A rationalist argument for libertarian free will

Stylianos Panagiotou

PhD

University of York

Philosophy

August 2020

Abstract

In this thesis, I give an a priori argument in defense of libertarian free will. I conclude that given certain presuppositions, the ability to do otherwise is a necessary requirement for substantive rationality; the ability to think and act in light of reasons. 'Transcendental' arguments to the effect that determinism is inconsistent with rationality are predominantly forwarded in a Kantian manner. Their incorporation into the framework of critical philosophy renders the ontological status of their claims problematic; rather than being claims about how the world really is, they end up being claims about how the mind must conceive of it. To make their ontological status more secure, I provide a rationalist framework that turns them from claims about how the mind must view the world into claims about the ontology of rational agents. In the first chapter, I make some preliminary remarks about reason, reasons and rationality and argue that an agent's access to alternative possibilities is a necessary condition for being under the scope of normative reasons. In the second chapter, I motivate rationalism about a priori justification. In the third chapter, I present the rationalist argument for libertarian free will and defend it against objections. Several objections rest on a compatibilist understanding of an agent's abilities. To undercut them, I devote the fourth chapter, in which I give a new argument for incompatibilism between free will and determinism, which I call the situatedness argument for incompatibilism. If the presuppositions of the thesis are granted and the situatedness argument works, then we may be justified in thinking that to the extent that we are substantively rational, we are free in the libertarian sense.

Table of contents

Preface

This is a thesis about libertarian free will and normative reasons. It aims to show that beings that are under the scope of normative reasons, are free in a robust libertarian sense. It takes insights from transcendental arguments against determinism and reframes them in a rationalistic framework, in order to free them from the frequently Kantian nature in which they are argued for and the restrictions it poses for the ultimate status of ontological claims. Instead of focusing predominantly on moral reasons, this thesis focuses on epistemic reasons as well. It is held that the conditions for being under the scope of normative reasons in the moral domain are structurally the same with the conditions for being under the scope of normative reasons in the epistemic domain. It is argued that such conditions have to involve the agent's access to alternative possibilities of action. Assuming that knowledge requires our access to epistemic reasons, we are led to the conclusion that to hold each other accountable to reasons, as well as beings capable of knowing that such and such is the case, we have to hold ourselves and others as beings with libertarian free will. Due to the limited amount of space that a PhD project allows for, certain presuppositions are made, which are not defended. Unfortunate though this may be, it is a necessity for a project of this sort. Such presuppositions are: i) the falsity of global epistemic skepticism, ii) an internalist conception of justification, iii) The Kantian OIC principle (OIC), iv) a verification transcendent conception of truth and v) a conception of reasons as irreducibly normative. The topics that this thesis deals with are many and some are treated in an inconclusive manner. All that I aim to achieve in this project is to develop the outlines of a philosophical position, according to which, rather than beings its enemies, rationalism and human situatedness are free will's greatest allies.

Acknowledgements

This work is dedicated to my parents, Matthew Panagiotou and Vicky Kolintza, and my partner Stella Antoniou.

First and foremost, I would like to express my gratitude to my supervisor, Prof Tom Stoneham. He gave me the benefit of the doubt at a time few were willing to and even fewer did. His support throughout the years, his willingness to discuss a wide range of philosophical topics and his emphasis on logical precision have made his contribution to my intellectual development, indispensable. Without any reservation, I can proudly call him 'teacher'. More than anyone, he knows the gravity that this term has.

Second, I wish to express my gratitude to several members of faculty, the administrative staff at the Department of Philosophy, as well as other PhD researchers at the University of York. Special thanks go to members of faculty for whose courses, I taught seminars and with whom we collaborated for administrative tasks. In alphabetical order, they are Dr Stephen Everson, Dr David Ingram, Dr Chris Jay, Dr Nick Jones, Dr Barry Lee, Prof Mary Leng, Dr Christian Piller and Dr David Worsley.

Special thanks go to Dr Janet Eldred for providing me with moral support throughout these years. On several occasions, one of them crucial for this thesis, she was willing to listen to what troubled me and offered her advice in a manner that shows she did and does care.

Third, I want to thank my friend, John Blechl. He has always been there for me and our philosophical discussions have always been breaths of fresh air.

Finally, I want to thank Prof Derk Pereboom (Cornell University) and Prof Alan Thomas (University of York) for agreeing to examine this thesis.

Author’s declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Introduction

This thesis defends libertarian free will. Its goal it to show that free will is a necessary requirement of substantive rationality –the ability to think and act in light of reasons- and our ability to know things to be the case. Various arguments have been given to that effect which are loosely called 'transcendental' and build on the insight that beings capable of reasoning must have a kind of freedom. Nevertheless, they are frequently argued for in a Kantian manner. Such a Kantian treatment of the argument seems to me to be quite restraining when it comes to claims about free will, due to the status of ontological claims in the context of critical philosophy. They are reduced to claims about how we must perceive the world; not how the world is. To do so would be to violate the very tenets of the critical project, according to which, pure reason cannot give us knowledge when it escapes the domain of sensible intuitions. It seems to me though, that those in the business of making ontological claims, such as claims about free will, do seem to make claims of a sort that Kantianism is in tension with. Precisely for that reason, I aim to reframe the insights of such arguments in a rationalistic framework which is more suitable to incorporate ontological claims about reality, among its confines. Unfashionable though it may be, I will argue that it is essential for maintaining the possibility of knowledge.

The rationale of the argument links substantive rationality with the conditions for being under the scope of normative reasons. In the domain of moral reasons, the fact that a person has not discharged a particular moral obligation is not enough to ground the claim that it has been violated. The necessary requirement for a person's violating an obligation is that the person is under that obligation's scope on the first place. Plausibly, epistemic conditions, as well as conditions about the agent's abilities are relevant in grounding claims about whether the person in question can be properly held accountable for discharging the normative obligation in question. Equally, and similarly, claims about normative epistemic reasons –at least, irreducibly normative ones- seem to pose the same problematic. The question arises as to: 'when is it proper

to hold one as rationally accountable?' or 'when can one be properly called irrational?' The thesis argues that sensitivity to the context in which an agent is situated is indispensable for answering these questions. It is a theme throughout the thesis that humans, just like all other living and non-living beings are situated beings; beings in particular circumstances, having one form rather than another, at a given time. Such circumstances present us with influences from a whole host of factors which are historical, social, physiological, etc. I will claim that proper sensitivity to a person's situatedness is indispensable for understanding that person as an agent, and consequently for making claims about when a person is justified or unjustified in thinking something or in some way, as well as justified or unjustified in performing a particular action.

Chapter I concerns reason, reasons and rationality. Reason is defined as a faculty in a rationalist sense; it consists in the faculty of grasping necessary facts about reality or necessary connections [A.1]. Reasons are treated as irreducible 'favourers'; normative considerations which render certain thoughts, evaluations and physical actions justified or unjustified [A.2, A.3]. As normative, reasons raise the question of their scope, which essentially asks for the conditions under which it is proper to judge a person as rational or irrational in their light. It is argued that such conditions involve the agent's knowledge of these reasons as well as the agent's ability to act as the normative reason favours [A.3, A.4] and that for such an ability to ground responsibility of a normative sort, the person who is under the scope of normative reasons must have alternative possibilities with respect to meeting or violating the demand, constituted or generated by the normative reason in question [A.4.1]. Furthermore, it is argued that to think and act in light of reasons cannot be understood in a strictly causal sense. Thus, to think and act in light of justifying reasons cannot mean that one is efficiently caused to think and act by one's mental states [A.4.2].

In the first chapter's section B, two challenges to the normativity of rationality are discussed. The first consists in the claim that there are no reasons to be rational

[B.1, B.2]. According to this challenge, the standards of rationality are not normative because they concern coherence among attitudes, as opposed to responsiveness to normative reasons. It is argued that the challenge is overstated. It is customary to draw a distinction between substantive/objective and structural/subjective rationality. Allowing that structural/subjective rationality is not normative, does not straightforwardly imply that there is no normative agency in rational agency. It would only imply that if a further claim were made, according to which, all instances of rationality are instances of structural/subjective rationality; i.e. if we never had access to substantive/objective epistemic reasons [B.2.2.1]. Furthermore, it is argued that a view that would hold all rationality to be structural/subjective, would have the unfortunate effect that conspiracy theorists would represent rational thinkers *par excellence* [B.2.2.2]. The second challenge to the normativity of rationality that is discussed, stems from moral error theory. According to moral error theory, normative properties are ontologically 'queer' due to their intrinsic 'prescriptivity'; they do not seem to fit well to a naturalistic worldview. Furthermore, knowledge of or access to such properties, would require 'queer' faculties; faculties of mind that do not seem to fit well into a naturalistic ontology of mind. It is argued that: i) if knowledge requires epistemic reasons, then an error theory about the normative faces troubles with accounting for the possibility of knowledge, and ii) that if there is something weird about the faculty of intuition, so worse for any theory that cannot or does not accommodate it [B.3, Ch. II]. The discussion of error theory in B.3 is sketchy. What is needed is a thorough exposition of an epistemology which accommodates the faculty of intuition, which will take place in the next chapter [Ch. II].

Chapter II argues for the epistemological superiority of a rationalist viewpoint over radical empiricism. As such, it plays a dual role. First, it argues for a rationalist epistemology which is committed to the possibility and actuality of a priori justification and to the centrality of the faculty of rational intuition in the pursuit of knowledge. Such a theory allows us to dispense with the mantle of critical philosophy that covers most of the arguments that link rationality with freedom. Second, it allows us to claim

that error theory about epistemic normativity –and moral normativity, I might add- is unmotivated. The concept of a priori justification is conceived of as justification that is independent of sense-experience, rather than 'experience' [A.2]. To do otherwise is to already presuppose what 'experience' means, which leads to a misrepresentation of the debate by assuming that its participants find themselves in agreement with respect to the nature of the faculties that are involved in experiencing the world.

In this chapter's section B, a deficiency argument against epistemologies that do not include rational intuition among their confines is outlined [B.1] and strategies with respect to it –rationalism, Kantianism, logical empiricism and radical empiricism- are presented [B.2–B.2.4]. Due to the limited space allowed for in the thesis, Kantianism and logical empiricism will not be discussed further. The chapter proceeds on the assumption that the strongest objections to the kind of rationalism that is developed and advocated in this project, stem from radical empiricism.

