be rationally justified for their perceived primary application to an action, since I argue that a wrong action committed in full freedom is impossible. However, as I already mentioned in §4.5, my view is tolerant toward the sort of blame that is proportional to careless or negligent behaviors. As for resentment, even according to our commonsense, even when we judge that a wrongdoer deserves to be resented by the victim, it is not as if we think that the victim owes it to anyone to resent the wrongdoer: it is all right if the victim does not resent. There is a commonsensible asymmetry between this and when we judge that an agent deserves recognition or to be thanked by the beneficiary: here, appreciation is judged to be owed the agent and there is also such a thing as being ungrateful. This suggests that circumstances can be such that it is understandable if the victim in fact resents someone for having wronged them, but not such that resentment is rationally required of the

victim. My view, which denies action-directed resentment the possibility of one main form of rational justification, can nonetheless accommodate a distinction between unreasonable and understandable resentment. It is certainly friendly, however, to the idea that it is rational and laudable of a victim not to resent the wrongdoer where any such resentment would be understandable. As for punishment, my view does not lend support for a retributive justification. If it is ever justified to punish someone for having done something wrong, the ground would likely be something else, on my account of free action.

APPENDIX

MUST INTENTIONAL ACTION MANIFEST AN UNTRIGGERED POWER?

There is another kind of argument for incompatibilism that is quite different from the ones discussed hitherto: a causally deterministic world does not even allow the possibility of intentional action. In the version of this argument advanced by Dawa Ometto (2016),¹ the more specific ground is that an agent is acting γ -ly intentionally (and therefore can have free will in so acting) only if their so acting manifests a power that has no trigger. Whether a power has a trigger in Ometto's sense comes down to this: a power to do F has a trigger just when what is in possession of this power, x, is such that x does F when... , "where the blank indicates something that *works on*, or *affects*" x (214). And, Ometto argues, "[a] subject precisely does not acquire practical knowledge via some object's affecting or working on her" (214). Here is my regimented reconstruction of his argument in 212-7, including only commitments I deem essential to it:

1. If one is doing A intentionally, that they are so acting is something they can have practical knowledge of.
2. Practical knowledge of one's doing A includes practical knowledge of why one begins

¹ It was thrilling to meet Dawa at Leipzig in summer 2016, who had just defended his dissertation, *Freedom and Self-Knowledge*. Like me, Dawa also brings the spontaneity of intentional action to bear on whether such action is possible in face of some traditionally perceived blockers of free will, but he focuses more on the metaphysics of powers. Even though I disagree with his incompatibilist conclusion, he argues powerfully and thoroughly, and reading someone well versed in the same traditions I build upon has helped me engage the dispute with a sharper focus.

doing A at the time one begins.

3. Insofar as the power to act manifested in one's doing A has a trigger, why this agent begins doing A at the time they begin is not something they can have practical knowledge of.
4. An agent is doing A intentionally (and therefore can have free will in so acting) only if their so acting manifests a power that has no trigger. [1,2,3]
5. Every power to act in a causally deterministic world has a trigger.
6. No one can be acting intentionally in a causally deterministic world. [4,5]

Charitably interpreted, premise (2) says that there is always *some contrastive description* of when one begins doing A, *one begins doing A at t1 rather than t2*, under which the agent's intellectual relation to their doing A covers why they begin doing A at t1 rather than t2. For example, as to why one begins adding broth to the skillet *after adding rice rather than before*, the agent may spontaneously possess the answer, '... because I am making a risotto – as opposed to e.g. a rice porridge – and this requires that the uncooked rice be fried a bit'. But as to why they begin adding broth to the skillet at 13:00:00 rather than 12:59:59, the same agent may possess no explanation. For the sake of argument, let us grant premise (2) thus construed.

Accordingly, premise (3) says that, insofar as the power to act manifested in one's doing A has a trigger, there is no contrastive description under which why this agent begins doing A when they do is possibly subject to their practical knowledge. The reasoning Ometto ultimately settles on seems to be this. If a power is triggered to manifest itself in a certain way, it is acted on by a trigger, and this triggering is other than the manifestation of the power itself. So, given that why an agent, x, begins doing A when they do is explained by a trigger's triggering their power to act, it is explained by a cause whose causality is other than the manifestation of their power to act. But what is available to one's practical knowledge must be the manifestation of a power for constitutive representation. Therefore, why x begins doing A when they do, regardless of how this timing is contrastively described, is not explained by something knowable to x practically.