The main challenge posed by radical empiricists towards rationalists consists in the charge that they conflate the limits of their imagination with the limits of possibility. That a contradiction is something that we cannot imagine does not allow us to claim that it is impossible. In the context of this research, this is referred to as 'Mill's thesis'. It is argued that Mill's thesis has quite skeptical ramifications. Eventually, an epistemology that is committed to it –and consequently, committed to a treatment of the law of non-contradiction as holding contingently- loses sight of truth [C.1.1]. Radical empiricists –both of an inductivist (Millian) and of a holistic (Quinean) kind- do accept the law of non-contradiction as the main epistemological principle that instructs us when a belief revision has to take place. A consequence of Mill's thesis is that we can never know whether the circumstances we treat as requiring a belief revision to take place, actually call for it or not. The principle might call for it, but absent a connection between the principle and the world, truth is lost from sight. We may navigate ourselves according to our epistemic compass, although we may never know whether it points north. Even worse, we never know whether revising a belief in

this kind of cases gets us closer to truth or further away from it. Furthermore, Mill's thesis can cause problems with respect to how we understand our own fallibility [C.1.2]. If the law of non-contradiction holds contingently, then we cannot be certain whether we can ever be right as opposed to mistaken, or mistaken as opposed to right, in each particular engagement of ours in a normative, rule-governed activity. Finally, certain more general objections to rationalism are discussed about its putative inadequacy in accounting for knowledge of necessity, as well charges to positing the faculty of rational intuition [C.1.3]. It is argued in response that the debate about a priori justification is an ontological debate about the faculties of mind –which concerns whether they are purely sensory or not- as much as it is an epistemological debate about the scope of the mind's faculties. To argue against the legitimacy of positing a faculty of rational intuition on grounds of it being 'queer' or a remnant of a discredited worldview, is either to misunderstand the debate or not to engage in it seriously.

Chapter III presents the main argument of the thesis, according to which substantive/objective rationality requires libertarian free will. The argument is laid out in section A and it puts together elements from previous chapters that concern the conditions that have to obtain in order for a person to be under the scope of normative reasons, which have been discussed previously [Ch. I/A.3– A.4.2]. Objections that are raised and discussed revolve around a reductionist treatment of the normativity of rationality [B.1], a compatibilist treatment of the principle of alternative possibilities –henceforth, PAP- and the "'ought' implies 'can' principle" –henceforth, OIC- [B.2], the claim that indeterminism is fatal for agency [B.3] and the wholesale or in-part denial of alternative possibilities as a necessary condition for being under the scope of normative reasons and obligations [B.4 –B.4.3]. The penultimate and the final objection stem from an eliminativist treatment of epistemic normativity and concern doxastic involuntarism [B.5], as well as the possibility of a-rationalism [B.6].

Chapter IV concerns free will and determinism. Its purpose is to shed more light into the concepts that are involved in our thinking about both. It involves an argument against several influential compatibilist treatments of free will and consequently,

amounts to an extended response to the objection to the thesis' main argument, raised in Ch. III/B.2, according to which, the alternative possibilities that are required for being under the scope of normative reasons and obligations can be given a compatibilist account. In section A, various conceptions of free will are raised and it is argued that the best one understands it as the ontologically fundamental/irreducible specific power to actively do otherwise. The main rival of such a conception is the conditionalist compatibilist one which claims that an agent can do otherwise, even in a deterministic world, if a conditional statement is true of an agent with respect to the effect that 'if the agent intended to do otherwise, then the agent would do otherwise'. After an exposition of traditional criticisms against such an understanding of the 'ability to do otherwise' [A.2.1] and its contemporary reformed formulations [A.2.2], an argument is given against the entire project of conditionalist compatibilism [A.2.3]. According to this argument, this project is ultimately misconceived because it eventually loses sight of the situatedness of human beings. As situated beings, persons are always situated in circumstances having one form rather than another at a given time. A fortiori, asking: 'Could this person have done otherwise?' is elliptical for asking: 'Could this person – situated in circumstances C having form (F n) at time t- have done otherwise?' I argue that to answer: 'yes, because if the person were situated differently, then the person would have done differently' is to answer a question which does not treat the agent's situatedness as essential. It fails because its very rationale renders something which cannot be actual for any situated person –namely, to be in circumstances having a form other than the one they actually have for this person at each time- as the necessary condition for doing otherwise. If the necessary condition for doing otherwise cannot ever be actual for a situated agent, then a situated agent is never able to do otherwise.

This chapter's section B is concerned with defining determinism and arguing that it is ultimately incompatible with free will. It concerns the final touch needed to argue against compatibilist understandings of the kind of agency which grounds normative responsibility. After several remarks about the attempts to capture the

meaning of determinism [B.1.1, B.1.2], two versions of determinism are defined [B.1.3]. Robustly causal determinism is conceived of as the kind of determinism that crucially involves causal necessitation, whereas neo-Humean determinism appeals to the purely epistemic relation of 'entailment' between propositions or statements, instead of referring to any relation of causal necessitation. In this chapter's B.2 section it is ultimately held that free will is incompatible with both robustly causal determinism [B.2.1] and neo-Humean determinism [B.2.2].

Chapter I: Reason and Normativity

In this chapter, I develop a conception of rationality as a normative engagement in thought and physical action. According to this view, rationality is substantive when we think and act in light of reasons, and structural when we engage in the attempt to achieve coherence among our attitudes, irrespective of reasons for and against them. In section A, I explore the notions of 'reason', 'reasons' and 'rationality'. Reason as a faculty will be conceived of in a rationalist sense, as the faculty that allows us to grasp necessary facts and relations. Reasons will be conceived of as irreducibly normative 'favourers' which require of a person to have alternative possibilities with respect to action, whether mental or physical, in order to be under their scope. Rationality, in its substantive sense, will be conceived of as responsiveness to reasons with the qualification that reasons are not efficient causes of how rational agents respond to them. In section B, I defend this view of rationality against two challenges to its normativity. The first maintains that rationality is not essentially connected with normative reasons because it is solely concerned with the achievement of coherence among our attitudes. I claim that this challenge is vastly overstated and can only represent a threat to the normativity of reasoning, as well as the normativity of thought in general, if all instances of rational thinking are conceived of as instances of structurally rational thinking. Such a view would make a conspiracy theorist the paradigm of a rational thinker. Furthermore, assuming that knowledge requires epistemic reasons, to claim that we can never have any access to epistemic reasons would lead straightforwardly to skepticism. The second challenge extends moral error theory to the epistemic domain. I claim that this objection leads also to skepticism and that it is ultimately unmotivated because it rests on a very restricted version of physicalism. The claim that an error theory about the normative is unmotivated will reach its completion in the second chapter where a rationalist alternative will be developed and defended.

Section A: ***Reason, reasons and rationality***

A.1 ***Reason***

The notion of 'reason' is one of the most central concepts of philosophy. It can refer to a faculty, a justifying consideration, a cause which is cited as an explanation of a particular phenomenon, a mathematical relation, or even an immanent force operating in the universe, like the Stoic 'Logos'. Insofar as it is conceived of as a faculty of mind, two conceptions of reason have been historically dominant and their opposition is nowhere more evident than in their respective treatments of its metaphysical and epistemological reach. Rationalists conceive of reason as enabling us to gain knowledge of the necessity of some facts and relations in particular domains, such as facts and relations about the world, its logical, mathematical, moral aspects, and so on. Most empiricists –if not all- share a more deflated approach with respect to reason's reach. Rationalists are interested in employing reason to secure the possibility of knowing what couldn't be otherwise, whereas the latter are frequently skeptical of such claims. According to Brand Blanshard, reason as an ability "commonly denotes the faculty and function of grasping necessary connections." (2013[1962]: 25) On the empiricist side, David Hume holds that reason cannot give us knowledge of the world because there are no necessary connections between distinct existences for reason to grasp. He writes: "There is no object, which implies the existence of any other if we consider these objects in themselves." (1985[1739]: 135, Book I, part III, section.6) Apart from necessary connections, Hume rejects the possibility that reason enables us to grasp necessarily true facts, because: "The contrary of every matter of fact is still possible; because it can never imply a contradiction, and is so conceived by the mind with the same facility and distinctness, as if ever so conformable to reality." (2007[1748]: 25, part I, section IV) If reason is concerned with grasping matters of fact, whose contraries are still possible, then it cannot grasp any necessary fact. The

reason is that what counts as a 'fact' for Hume, is by definition contingent. Thus, Hume holds that reason in the rationalist sense has nothing to grasp and a fortiori cannot tell us anything about the world, in the way that pre-Kantian rationalists hold it to.[1]

One of the main themes of this thesis is that the debate between rationalism and empiricism in epistemology is not just historically interesting. Instead, it directly affects our contemporary discussions about philosophy of mind, epistemology and normative branches of philosophy in general. I will claim that a version of rationalism, particularly one which doesn't extend the principle of sufficient reason to the realm of action, is a viable option in epistemology with interesting ramifications for ontology and the discourse concerning normative reasons.

In contemporary analytic –and arguably continental- philosophy, the power of reason to tell us anything about the world is widely contested.[2] The consequences of such skepticism leave almost no branch of philosophy untouched. Perhaps the most significant ramification of such skepticism is that it leads to skepticism of the very autonomy of philosophy as a discipline. Philosophy's autonomy is challenged on the grounds that it does not possess a subject matter that is not in principle addressed by the methodologies of other disciplines; for instance, the methodology employed in the

[1] Hume does hold that among the objects of reasoning are 'matters of fact' and 'relations of ideas' (ibid: 25, part I, section IV). If we focus on that sentence alone, it would seem that Hume thinks that reason can indeed give us knowledge about matters of fact. The tension can be resolved once we bear in mind that the objects of reasoning that Hume has in mind are not objects of reasoning *in the rationalist sense*.

[2] This seems to me to be a dominant tendency which is shared both by Kantians who associate the position with uncritical dogmatism, logical empiricists who broadly equate apriority with necessity and analyticity (Carnap 1937; Quinton 1964) and radical empiricists. Due to limitations of space I will have to pick radical empiricists like W.V.O. Quine (1953, 1969) and Gilbert Harman (1986, 2001) as representing the main challengers to the rationalism I defend. The reasons I pick them is twofold: i) Motivationally speaking, Kantianism does not seem to me to be hostile to free will, and ii) after Saul Kripke's (1980) *Naming and Necessity*, the conception of apriority as identical with necessity and analyticity has been widely recognized as suspect.

natural sciences.[3] In order to secure the autonomy of the discipline, it suffices to restore reason to its proper position. That is why I will focus predominantly on how skepticism about reason affects our conception of the relation between reason and reasons.

One of the main implications of skepticism about reason consists in the fact that theoretical and practical reasoning become increasingly portrayed as distinct from normative reasons.[4] The rejection of pure reason leads many contemporary philosophers to treat reasons as facts that, if they are to exist, have to be grasped by our sensory faculties. This leads many thinkers to feel an increasing pressure to reduce their normativity into descriptive terms. Statements about value, rightness and wrongness must go. The very idea of irreducible normativity becomes suspect because it does not seem to feature well in a physicalist worldview. According to the view of reason and rationality I develop, knowledge, reason and normativity stand or fall together.

Several ways have been proposed to bridge the gap between rationality and normative reasons. Christine Korsgaard (1996, 2009a, 2009b) has proposed a constructivist/constitutivist response to the challenge, according to which, to accept – not merely to conform to- the binding force of the principles of theoretical and practical rationality constitutes personhood.[5] Korsgaard understands reason as "the capacity for normative self-government." (2009a: xi) To possess rational powers is to possess the powers to think and act in light of reasons. Those we recognize as possessing such powers, we also recognize as accountable to various sorts of considerations like rational, moral and prudential ones. Along such lines, to be rational is essentially connected with being a person, in the precise sense that physical objects aren't persons.

[3] Some influential alternatives are presented by logical empiricism and Wittgensteinian quietism. Nevertheless, their dominance seems to have completely faded.

[4] For instance, Harman who embraces radical empiricism, draws the distinction between rules of argument and rules of inference (1986: 3). He counts among them principles like Modus Ponens and the principles of deductive logic and conceives of them as purely formal principles of implication that are ultimately insufficient to drive belief revision. They do not give us reasons to embrace or reject what they support at each time. A criticism of this view, can be found in Robert Hanna (2007: Ch. 7)

[5] For a criticism of Korsgaard's constitutive response, see David Enoch (2006).

Another promising response to the normative question about rationality lies with conceiving of rationality as a kind of responsiveness to reasons. According to this strategy, to be rational, i.e. to conform to the principles of rationality, is always accompanied by reasons because it constitutes a response to them. Errol Lord writes: "According to the Reasons Responsiveness account, rationality consists in responding correctly to the normative reasons you possess, where the reasons you possess are the facts that count in favour of acts and attitudes that are within your ken." (2017: 1111)

Throughout the thesis, I will conceive of reason as an intellectual power whose main function is to grasp necessary facts and relations. Its scope, as well as the modal status –the necessity or contingency– of its objects will be discussed in chapter II. I will claim that it is implausible to hold that rationality and normative reasons come apart entirely and that efforts to link them in the epistemic realm that do not embrace a rationalist conception of reasoning are incomplete when it comes to accounting for how reason is relevant to the pursuit of knowledge.

A.2 ***Kinds of reasons***

The forms that reasons may assume are numerous and depend as much on the domain in which they are encountered as on our manner of engagement with them. Although we will be concerned with four ways to conceive of reasons, such ways are not by any means exhaustive. Insofar as reasons are conceived of in terms of the domain in which they are encountered, they can be epistemic, practical, evaluative, etc. Insofar as they are conceived of in terms of their comparative strength, they can be conceived of as defeasible or indefeasible, whether in a particular domain or across domains. Insofar as reasons are conceived of in terms of the source of their force, they can be understood as instrumental or categorical. Insofar as reasons are conceived of in light of how they relate to action, they can be conceived of as motivating,

explanatory or justifying. Let us unpack each of these distinct approaches in understanding reasons.

Reasons are always encountered in a domain. Epistemic reasons are reasons that are encountered in the epistemic domain. They concern what we are to think as being the case and how we are to engage in inquiry. They are distinct from practical, moral or prudential reasons in virtue of the fact that the latter are encountered in the practical domain and are concerned with what we are to do. Our predicament as beings endowed with rational powers is not a bed of roses. We are frequently presented with dilemmas that stem out of the potential inconclusiveness of reasons we have for either option we consider, or the seeming conclusiveness of incompatible reasons we encounter. The fact that not all reasons support the same thing, and consequently don't favour the same attitude, reveals the need for comparative treatments of such reasons.

Placed in the scale of reasons, not all reasons have the same strength[6] or weight. Some are stronger than others. Furthermore, their force is not always the same across all domains. Reasons can be weighed within a domain as well as across domains. An epistemic reason can be conclusive in the epistemic domain but reasons of a different sort may outweigh it in the kind of weighing of reasons that takes place when we deliberate about what we have an overall reason to do and we consider whether we should engage in further inquiry or do something else instead. Reasons that can be outweighed are called defeasible whereas those that cannot be outweighed are called indefeasible. Depending on the kind of deliberation in which their force is compared, reasons can be defeasible or indefeasible within their domain, as well as across domains. An epistemic reason may be indefeasible within the epistemic domain, but defeasible across domains. For instance, practical reasons may outweigh epistemically

[6] By the 'strength' of a reason, I understand not its causal power, but the degree to which it justifies the adoption of a particular attitude. It will be mentioned in section A.4.2 that the rational strength of a reason is independent of the causal force of mental states that many thinkers treat as reasons.

indefeasible reasons by making it imperative in some occasions to temporarily cease engagement in further inquiry.

Apart from their domain and force, reasons can be further understood in terms of the *sources* of their force. Various conceptions of the nature of their strength's origin shed light to various distinct conceptions of the sources of normativity. If the source of a reason's force lies with the agent's desires or some other aspect of the agent's psychology, then such a reason is instrumental. If not, the reason in question is categorical. If the force of a reason expires with the cessation of desire, then such a reason is an instrumental one. If the desire ceases to be and the reason remains, such a reason is categorical.

Their relation to action renders reasons motivating, explanatory or justifying. When an agent performs an action in order to achieve a goal, that goal may be called the agent's motivating reason. If the action is treated as an effect, then its cause can be held to be its explanatory reason. Furthermore, depending on whether reasons render an action justified or unjustified they can be conceived of as justifying reasons.

Thinking about reasons in this way reveals something important about the human predicament. We are simultaneously present in various distinct domains in which distinct aspects of reality manifest themselves. Our presence in each domain and our concern with what we encounter in it, is not to be subordinated to our presence in other domains and our respective concerns in them. Insofar as we engage in inquiry, the search for truth, we inhabit the epistemic domain. Each distinct element we may think of, can be seen as an epistemic reason for a thinker to infer that which it supports. Insofar as we engage in practical deliberation, we inhabit the practical domain. Each element we may think of, is seen in light of how it relates or would relate to the achievement of our ends and well-being or the ends and well-being of others, construed in a very broad sense, impartial to distinct ethical theories. If I am stranded in a remote island due to a shipwreck and I come across a settlement, the fact that there is a settlement there gives me a defeasible epistemic reason to think that other humans live

or lived in the island. Seen in light of how it relates to my ends and well-being, the same fact may give me a defeasible practical reason to hide and carefully observe whether it would be safe for me to introduce myself to them and ask for their help or offer mine to them. Suppose some people come back to the settlement holding someone captive who was on the same ship with me and they amputate him by cutting his left foot. It seems plausible to say that this shocking sight would give me a defeasible reason to infer that they are hostile to strangers. Nevertheless, if further evidence suggested that they did this horrible act to prevent a fatal infection, the reason to believe they are hostile to strangers may be outweighed. Perhaps more observation is required to make a wise decision. But sooner or later, a decision has to be made. In all cases, the way in which we engage with an aspect of the world is closely linked with what kind of a reason it is or we take it to be.

Throughout this project, I will focus on epistemic reasons and hold them as categorical within the domain of inquiry. Moral reasons will be discussed but for reasons of space, evaluative reasons will not be discussed. As mentioned in the preface, this thesis presupposes an irreducibly normative conception of reasons. I will argue for the view that although defeasible in most cases, their strength within the domain of inquiry is indefeasible in some cases, such as when we are given a reason to reject something as false when we encounter a contradiction. As categorical, such reasons do not depend on our desires for their existence. Attempts to reduce epistemic reasons to instrumental reasons seem to me to be driven by the pressure to subdue our presence in the epistemic domain to our presence in the practical realm, where the latter is conceived of as fundamentally concerned with the satisfaction of desires or other contingent elements of our psychology.[7] It seems that the instrumentalist conception of epistemic reasons conflates the non-existence of an epistemic reason with its being

[7] A relatively recent defense of the position that epistemic reasons are categorical is given by Thomas Kelly (2003, 2007) and a counterargument can be found in Adam Leite (2007).

outweighed by a reason of another kind, *while still remaining an epistemic reason*.[8] But due to considerations of limited space, this point will not be defended in this thesis.

A.3 ***Reasons and normativity***

The concept of 'a reason' is one of the central normative concepts. In the attempt to classify normative concepts and explore their interrelations, two positive strategies are adopted: normative monism and normative pluralism (Robertson 2009). Normative monists hold that there is one fundamental normative concept such as value, reason, obligation, etc. and any other normative concept is to be understood in its terms. Normative pluralists allow for more than one fundamental concepts, irreducible to each another. Reasons-fundamentalists (Scanlon 1998, 2014; Raz 1999, 2011; Skorupski 2010) follow normative monism and hold that statements about reasons are the fundamental normative statements, which form the ground of reduction for every other normative statement. Joseph Raz writes: "The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons." (1999: 67) For instance, statements about what one 'ought' to do, are reducible to statements about what one has decisive reason to do.[9] Non-reasons-fundamentalists may take reasons to

[8] For instance, to the extent that Harman's claim that we have no reason to clutter our minds with trivialities (1986: 12) amounts to the claim that whether we have epistemic reasons or not depends on our interests, it may be countered in the two following ways: first, not all instances of rationality are connected with reasons –e.g. structural/subjective rationality- and second, it is entirely plausible that there is an infinity of epistemic reasons, some of which are constantly outweighed by reasons of another kind due to our situatedness and our simultaneous presence in several domains. For instance, getting some rest may constantly outweigh the epistemic reasons that Harman has in mind, without that implying that they cease to be epistemic reasons. As such, it is not plausible to hold that they come and go in and out of existence, depending on the agent's desires or equivalent proposals.

[9] An account that treats a qualified notion of an obligation as the ground on which normative reasons are to be understood, can be found in John Broome (2013: ch. 4)

be either non-fundamental, or equally fundamental with other normative notions. In what follows, I will not make a claim about what the central normative concept is, or if there is one *a la* normative monism. I will mostly refer to reasons, instead of obligations, without implying that they are more fundamental than the notion of an obligation or vice versa. I will assume the existence of normative reasons and motivate the claim that such reasons can only be normative if the agent possesses them and is able to act in accordance with them or violate them.[10]

Grasping the notion of 'a reason' is integral to understanding rationality as well as 'reason' as a faculty. It can be understood as a *weakly* normative notion in the sense that it is associated with conditions of correctness. Inferences and actions can be in accordance with certain requirements, as well as not in accordance with them. Furthermore, reasons can be understood as *strongly* normative in the sense that they *justify* the responses they favour, as well as render *unjustified* the attitudes they count against. They do not just explain why we do something. They explain why it is right or wrong to do it. As Maria Alvarez writes: "I therefore take the idea that reasons can have normative force to involve the idea that reasons can be invoked to support claims about what it would be right for someone to do, believe, want, feel, etc." (2010: 9)

Normative reasons are justifying reasons *for* someone *to* do[11] something *in their light*,[12] whether in thought, evaluation or physical action. The agent's responsibility to act in light of such reasons is the agent's normative responsibility. A particular 'reason', as a justifying consideration or set of considerations, is characterized by the

[10] Since agents are situated beings, I will claim that such an ability to follow or violate requirements generated by reasons is possessed by a situated agent and requires the physical ability to do either, at that time and in those very circumstances.