But why might one think that the explanandum, why x begins doing A when they do,

under some contrastive description, cannot have both a trigger-involving and a practically knowable explanation? Ometto's argument, in turn, seems to be this. The power to act intentionally enters into the activity of doing A only if it so comes into activity *because* the subject judges that it is the right time to do A (214, 220). Further, Ometto seems to assume that, if something acts on a subject so as to trigger indefeasibly their power for acting, at t1, to do A, then even if the subject at any point also judges that t1 (under whatever description) is the right time to do A, the onset of their doing A is not properly connected – in this sense *accidental* – to this judgment, so that the judgment cannot figure in an etiological connection practically knowable to the agent in such a way that the connection explains why they begin to do A when they do under any contrastive description.

This resembles the argument for the Paradox we saw in §4.2: both that and the present argument contend that, if one's behavior is necessitated by a factor other than oneself in a certain way, one cannot be acting genuinely out of one's representation of the behavior's rightness. My response here is accordingly similar. Even if a cause causes indefeasibly, even if it is other than the agent and precisely other than their power to act, and even if its causality is merely discoverable to the agent and precisely other than the manifestation of that power, its effect can still be the agent's appreciation of what they should appreciate. And even if what they thereby appreciate is something they discover, they can nonetheless constitutively represent an etiological connection in which what they appreciate figures as a cause and a bit of their behavior figures as the explanandum. Here is an example of the possibility I have in mind. Suppose Anita is such that she always acts honestly in her situation as soon as something in the situation has affected her in such a way that she has come to know and judge that honesty is the only fair way to act now. So, by Ometto's definition, now Anita's power to act has a trigger. But this triggerableness does not preclude Anali from acting-honestly-because-it-would-otherwise-be-unfair. The fact that acting dishonestly would otherwise be unfair in a situation can figure as a cause in an etiological connection that is practically knowable to Anita so that this connection explains why she begins to behave honestly when she has come to know that it would otherwise be unfair. Nor does the triggerableness preclude the onset of Anita's honest behavior from being explained by her judgment that now, rather than any later, is the time to start acting honestly.

BIBLIOGRAPHY

- Adams, F. and A. Mele (1989, December). The role of intention in intentional action. *Canadian Journal of Philosophy* 19(4), 511–532.
- Albritton, R. (1985, November). Freedom of will and freedom of action. *Proceedings and Addresses of the American Philosophical Association*, 239–251.
- Anscombe, G. E. M. (1963/2000). *Intention*. Harvard University Press.
- Boxer, K. (2013). *Rethinking Responsibility*. Oxford: Oxford University Press.
- Bratman, M. (1984, July). Two faces of intention. *Philosophical Review* 93(3), 375–405.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, Massachusetts: Harvard University Press.
- Chisholm, R. (1964). Human freedom and the self. The Lindley Lecture, pp. 3–15.
- Davidson, D. (1978). Intending. In *Essays on Actions and Events*. Oxford: Oxford University Press.
- Falvey, K. (2000). Knowledge in intention. *Philosophical Studies* 99(1), 21–44.
- Fischer, J. M. (2006). *My Way: Essays on Moral Responsibility*. Oxford University Press.
- Fischer, J. M. and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Ford, A. (2011). Action and generality. See [Ford et al. \(2011\)](#).
- Ford, A., J. Hornsby, and F. Stoutland (Eds.) (2011). *Essays on Anscombe's Intention*. Harvard University Press.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy* 68(1), 5–20.
- Franklin, C. E. (2011). Neo-frankfurtians and buffer cases: The new challenge to the principle of alternative possibilities. *Philosophical Studies* 152(2), 189–207.