[11] Although 'doing' is considered to be closely tied to physical action, I will use it throughout the thesis as characterizing a form of active engagement in any of the domains of existence we inhabit. So to think to evaluate and to act, seem to me to represent *doings* of a distinct sort. There is a prevalent idea which holds all our epistemic attitudes to be non-voluntary. The purpose of the thesis is by and large to deny this claim.

[12] What I mean by that is that a normative reason R for an agent S that justifies an action A is a reason for S to A *in light of* R; not blindly.

fact that it is addressed to a person, and on the basis of what it supports, favours an attitude. The recognition of such support constitutes the recognition of a relation between distinct elements.[13] For instance, the fact[14] that smoke is related to fire, gives one a defeasible epistemic reason to think that a fire burns yonder, in light of the sight of plumes of smoke rising in the sky. Such support can take many forms. A can necessitate B, without B necessitating A. In other cases, it might be that A necessitates B and vice versa. Furthermore, that which an epistemic reason justifies is different from what it supports. The first consists in an attitude whereas the latter consists in a fact or a set of facts. An epistemic reason supports facts and on the basis of such support, justifies our believing in them or inferring them, without necessarily believing in their truth. T. M. Scanlon (1998) writes: "'a reason for something' is 'a consideration that counts in favour of that thing.'" (1998: 17) A particular reason to believe that a state of affairs obtains is that which justifies a thinker in believing that this state of affairs obtains. In this case, the epistemic reason consists in a simple or complex element in the rational discourse we engage in, which serves as the ground upon which we apply rational principles in order to grasp what it supports and consequently, to guide our thinking in light of the attitude it favours.

The normativity of reasons is closely linked with the possibility of their violation. Absent such a possibility, we cannot maintain that we can be responsible for acting or not acting in accordance with this or that normative reason. Therefore, if an

[13] An existent relation is not the same thing as an epistemic relation; the epistemic relation consists in the identification of an existent relation as providing us with an epistemic reason to think what it supports. So, claims about what reasons favour as well as what is related with what else, seem to be inherently connected with claims about epistemic reasons. If there is a semantic circle, it is not a vicious one. It indicates that existent relations, their identification and reasons to think are all irreducible aspects of engaging in inquiry; in more general terms, of the relation between mind and world.

[14] I will adopt Derek Parfit's (2001, 2011) reluctance to identify reasons with facts and will follow his claim that "reasons are provided by facts" (2001: 17). To be fair, Parfit admits that the possibility still exists that such a distinction may not be tenable for the reasons-fundamentalist (ibid: 18).

agent cannot violate a demand constituted or generated by normative reasons, that agent cannot be under their scope; the range of their authority. Douglas Lavin writes:

> Now, a putative principle formulable as "Do A or don't do A" or "You must either do A or else not do A" isn't something which an agent can violate. Its logical form prevents us from describing or thinking anything at all that would count as a deviation from it: for this reason we want, I think, to say that its violation is logically impossible. And for this reason in turn we want to deny that an agent can be under it. (2004: 426)[15]

Since normative reasons are justifying reasons, lack of the possibility of their violation implies lack of the possibility of being unjustified in their light. Being justified where someone cannot be unjustified amounts to winning in a game in which one cannot lose.

Normative reasons are inextricably interconnected with the possibility of normative criticism. To say that an agent is unjustified[16] in doing something but cannot be criticized normatively for doing it, is to voice an absurdity. Correct normative criticism consists in judging fairly whether persons do the best they can with what they have in the circumstances they are situated in at a given time. Incorrect normative

[15] In this article, Lavin maintains that the possibility of error is essential to normative considerations, but claims that there are two ways to understand such a proposition: the logical interpretation and the imperatival one. The first one requires only the logical possibility of a mistake whereas the latter requires also the physical ability of the agent to do something that would violate the principle. He argues in favour of the first, whereas I will be arguing in favour of the second. The reason why I will not deal with Lavin's point more directly in what follows is that he takes the entailment of libertarian free will by the imperatival interpretation of the error constraint as the grounds for rejecting it. The thesis can be seen as a defense of that very account by showing why Lavin's rejection of it is premature.

[16] In this context, lack of positive justification differs from negative justification. There is a difference between that which is permissible by reasons and that which is forbidden by them. Being unjustified in the sense of not having justification for doing something *while not violating any available reasons* differs from being unjustified in the sense of having reasons not to do something and doing it. It is the latter sense that is involved in this claim.

criticism consists in expecting too much or too little from the agent. In the first case, criticism is unfair because it holds the agent accountable to normative reasons without being sensitive to the agent's abilities. In the second case, criticism is unfair because it doesn't pay tribute to the agent's potential. By and large, normative criticism consists in an interpretive process and requires us to engage in charitable attempts in making a whole nexus of judgments about the case of the person we judge.[17]

The normativity of reasons and the criticizability they entail for persons gives rise to the question of their scope. Robertson writes: "The scope of reasons concerns the agents to whom reasons apply –those agents to whom we correctly attribute reasons or about whom we make true reason claims, either reasons in general or specific reasons." (2009: 11) The scope of normative reasons concerns the conditions that have to obtain in order for a person to be actually subject to their force or authority.[18] If these conditions do not obtain, the agent is not normatively responsible for meeting them or not, and as such they cannot be reasons that justify an agent's doing something; in short, they cannot be normative reasons for that person because they do not apply to that person. For instance, it seems commonly accepted that there is no normative reason to do the impossible (Streumer 2007; Haji 2012; Wedgwood 2013, 2017).[19] If it is

[17] As such, it requires what Adam Smith refers to as the traits of a virtuous spectator, upon whose attempt to understand the sentiments of another person, rest "The soft, the gentle, the amiable virtues, the virtuous of candid condescension and indulgent humanity…" (2009[1759]: 30, Part I, Ch.V)

[18] Whether one is under the scope of normative reasons at a particular instance differs from whether one takes oneself to be under the scope of normative reasons at that very instance. This is evident if we take into account phenomena in which some people feel an overwhelming amount of regret and shame, as well as blame themselves for something they could not have prevented. In this case, understanding the scope of normative reasons seems a salient feature of understanding humans and how we respond to certain situations. Understanding of our own limitations and contingencies might have a healing effect.

[19] Some rejections of this widely held principle can be found in Sinnott-Armstrong (1984) and recently, Graham (2011). A response to them, as well as a wider defense of it, can be found in Vranas (2007, 2018). This project explicitly presupposes the OIC principle.

impossible for an agent to do something, the agent is not normatively responsible for not doing the impossible. We can say that the agent's inability to act in a way or inability to avoid acting in another way, provide us with reasons to maintain that the agent's responsibility in both cases, is not one of a normative kind. It may be merely causal, but it cannot be normative. We cannot be unjustified in not doing what couldn't be done. It seems plausible to hold that there are no normative reasons to do the impossible. Not doing something is not straightforwardly the same as failing to do it *while being under the scope of normative reasons that render it justified*.

Considerations about an agent's abilities and knowledge are considered as potentially responsibility-absolving. It is intuitively plausible that human beings cannot be blamed for the entirety of what happens in the domains they inhabit. Although they might serve as the needle that points north in our evaluative compass, omniscience, omnipotence and omnibenevolence do not seem to be realizable for beings like us; we cannot expect human beings to be Gods. Due to the severe limitations that are presented to us by our situatedness, many thinkers are prepared to admit that statements about what should ideally happen cannot generate normative reasons, unconditionally. As such, they maintain that agents cannot be normatively responsible for meeting these demands. Do such statements about reasons or obligations hold whether we know of them or not? Do they hold independently of whether we can realize them?

Treatments of the scope of such reasons and obligations fall under two broad categories.[20] This marks an important distinction between the objectivist and the perspectivist conception of the scope of normative reasons and obligations. Advocates of the former view, like Moore (1969[1912]), Thomson (2008) and Graham (2010) hold that what we ought to do is independent of facts about our perspective. We *ought* to do something or refrain from doing it, independently of whether we know of it or not. They claim that facts about what we ought to do or have reason to do and facts

[20] For reasons of space I will not analyze the prospective view of moral responsibility, associated with Michael Zimmerman (2008).

about blameworthiness and praiseworthiness can be distinct and maintain also that the former do not entail the latter (Lord, 2017: 1136). Advocates of the latter view, like Shafer-Landau (2009), Kiesewetter (2011, 2017) and Wedgwood (2017) hold that what we have reason to do and what we should do is essentially dependent on the agent's perspective, such as our beliefs or evidence for what we have reasons to do.[21] As such, there can be no distinction between facts about reasons and obligations and facts about blameworthiness and praiseworthiness. In the next section [A.4] I will propose a view that blends insights from both sides. Against objectivism, I will hold that facts about the agent are necessary conditions for something to be a reason or an obligation for an agent. Against perspectivism, I will hold that facts about the agent are not sufficient conditions for making something obligatory, or a normative reason for the agent. But before delving into that, let us examine the dialectic between objectivism and perspectivism.

Objectivist intuitions about moral obligations and reasons are radically distinct from perspectivist ones. To highlight such a divergence, I will modify an influential thought-experiment, expressed by Frank Jackson who writes:

> Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B, and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of drugs B and C will completely cure the this skin condition; the other though will kill the patient, and there is no way that she can tell which of the two is

[21] Most, if not all perspectivist views about what the agent ought to do, fall under two main categories: belief-based perspectivism and evidence-based perspectivism (Kiesewetter 2011). The former generates obligations from what the agent believes, whereas the latter understands obligations to be generated by the evidence an agent possesses or by beliefs *for which the agent has evidence*.

> the perfect cure and which is the killer drug. What should Jill do? (1991: 462-463)

For purposes of simplicity, suppose that Jill had access only to drugs B and C; in our thought experiment drug A is not an option. And suppose also that she has to decide quickly and that all the records of medical history indicate that she should give drug B. Incidentally, John suffers from a weird condition, undiscovered yet by medical practitioners. And because of that condition, drug B would be lethal to John, whereas drug C, contrary to hitherto medical knowledge, would save him. Objectivists about obligation argue that Jill should give John, drug C. According to them, what we ought to do is independent of what we know and is determined by facts, whether we know of them or not. Perspectivists about obligations argue that Jill *should* give drug B. For them what we ought to do is based entirely on facts an agent is aware of, such as beliefs or evidence about what is the best thing to do. We can only be judged in terms of standards that we can know of. It might be good for her and John not to give drug B, but what would that show for her as a person, if she gave John the very drug she thought would kill him?

The standard objectivist response to this would be to distinguish between facts about what we ought to do and facts about blameworthiness and praiseworthiness. No one denies that statements of the latter form say something important about us. The question seems to be whether they can form the ground of *normative* criticism. A perspectivist view about reasons and obligations would not draw the same conclusion. Since normative standards should provide us with normative guidance, we cannot be guided by that which we do not, and in many cases, cannot know. To engage in normative criticism of an agent is to judge whether in particular cases, they did the best they could in light of the circumstances they were situated in. But what is actually good and what is thought to be good, may diverge. It is one thing to engage in discourse about evaluative facts and quite another to engage in normative criticism of an agent. The perspectivist may not deny the possibility of a link between what is good and what

is right, but they may deny that evaluative claims generate claims about obligations and normative reasons, unconditionally. And even if the perspectivist concedes to the objectivist that there might be such ought-claims, grounded on what we don't know, they can still maintain that they are not normative in the responsibility-implying sense, but that they just express ideal or desirable states of affairs for which the agent is not normatively responsible; i.e. not normatively criticisable.

A recent account of such a distinction between two uses of ought-claims is given by Ralph Wedgwood (2007: ch. 4, 5, and 2009). Wedgwood claims that there is a distinctive sense of 'ought' which he calls the 'ought of general desirability', which expresses an ideal state of affairs without implying that there is any responsibility which burdens the agent's shoulders if the ideal state of affairs that is expressed by it isn't realized. An example he gives is: "We ought to hire that man, David Lewis." (2007: 89) Although that would be generally desirable, it doesn't imply responsibility for us if we do not hire David Lewis, because we cannot do so. Therefore, this is not a genuinely normative 'ought'. It might be ideal, but its lack of realization would not imply blameworthiness. Violation of the 'ought of general desirability' doesn't imply the strong criticizability that the violation of a normative reason would. As such, it seems absurd to call it a violation, rather than a lack of realization. The privation of an ideal state of affairs straightforwardly implies neither the existence of a bad state of affairs nor the existence of a mistaken action on the part of a person that is normatively responsible for it.

In light of such considerations, we may distinguish between normative and non-normative uses of the notion. Normative 'oughts' imply normative responsibility in meeting them. Non-normative uses of 'ought' do not imply such responsibility. But this notion of responsibility doesn't presuppose only the ability to meet what is required of us; it also requires us to know what we are doing. It is possible for me not to move my right hand but I have to do it for the right reason, if not moving it is to count as an action for which I am normatively responsible. Normative standards are supposed to

provide us with guidance about how we should engage in theoretical and practical reasoning. And to guide us they have to guide us *in light of* themselves, as opposed to blindly. As Raz writes: "if some people cannot know of a fact it does not constitute a reason for them, even though other people can know about it." (2011: 110)

What we do not know cannot provide us with guidance. Provided one is not normative responsible for not knowing, being subjected to normative criticism in terms of what one does not know is for that person to be judged unfairly; i.e. not to be judged in light of whether she or he did the best they could with what they had. It is unfair because it either requires us to have knowledge that we do not have or it requires us to do what is impossible for us to do at that time. Therefore, reasons that are not known by us cannot be normative reasons for us to do what they demand. We cannot be responsible for meeting them or not because we cannot be guided by them. We cannot think and act in light of reasons that we do not have access to. If we end up doing what we have such a reason to do, it is a matter of luck. So, if we can allow that we have an 'objective' reason not to move our right hand, and therefore, we 'objectively ought' to abstain from moving it, we are not responsible in a normative sense if we end up moving it because we cannot be guided by that which we do not know. In this sense, objectivist reasons and statements about objectivist reasons do not automatically translate to normative reasons and statements about normative reasons. To answer when they do, is the topic of A.4.

A.4 ***The double threshold theory of normative responsibility***.

In this section, I introduce the double threshold theory of normative responsibility. It is concerned with the standpoint we must occupy in order to be under the scope of normative reasons, and consequently, the subject of fair normative criticism; the standpoint in which evaluative claims may generate normative claims. Its upshot is that if we accept the Kantian 'Ought implies can' principle (OIC), we will

also have to embrace the principle of alternative possibilities (PAP) [A.4.1], not only in the case of practical reasoning but in theoretical reasoning as well [A.4.2].

The double[22] threshold theory of normative responsibility is structured around the two conditions that are traditionally thought of as determining whether one is morally responsible or not. These conditions concern the awareness of reasons as well as the ability to meet rational demands in their light; not blindly. Based on several combinations of the actuality or lack of each condition with the actuality or lack of the other, for a situated person, we can identify four contingencies, three standpoints, three thresholds –only two of which can be crossed by an agent at a time- and ultimately, one normative standpoint.

Emphasis on awareness of reasons and the ability to meet the demand constituted or generated by them, leads us to recognize four contingencies.[23] First, a person can be unaware of reasons and be unable to meet their demand even if she became aware of them. Second, a person can be unaware of reasons but possess the ability to meet their demands if she became aware of them. Third, a person can be aware of reasons and be able to meet their demands. Fourth, a person can be aware of reasons but be unable to meet their demands. If we symbolize awareness of reasons with 'K' and the ability to act in their light with 'A', we can have the following matrix of the four contingencies:

K-, A- (standpoint of ignorance)	K+, A+ (normative standpoint)
K-, A+ (standpoint of ignorance)	K+, A- (evaluative standpoint)

[22] I call this the double-threshold theory of normative responsibility, instead of the triple-threshold theory. The reason is that although it recognizes three thresholds, only two of them can be crossed by a situated agent at each time.

[23] Introducing a further distinction between the ability to try to act in light of a reason and the ability to meet the demand constituted or generated by that reason would make the exposition of this point, unnecessarily complex.

These four contingencies demarcate three distinct standpoints. These are the standpoint of ignorance, the normative standpoint and the evaluative standpoint. The first two contingencies represent the standpoint of ignorance. Irrespective of whether the agent has the ability to meet the demands that these reasons would generate if they were normative, the agent is unaware of them. Questions can be raised as to whether the person in this state is normatively responsible for being in that state. For instance, someone might not know the rights and responsibilities of her citizenship. Was that person in a position to be aware of them or not? If not, then being in that standpoint might not be normatively criticisable. If that person was in a position to know but did nothing about it, then this person has derived normative responsibility with respect to not knowing the rights and responsibilities of her citizenship. The third contingency represents the normative standpoint. If a person is aware of normative reasons and is also able to meet their demands, then at the absence of stronger reasons that would favour another action, this person is normatively responsible to meet them. When an agent occupies the normative standpoint, it is fair to criticize her in a normative sense. She is under the scope of normative reasons. The fourth represents the evaluative standpoint. This is the standpoint one occupies when one is aware of certain reasons but is unable to meet the demands they constitute or would generate if they were directed to an agent who occupied the normative standpoint. Normative criticism of a person who occupies the evaluative standpoint with respect to a reason, in light of that reason is inherently unfair. It violates OIC in that it is not sensitive to the fact that the person in question is unable to act in the way this reason requires. Again, questions arise as to whether this person occupies this standpoint on purpose, such as when people voluntarily incapacitate themselves with respect to discharging an obligation, in order to be freed from it. In the same way that there can be derivative responsibility for occupying the standpoint of ignorance, there can be derivative responsibility for occupying the evaluative standpoint. But these cases do not exhaust the entirety of cases in either. Although some who occupy the standpoint of ignorance or the evaluative standpoint with respect to a reason, are normatively responsible for it in a

derivative way, it is not plausible to hold that everyone who occupies one of these standpoints with respect to such reasons is normatively responsible for it.

These three standpoints indicate three thresholds, only two of which can be crossed at a time. Since we have three standpoints and each person always has to be in one standpoint with respect to a reason, each person who occupies one of the three standpoints with respect to a reason, can either remain in that standpoint or cross either of the two thresholds that link that standpoint with the other two. In transitioning to the normative standpoint from the standpoint of ignorance or the evaluative standpoint, a person crosses the normative threshold. Within this threshold, the person is a normatively responsible agent who can be properly held accountable in light of how and whether she meets the demands generated or constituted by the relevant reasons. It is in virtue of her occupying the normative standpoint with respect to a reason, that this reason can be called a normative one and statements about these reasons can be called statements about normative reasons. In transitioning to the evaluative standpoint from the standpoint of ignorance or the normative standpoint, one crosses the evaluative threshold. In transitioning to the standpoint of ignorance with respect to a reason from the normative or evaluative standpoint with respect to that very reason, one crosses the threshold of *lethe* or oblivion. Although the thresholds are three, only two can be crossed by an agent at each moment. The reason is that these thresholds connect standpoints and a person has to occupy one of them with respect to a reason at each moment. Such itineraries can be represented in the following way:

From (standpoint of ignorance or the evaluative standpoint) to the normative standpoint	Normative threshold
From (standpoint of ignorance or the normative standpoint) to the evaluative standpoint	Evaluative threshold

From (normative standpoint or evaluative standpoint) to the standpoint of ignorance	Threshold of oblivion/*lethe*.

The bottom line is this. Statements about reasons or about what an agent ought to do, do not generate normative demands unconditionally. They do not necessarily represent normative reasons or obligations. They can only do so for an agent who occupies the normative standpoint with respect to them. Furthermore, if an agent does not occupy the normative standpoint with respect to a reason, the question arises as to whether she is normatively responsible for not occupying it with respect to that reason. For her to be normatively responsible for not occupying the normative standpoint with respect to that very reason, it would have to be the case that she is responsible in a derivative way for it. She *should* have been aware of that reason or she should not have purposefully incapacitated herself with respect to meeting the demand that is constituted or generated by that reason, if she were aware of it. In one way or another, normative responsibility can only be grounded in something that one did or failed to do, while occupying the normative standpoint.