- Ginet, C. (1996). In defense of the principle of alternative possibilities: Why i don't find frankfurt's argument convincing. *Noûs* 30, 403–417.
- Grice, H. P. (1971). Intention and uncertainty. In *Proceedings of the British Academy*, Volume 57, pp. 263–279.
- Haji, I. (1998). *Moral Appraisability: Puzzles, Proposals, and Perplexities*. New York: Oxford University Press.
- Hunt, D. and S. Shabo (2013). Frankfurt cases and the (in)significance of timing: A defense of the buffering strategy. *Philosophical Studies* 164(3), 599–622.
- Hunt, D. P. (2000). Moral responsibility and unavoidable action. *Philosophical Studies* 97(2), 195–227.
- Hunt, D. P. (2005). Moral responsibility and buffered alternatives. *Midwest Studies in Philosophy* 29(1), 126–145.
- Kane, R. (1985). *Free Will and Values: Adaptive Mechanisms and Strategies of Prey and Predators*. SUNY Press.
- Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Klein, M. (1990). *Determinism, Blameworthiness, and Deprivation*. New York: Oxford University Press.
- Langton, R. (2004). Intention as faith. In J. Hyman and H. Steward (Eds.), *Agency and Action*, pp. 243–258. Cambridge: Cambridge University Press.
- Marcus, E. (2012). *Rational Causation*. Harvard University Press.
- Marušić, B. (2015). *Evidence and Agency: Norms of Belief for Promising and Resolving*. Oxford University Press, USA.
- McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, J. (2013). Zum verhältnis von rezeptivem und praktischem wissen. *Deutsche Zeitschrift für Philosophie* 61(3), 387–401.
- Mele, A. R. (1995). *Autonomous Agents: From Self Control to Autonomy*. New York: Oxford University Press.
- Mele, A. R. (2006). *Free Will and Luck*. New York: Oxford University Press.
- Moore, G. E. (1912). *Ethics*, Volume 52 of *Home University Library*. New York: Holt & Co.
- Ometto, D. (2016). *Freedom and Self-Knowledge*. Ph. D. thesis, The Department of Philosophy and Religious Studies, Utrecht University.

- O'Shaughnessy, B. (1980). *The Will: A Dual Aspect Theory*. Cambridge University Press.
- Palmer, D. (2011). Pereboom on the frankfurt cases. *Philosophical Studies* 153(2), 261–272.
- Paul, S. K. (2009). Intention, belief, and wishful thinking: Setiya on “practical knowledge”. *Ethics* 119(3), 546–557.
- Pereboom, D. (2000). Alternative possibilities and causal histories. *Philosophical Perspectives* 14(s14), 119–138.
- Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.
- Pereboom, D. (2005). Defending hard incompatibilism. *Midwest Studies in Philosophy* 29(1), 228–247.
- Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Schafer, K. (2013). Perception and the rational force of desire. *The Journal of Philosophy* 110(5), 258–281.
- Setiya, K. (2008). Practical knowledge. *Ethics* 118(3), 388–409.
- Setiya, K. (2009a). Practical knowledge revisited. *Ethics* 120(1), 128–137.
- Setiya, K. (2009b). *Reasons without Rationalism*. Princeton University Press.
- Setiya, K. (2011). Knowledge of intention. See [Ford et al. \(2011\)](#).
- Setiya, K. (2012). Knowing how. In *Proceedings of the Aristotelian Society*, Volume 112, pp. 285–307.
- Setiya, K. (2017). *Practical Knowledge: selected essays*. New York: Oxford University Press.
- Strawson, P. F. (1962/1974). *Freedom and Resentment*, pp. 1–25. Methuen & Co. Ltd.
- Stump, E. (1996). Libertarian freedom and the principle of alternative possibilities. In J. Jordan and D. Howard-Snyder (Eds.), *Faith, Freedom, and Rationality: Philosophy of Religion Today*, pp. 73–88. Lanham: Rowman; Littlefield.
- Stump, E. (1999). Dust, determinism, and frankfurt. *Faith and Philosophy* 16(3), 413–422.
- Stump, E. (2003). Moral responsibility without alternative possibilities. In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate.
- Thompson, M. (2004). Apprehending human form. *Royal Institute of Philosophy Supplement* 54, 47–74.

- Thompson, M. (2008). *Life and Action*. Harvard University Press.
- Thompson, M. (2011). Anscombe's *Intention* and practical knowledge. See [Ford et al. \(2011\)](#).
- Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- Watson, G. (1975, April). Free agency. *Journal of Philosophy*, 205–20.
- Widerker, D. (1995a). Libertarian freedom and the avoidability of decisions. *Faith and Philosophy* 12(1), 112–118.
- Widerker, D. (1995b). Libertarianism and frankfurt's attack on the principle of alternative possibilities. *The Philosophical Review* 104(2), 247–261.
- Wolf, S. (1980). Asymmetrical freedom. *The Journal of Philosophy* 77(3), 151–166.
- Wolf, S. (1990). *Freedom Within Reason*. Oxford: Oxford University Press.
- Wyma, K. D. (1997). Moral responsibility and leeway for action. *American Philosophical Quarterly* 34(1), 57–70.