The double threshold theory of normative responsibility allows us to combine insights from both objectivists and perspectivists about moral responsibility. The facts that the objectivists about normative reasons treat as reasons can be plausibly seen as evaluative facts that hold independently of whether we know them or not.[24] Assuming the existence of creatures that possess a well-being, such facts about their well-being are independent of their knowledge of it. Independently of what I know right now, there are facts of the matter as to whether certain states of affairs will be good for me or not. Such facts cannot form the ground of fair normative criticism, unconditionally. To be fair in criticism, we have to have in mind what is the ideal of a rational agent,

[24] I don't see any reason why a supporter of objectivism about what an agent ought to do, would disagree with this. This seems to me to be the distinction that Thomson (1992) draws between what is good for an agent and what is good from an agent's point of view.

realistically speaking; an ideal that is achievable in the right way; a way that is informed by the person's understanding of the situation. In general, the ideal rational agent is the agent that does the best she can with what she has, in every case she is situated in. To engage in normative criticism of an agent in light of the 'ought of general desirability' is not to criticize the agent in light of doing the best she can with what she has, but instead to criticize her in terms of what we would want her to be able to do, despite her inability to do so. In that case, the error lies not with the person who cannot act in light of it but with the person who holds the agent accountable to normative criticism in light of the ideally desirable state they cannot bring to fruition. And since there are cases where these two roles are assumed by the same person, it is not in not doing the impossible that error lies but in thinking that you can do and should do, in a normative sense, what you are ultimately unable to do. To be fair, normative criticism has to be criticism of a person in light of reasons or obligations, with respect to which she occupied or occupies the normative standpoint.

The objectivist might object that such a theory downplays the importance of certain statements that express unattainable, yet ideal states of affairs. But this doesn't follow because such claims can still maintain a central position in normative inquiry. Despite the fact that they are unattainable, they can still maintain a salient position in the structure of our understanding of the source of obligations and normative reasons. Expressing this idea in the realm of political philosophy, L.T. Hobhouse writes: "At the other extreme, true harmony is an ideal which it is perhaps beyond the power of man to realize, but which serves to indicate the line of advance." (1911: 59) Of course, this predicament may feature in normative inquiry in various ways. For instance, we could either claim that social harmony is the ultimate social good or we could also say that it is one among many values; or that it isn't valuable at all. This seems to show that we can reject the idea that we have normative reasons to do the impossible while maintaining the intuition that such statements frequently indicate something important. In the same way, statements that claim that we have normative reasons to do something

that is physically impossible for us to do, cannot be the grounds of normative criticism but they can function as a sort of normative compass which shows us what to strive for *to the extent that we can*. Their importance is not downplayed at all. That perfection might be unattainable for us, does not imply that we should not care about it. This is another way of saying that in these cases, we occupy the normative standpoint as evaluators, instead of as practical agents, with respect to something in whose respect we occupy, as practical agents, the evaluative standpoint.

The upshot of this discussion consists in the claim that to be under the scope of normative reasons, a person has to occupy the normative standpoint with respect to them. And if one doesn't, certain questions are raised as to whether one occupied the normative standpoint with respect to not being outside it, in that particular case. To occupy the normative standpoint with respect to a reason, one has to be aware of it, be able to act in light of it, as well as be able to meet the demand constituted or generated by it. To act in light of normative reasons is not to act blindly. Normative responsibility is grounded on the obtaining of the conditions of normative criticism. To be fair in judging someone normatively, that person has to have awareness of a normative reason, the ability to understand what it supports, what it favours, as well as the ability to achieve what that reason favours, in light of that reason; not blindly. Still, much hangs on how we understand what it means to be guided by a reason. And to shed more light to such a concept, it is time to discuss the abilities that are relevant in grounding responsibility of a normative kind [A.4.1] as well as those involved in theoretical rationality; the power to think and act in light of reasons [A.4.2].

A.4.1 ***OIC and PAP***

Hitherto, our theory of normative responsibility is silent on the topic of an agent's alternative possibilities. In this subsection, I argue that without alternative possibilities with respect to an action that the agent has reason to perform, an agent cannot be under the range of that reason's authority; i.e. under its scope. Harry

Frankfurt (1969, 1988) has convinced many thinkers to hold that the OIC principle does not entail PAP, and that one can be responsible for discharging a moral obligation even if she or he lack the ability to do something else. To be morally responsible for doing A does not require of an agent the ability to do anything other than A. An influential response to Frankfurt's point builds on the idea that OIC does in fact entail PAP, at least with respect to blameworthiness (Widerker 1991; Copp 1997, 2008; Schnall 2001). I will take another route in defending PAP and claim that the question of entailment becomes a secondary one because without PAP, OIC loses its prescriptivity.

As maintained above, [A.3-A.4] statements about what ought to be the case can be used in ways that imply normative responsibility as well as ways that do not. Furthermore, the discussion revolving around the connection between OIC and PAP concerns also the notion of ability that is relevant in grounding the claim that an agent is under the scope of particular normative reasons and obligations. Does lack of alternatives in action render it inappropriate to hold an agent as being under the scope of normative considerations or not? Those who think that it is not inappropriate to do so, claim that as far as abilities are concerned, a 'weak' notion of ability is all that is needed for an agent to be normatively responsible. Those who think otherwise, hold that a 'strong' notion of abilities is necessary to ground claims of such responsibility. They hold that only a 'strong' kind of ability –i.e. the ability to do otherwise- can render it appropriate to hold a person as being normatively responsible.

When we claim that 'ought' implies 'can' we can mean many things. If it is the notion of the 'ought' of general desirability that we have in mind, then it is just a descriptive claim about conditions that would have to obtain in order for that state of affairs to be generally desirable; a most probably false claim. It does not imply that anyone has any obligation or reason to realize it. In another sense, we mean to say something that is essentially related to prescriptive considerations. The responsibility that we are thinking of ascribing to the agent is of a normative kind; not a merely causal

responsibility which can even be exemplified by objects. In that sense, the 'ought' of the OIC principle has to be a prescriptive 'ought'. A prescriptive 'ought' represents a normative consideration that implies normative responsibility in meeting it. A non-prescriptive 'ought' is an 'ought' whose lack of realization does not imply responsibility of a normative sort; it is just like Wedgwood's notion of an 'ought of general desirability'.

The notions that are involved in OIC admit of various interpretations. Thus, it is prudent to unpack the possible combinations of their meanings. The relevant constituent notions are the notions of 'ought' and 'can'. 'Ought' may either be understood as prescriptive or as non-prescriptive, whereas 'can' may be understood as either strong or weak. The possible combinations of their constituent notions lead us to the following four interpretations of the OIC principle:

OIC (i1): prescriptive 'ought' implies 'strong can'
OIC (i2): prescriptive 'ought' implies 'weak can'
OIC (i3): non-prescriptive 'ought' implies 'strong can'
OIC (i4): non-prescriptive 'ought' implies 'weak can'

The first position –OIC (i1) - is the position which claims that to be under the scope of a normative obligation to do A, an agent must be able to do A or not-A. At this point, I will introduce the symbol 'C (F n) at t' to stand for an agent being situated in circumstances having form n rather than another form at time t.[25] Since human beings are always situated in specific circumstances, which have a specific form rather than another at a given moment, OIC (i1) entails that an agent in circumstances C having form (F n) at a time t, can be under a moral obligation to do A at t+1, if in C (F n) at t, she is able to do either A or not-A at t+1.

[25] Considerations about the agent's situatedness will be a theme throughout the thesis, but will be discussed more extensively in Ch. III/B.2 and Ch. IV/A.2.3.

The second interpretation –OIC (i2) - represents the position that many compatibilists and semi-compatibilists embrace. According to OIC (i2), in order to be under the scope of a normative obligation to do A the agent needs only the ability to do A; she needs no alternatives accessible to her. To be under the scope of an obligation to perform action A at t+1, in C (F n) at t, she must be able in C (F n) at t, to A at t+1. No ability to refrain from A or doing something other than A is needed.[26]

Since our discussion about OIC and PAP concerns the conditions under which an agent is under the scope of a prescriptive obligation or reason, we can safely eschew OIC (i3) and OIC (i4). They do not refer to prescriptive obligations. So we have to focus at OIC (i1) and OIC (i2). I will provide an argument for the claim that OIC (i2) is a confused concept. It cannot really be a prescriptive 'ought' that is involved in its formulation.

A prescriptive consideration is one that can be met or violated. To be met or violated by an agent, that agent has to be under the range of its authority; its scope. One cannot violate an obligation to do A at t+1, if one is not already, say at t, under its scope. The problem with advocates of OIC (i2) is that they ultimately conceive of 'being under the scope of an obligation' as identical with 'being in conditions in which one *will* proceed to discharge that obligation'. Thus, they leave no room for the possibility of that obligation's violation. What we end up with is a conception of OIC which implies that obligations can only be met, but never be violated. We can win in a game that we cannot lose in.

To see this more clearly, let us gather all our clues. A prescriptive obligation is one that is addressed towards a situated agent; an agent who inhabits circumstances that have a particular form rather than another at a given time. A prescriptive obligation demands of an agent in C (F n) at t, to perform an action A or to abstain from

[26] I write this because not-A is ambiguous between an omission with respect to A and a description of an action contrary to A which is not an omission. This affects definitions of the ability to do otherwise as the ability to do something or refrain from it. This usual definition is blurry.

performing it at t+1. Assume an agent S in world w, in circumstances C having form (F n) at time t. To do justice to the point that the agent lacks alternatives, assume w to be a deterministic world. Take O to be a candidate prescriptive obligation to A at t+1. Given C with form (F n) at t, the agent at t+1 will either A or not-A. In this case, not-A stands for a description of states of affairs that do not include A. If the agent A-s at t+1, then the agent in C (F n) at t is under the scope of O; because S in w in C (F n) at t can –in the weak sense of lacking alternatives- perform A at t+1. If the agent does not A at t+1, then S was not under the scope of O in C (F n) at t, to begin with; it follows from OIC and the assumption that w is a deterministic world, that if the agent did not A at t+1, then in C (F n) at t –in the exact same circumstances- the agent couldn't proceed to A at t+1. To violate an obligation, an agent has to be under its scope. Therefore, the agent was not under its scope at t. Without PAP, being under the scope of an obligation is eventually defined as being in conditions in which one *will* proceed to discharge that obligation. Therefore, without PAP, OIC cannot be really prescriptive because the 'ought' it refers to cannot be violated. Our argument can be formalized in the following way:

i) For an obligation to be prescriptive, it has to be possible to violate it.

ii) To violate a prescriptive obligation, the agent towards which it is addressed must be under its scope.

iii) Advocates of OIC (i2) are committed to treating the conditions under which one is under the scope of an obligation as identical with being in conditions in which one *will* discharge that obligation.

iv) Therefore, advocates of OIC (i2) do not leave room for the possibility of the violation of a prescriptive obligation; one either discharges it or is not under its scope.

v) From i) and iv) we get that the 'ought' involved in OIC (i2) cannot be prescriptive.

vi) The minimum necessary move to secure room for the possibility of violation of a prescriptive obligation while being under its scope, is to reject the equation described above in premise iii). Such a rejection requires at least, the recognition of the agent's alternative possibilities as a necessary condition for being under the scope of a normative obligation; PAP.

vii) Thus, without PAP, OIC cannot involve a prescriptive 'ought'.

The upshot is that normative reasons, in virtue of the fact that they render actions – whether mental or physical- justified or unjustified, require alternative possibilities accessible to the agent. Perhaps this argument seems more plausible in the case of practical reasons. It is precisely for this reason that we should now proceed into an examination of theoretical rationality.

A.4.2 ***Rationality as reasons-responsiveness***

In this subsection, I will argue that theoretical rationality consists in a particular kind of responsiveness to reasons. The distinction between substantive and structural rationality will be discussed in section B.2.2. In contrast to causal conceptions of responsiveness to reasons, it will be maintained that normative reasons cannot be conceived of as efficient causes of reasoning. Those who aim to understand responsiveness to reasons in terms of a thought process being efficiently caused by normative reasons fall into a trap that is structurally similar with the one that those who embrace OIC fall into when they deny PAP's importance in thinking about when an agent is morally responsible [A.4.1]. The solution that is proposed is the same. The discussion of this section might seem premature. The reason is that it doesn't discuss in detail several contemporary endeavours that aim to develop a naturalistic theory of content and consequently, misrepresentation (Dretske 1981; Millikan 1984; Fodor 1990). To discuss them here would be to divert significantly from the purpose and

nature of this project, which involves a rationalist theory of the mind, argued for in Chapter II. If the rationalistic theory developed there is correct, then no naturalistic account of thought will be tenable.

The complexity of our mental lives leads us to develop categorizations of various mental events, states and processes. A fortiori, inquiring about our very own rationality requires us to understand how we think about the criteria that a process has to satisfy in order for us to count it as a process of reasoning. If the criteria that a mental process has to satisfy in order to count as a process of reasoning consist in whether or not that mental process is efficiently caused by reasons, conceived of as mental states, then at pains of dissociating the rational force of reasons from their causal force, a process will either count as rational or as a-rational. Ultimately, those in favour of causal conceptions of rationality leave no room for incorrect reasoning; no error is possible in the case of reasoning.[27]

It is commonly accepted that there is a distinction between conceptions of rationality that treat it as a power and conceptions that treat it as a success term. As a power, theoretical rationality consists in one's ability to think in light of reasons. In exercising one's power to engage in theoretical reasoning, one may be justified or unjustified in thinking something to be the case or in engaging in inquiry in particular ways. Such rationality is contrasted with a-rationality; the lack of rational powers.[28] As

[27] Similar arguments can be found in various philosophers who embrace a broadly Wittgensteinian framework of thinking about reasoning and mental representation in general. In a relatively recent formulation of an equivalent objection, Julia Tanney claims that "the explanatory connection between reason (what is represented) and action (representation) cannot be too close – reason cannot necessitate or determine action – or irrationality (or *picking* as opposed to *choosing*) will be impossible." (1995: 122) In the free will literature, a similar argument is given by E.J. Lowe (2008: Ch. 9)

[28] Having rational powers does not imply that all aspects of an agent's life can be rationally assessed. A person possessing rational powers may engage in a-rational activities, or undergo certain processes like the production of anti-bodies or digestion which have nothing to do with rational powers. On the contrary, lacking rational powers does imply that all aspects of one's life are a-rational. Therefore, the possibility of irrationality requires the possession of rational powers.

a success term, rationality concerns an agent's *correct* reasoning. Such rationality requires the agent's possession of rational powers and is contrasted with irrationality; the incorrect exercise of reasoning. As Donald Davidson writes: "For the irrational is not merely the non-rational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason." (2004: 169)

This distinction allows us to disambiguate statements about the criteria that a process has to satisfy in order for it to count as a rational process. Unless we are aware of it, it is not obvious whether the relevant criteria we refer to are criteria that a process has to satisfy in order to count as a process of reasoning, as opposed to being an arational process, or as a process of *correct* reasoning as opposed to a process of *incorrect* reasoning. Ultimately, those who conceive of rationality in strictly causal terms, cannot make room for incorrect reasoning. A mental process counts as a process of correct reasoning or as an entirely a-rational process.

To shed more light on this, it is important to lay out the main motivation behind a causal theory of rationality. Let us suppose that a mariner in the ancient times sailed to the unknown in search of hitherto undiscovered lands. On a long piece of scroll, he inscribed certain sentences. Suppose such sentences represent his thoughts at the time. During a storm, he put pieces of that scroll in a bottle that was lost at sea. At some point in time the bottle was picked up by other mariners, lost in the ocean. In opening it and reading it, they find out a list of thoughts. For ease of exposition, suppose each inscription corresponds to a thought and such thoughts are registered in the exact chronological order in which the ancient mariner entertained them. The mariners who read the scroll try to interpret it. How are they to understand from this list whether the mariner is rational?

A first reaction might be to look at the rational relations between the contents of thought. It seems fairly obvious that just the rational relations of the contents of thoughts cannot settle the matter. Rational interconnections of the contents of one's

thoughts are a necessary condition for rationality but not a sufficient one. Among the contents of the mariner's thoughts are the following:

T (n): The Indian Ocean is the largest ocean

T (n+1): The Indian Ocean is not the largest ocean

What can we infer from this? Can we say that the mariner is inconsistent? We cannot for the following reasons: i) the fact that thoughts are entertained doesn't imply a thinker's commitment to their truth, and ii) the fact that inconsistent thoughts are entertained doesn't imply simultaneous commitment to their truth. So, to criticize a person as irrational we cannot rest satisfied with the rational relations of contents *alone* as a criterion for whether a person engages in reasoning or not. Inconsistent contents of thoughts don't make one necessarily irrational. It seems that to find out whether someone is rational or irrational, we must also know how these thoughts feature in one's life. Such rational relations are a necessary condition for rationality, but they are far from sufficient. And just in case one wondered, they can never be sufficient, despite not being necessary.

Advocates of causal conceptions of rationality hold that appeals to causation can finally settle the matter. Causation by contentful mental states can account for the missing piece of the puzzle which concerns how thoughts feature in one's mental life. To engage in reasoning is to engage in a process in which one's contentful mental states are efficiently caused by other contentful mental states, whose contents stand in rational relations with the content of the mental state which is the effect of that causal process. As Wedgwood writes: "Reasoning is a causal process, in which one mental event (say, one's accepting the conclusion of a certain argument) is caused by an antecedent mental event (say, one's considering the premises of an argument)." [29]

[29] In this paper, Wedgwood is primarily concerned with the problem of deviant causal chains.

(2006: 660) If we follow this as the spinal cord of rationality, then we can claim that rather than merely possessing contents which may happen to stand in rational relations, our thoughts can be rational, if they are related in the way we specify above. Causation is supposed to provide the missing element we have been searching for. In reading the ancient scroll, we find the following three registrations:

MS1: All whales are mammals

MS2: Orcas are whales

MS3: Orcas are mammals

According to our new addition, we can examine the mariner's process of thinking as rational, in the power sense, if MS3 was efficiently caused by MS1 and MS2. If it was caused by them, then the process counts as a process of reasoning. If not, then the process counts as an a-rational one. In that case, the criteria that have to be satisfied for a process to count as a rational one aren't satisfied in the case of MS1, MS2 and MS3.

This brings us back to the problem of error. The problem arises when 'engaging in reasoning' is identified with 'reasoning correctly'. If reasoning consists in responding to a reason and responding to a reason consists in reasoning correctly, then there is no room for error; there is no possibility of reasoning incorrectly. By definition, if something doesn't qualify as engaging in reasoning correctly, it cannot be judged in light of reasons and the principles of rationality.

What about the following causalist response? The causalist about reasons might protest: You neglect that other forces are operant in the mind, like desires or other urges. So, we can account for epistemic reasons as mental states that cause other mental states, despite being occasionally interrupted in their exertion of causal power. What the causalist tell us is this: when left uninterrupted, reasons/mental states cause other

mental states whose content rests in rational relations with the content of the mental states that constitute the epistemic reasons. There is a harmony between the causal force of reasons/mental states and the rational force of their contents.

Let us lay out two crucial possibilities: i) reasons MS1 and MS2 are uninterrupted in producing MS3, and ii) reasons MS1 and MS2 are interrupted in the production of MS3 by other forces; typically non-rational ones like desires or urges. In the first case, we have what the causalist pronounces to be reasoning par excellence. Reasons MS1 and MS2 cause MS3. In the second case, the causalist equates irrationality with the operation of a non-rational force which overrides the power of reasons MS1 and MS2. This would appear to make irrationality as the outcome of a non-rational force (which has to involve the causalist's relaxation of the definition of reasoning) and it also violates the causal element that the causalist barricades behind, in order to give a plausible account for when a process counts as a process of reasoning. For, if the mental state that results is not caused by reasons, it is not a process of reasoning that we are talking about, by definition. In that case the causalist has a following response. He can protest that it might be the case that lesser strong reasons interfere and override causally the exertion of force by stronger reasons, whose strength is rational; not causal. This threatens the whole of the causalist position because it dissociates the causal force of a mental state with its rational force. Such a dissociation leads into a treatment of mental properties, such as mental content, as irrelevant to causation.[30] That way, causation becomes a wholly non-rational force, and to be caused by a reason, is to be caused by the non-rational force of a mental state which has contents which can stand in rational relations. This puts us back into accounting for when a process counts as rational, in terms of the rational relations of contents of thoughts, irrespective of how they feature in a person's mental life. For, the way that they do in this account is wholly nonrational.

[30] An interesting discussion about this problem can be found in Segal and Sober (1991).

So, along the lines of the previous causal theory of reasoning, the thinker counts as either rational in the success sense or as a-rational. This causes a conflation between the conception of rationality as power and its conception as a success term. This is a result of the one-way power that is involved in the dominant conception of what it means to be guided by reasons; i.e. to think in light of reasons is to be in a mental state that is caused by other mental states, whose contents stand in rational relations with the content of the resulting mental state. I maintain that such a problem can be solved by conceiving of such a power as a two-way power. That means that to respond to reasons, reasons cannot be efficient causes of the passive outcome which constitutes the response to them.

If the considerations laid out in this section are correct, then causal conceptions of responsiveness to reasons are problematic. Responsiveness to reasons cannot be understood as a passive outcome of the causal force of reasons, conceived of as mental states. Instead, I maintain that it is an active engagement in thinking or overt action where we think and act in light of a reason, where that reason is not an efficient cause of our thinking in light of it. Models of reasons-responsiveness that construe reasons as efficient causes are unable to sustain the normativity of thinking. Thus, along these lines, I claim that the dominant conception of 'motivation by a reason' as 'being set in motion be a reason/efficient cause' is mistaken. To be motivated by a reason cannot consist in being set in motion by that reason.

Section B: ***Rationality and the normative question***

In section B, I discuss two objections against the idea that rationality is normative. The core of both will be discussed in B.1 and amounts to the normative question, which asks whether there are any categorical reasons to follow the principles of rationality. In section B.2, I will discuss the first objection which claims rationality to be concerned merely with attitudes, irrespective of reasons for and against them. As

such, it treats rationality as essentially separated from reasons. In section B.3, I will discuss the second objection to the normativity of rationality which concerns the wholesale denial of all kinds of normativity. In replying to both objections, I will assume the falsity of global epistemic skepticism and the position that knowledge requires internalist justification. If these presuppositions are made, it is hard to see how we can accept either objection without committing ourselves to global epistemic skepticism. If knowledge requires internalist justification and we can never have access to reasons that justify us in thinking something to be the case, we cannot know anything because we cannot be justified in thinking anything.

B.1 ***What is the normative question?***

The 'normative question', after Korsgaard (1996, 2009a), asks whether there are categorical reasons to follow the rules of an activity. In the context of this question, the 'normativity' of an activity does not refer to the fact that the activity in question has conditions of success or correctness. It goes one step further and asks whether human agents have reasons that are authoritative or binding (Hampton 1998) to follow such rules, if the activity we inquire about is in fact rule-governed. With respect to morality, Korsgaard writes: "When we seek a philosophical foundation for morality we are not looking merely for an explanation of moral practices. We are asking what justifies the claims that morality makes on us. This is what I am calling 'the normative question'." (1996: 9-10) In this case, it is clear that 'reasons' are not meant as causes, but as justifying considerations. Whenever we ask the normative question about any activity's standards, we ask not for what explains their adoption in practice, or what would be in accordance with them. What we ask for instead is a justification of the force that these standards claim to have on us.

Those who take rationality to be strongly normative think that rational principles are accompanied by categorical reasons to comply with them. Such reasons can be defeasible or indefeasible, depending on the particularities of the case we examine.

Violating such standards is strongly criticisable and implies something more than saying that you do not act in a certain way; it implies that you do not act in the way you *should* be acting. For those who embrace the normativity of rationality, to say that one is irrational is to say that they have done something that in a strongly normative sense they shouldn't; claiming that one is irrational is a normative criticism and finds the irrational agent as not doing the best she can in the circumstances she finds herself in. Concerning the criticism of irrationality, Nicholas Southwood (2008) writes:

> Local grammatical requirements are merely constitutive rules or conventions that do not possess any kind of intrinsic normative status. In violating a local grammatical requirement, we are guilty of nothing more than a conventional breach. Rational requirements, by contrast, seem to be normative in a deeper sense. If we fail to comply with them, it seems that we've necessarily gone wrong in some deeper way. (2008: 11)[31]

Those who object to the idea that rational principles are normative deny the above. For them, violating a rational standard implies irrationality, but irrationality is not strongly criticisable. When directed against the standards of rationality, the normative question asks whether we have any prescriptive reason to be rational, i.e. whether we should apply the principles of rationality correctly. And they maintain that not acting in light of rational standards does not merit normative criticism. To say that one is irrational is to say that she or he do not exhibit a property, without that implying that they should exhibit such a property. It becomes a purely descriptive claim. The reaction against the position that rationality is normative takes two main forms: i) the

[31] This can be seen also in terms of how people react to criticism, which is often experienced as implying not just that the person who is criticised doesn't conform to a standard in the same way that we say: your clothes aren't yellow. For a very informative discussion about the normative dimension of criticism, see Kiesewetter (2017: ch.2)

divorce between rationality and normative reasons, and ii) the wholesale rejection of normativity in the prescriptive sense. In section B.2, I will discuss the first and in section B.3, I will discuss the second objection.

B.2 ***Divorcing rationality from normative reasons***

B.2.1 ***Requirements of rationality***

It has become customary in contemporary discussions about reasons and rationality to treat rationality as codifiable in terms of certain rational requirements. These are rational requirements of coherence among attitudes.[32] For instance, if one believes that Australia is on fire, it is incoherent[33] for this person to also believe that there are no fires raging in Australia. If on the one hand, one intends to discharge one's responsibility as a citizen to safeguard public hygiene from a highly contagious virus and believes that staying at home is a necessary means for doing so, but on the other hand, does not intend to take the means necessary to achieving that end, then one has incoherent practical attitudes. If one believes that: p, p is sufficient evidence for q and also believes not-q, then one has incoherent epistemic attitudes.

It seems intuitively plausible to think that there is something wrong with incoherence. And such an intuition ultimately rests on the idea that rationality is normative; i.e. we have reason to be rational. And if rationality is coherence, we have reason to be coherent. To think that we have incoherent attitudes seems to give us a prima facie reason to revise them in order to restore balance between our attitudes. And to know how to revise them, we need to pay close attention at the requirements we use

[32] There are lots of distinctions that concern rational requirements, such as the one between state requirements or process requirements (Kolodny 2007). Due to lack of space, I will not deal with them in this thesis. A thorough discussion of requirements is given by Broome (2013: ch.7, 8).

[33] If incoherence is ultimately defined in terms of inconsistency, it doesn't seem possible to maintain that coherence is entirely an internal matter.

to move from one attitude to the next. On a first notice, we may codify the following requirements of rationality:

Consistency requirement – NS: if person S believes that p, then S is rationally required *not to believe that not-p*.

Instrumental requirement – NS: if person S adopts goal G, and believes that M is a necessary means to G, then S is rationally required to *intend to M*.

Evidential requirement – NS: if person S believes that p, and believes that p is sufficient evidence for q, then S is rationally required to *believe that q*.

These requirements are narrow in scope. This means that they are requirements of coherence among attitudes, which require of the agent to adopt a specific attitude.

Unfortunately, the narrow scope conception of rationality seems to be incompatible with the idea that there is something wrong about incoherent attitudes. As mentioned above, the idea that there is something wrong about incoherent attitudes is closely linked with the idea that we have reason to be rational; i.e. we have reason to do what rationality requires us to do. And if rationality is a set of requirements of coherence among attitudes, then we have reasons to conform to these requirements. According to the narrow scope conception of requirements of rationality, conformity with rational requirements is achieved or realized when the person whose rationality we assess, adopts the specific attitude that a requirement of coherence requires of her. And this has some very counterintuitive implications.

A significant problem that the narrow-scope treatment of requirements of rationality faces has become known as the 'bootstrapping objection', posed first by Michael Bratman (1981, 1987). The bootstrapping objection concerns the implausibility of the idea that it is possible to generate a reason to do something just

by adopting some attitude. For instance, just by believing X due to wishful thinking, it seems implausible to maintain that I have a reason to believe anything that follows from X. It is as implausible as saying that wishful thinking can justify theoretical beliefs. It seems not to be that easy to generate reasons for belief. In the practical domain, things are equally implausible. All that has to be in place for there to be a practical reason to do something is that the agent adopts a particular goal and believes that some means M are necessary for realizing that goal. This would allow us to generate reasons for intention and action that are quite implausible. Assume you suffer from a momentary lapse of reason. Your knee-jerk reaction is to hit someone. It seems implausible that it is so easy to generate a practical reason and a reason for intention. Thus, it seems that we cannot plausibly combine a narrow scope conception of rationality with the initially plausible assumption that there is something wrong with incoherence of attitudes. Treating such narrow-scope requirements as normative, generates reasons to engage in activities that from a common sense point of view seem entirely implausible.

Dissatisfaction with narrow-scope conceptions of rationality has led many thinkers to treat the requirements of rationality as wide in scope. This still makes them requirements of coherence among attitudes, but the main difference is that rationality requires not a specific attitude, but a coherent pattern of attitudes. According to the wide-scope treatment of rationality, rational requirements take form of the following kind:

Consistency requirement –WS: A thinker S is rationally required to [believe that p and not to believe that not-p ***or*** believe that not-p and not to believe that p]

Instrumental requirement –WS: A thinker S who adopts goal G and believes that M is a necessary means to G is rationally required to [reject G ***or*** intend to M]

Evidential requirement –WS: A thinker S who believes that p and that p is sufficient evidence for q, is rationally required to [reject that p ***or*** reject that p is sufficient evidence for q ***or*** believe that q]

Wide scope treatments of rational requirements are requirements of patterns of attitudes. Their main difference from their narrow-scope counterparts concerns the fact that they do not require the adoption of a specific attitude; they can be satisfied by any move that restores coherence to an incoherent pattern of attitudes or achieves coherence when a new element is being encountered.

Wide scope treatments of rational requirements initially seem to be more plausible than their narrow scope counterparts. Particularly, they seem to be quite compatible with our intuitions about reductio ad absurdum arguments or syllogisms. The person who begins to draw implications from a particular belief that begin to seem outlandish, can also be rational by rejecting that very belief. This cannot happen in the narrow scope treatment of rational requirements where one particular attitude is required of a person. This is not the case in the wide scope treatment of rational requirements because we have considerably more leeway in the way in which we can achieve coherence among a pattern of attitudes. If we stumble upon the realization that our attitudes require us to adopt an attitude that is counterintuitive to think that we should adopt it, wide scope requirements allow us also to drop the initial attitude that in a narrow scope reading of theirs, generates a reason to adopt the implausible or absurd attitude. It is rational for the agent to revise the intention or belief, if it leads to that implausible consequence, acceptance of which constitutes the reason why the normativity of narrow-scope requirements seems questionable.

But this idea seems short-lived. The issue with the narrow-scope treatment of rational requirements lies in its uneasy combination with the normativity of rationality; the idea that we have reason to be rational. It seems that the wide-scope treatment falls into the same trap, albeit in an indirect way. It seems that a normative understanding

of wide-scope requirements of rationality will still generate reasons that might favour the adoption of attitudes, for which it seems implausible to think that we have reasons to adopt. The reason is that in such problematic cases as the ones that cause problems for the narrow scope treatment of rational requirements, it implies that it is *equally* rational for the agent to adopt the consequent attitude as it is to drop the antecedent attitude. It is equally rational to adopt the intention to hit someone, as it is to drop the belief that we ought to hit him.[34] It is equally rational to embrace a belief that fossils are evidence that favour belief in the fact that there has been past life on Earth, as it is to believe that such fossils are fake and were placed there by those who would benefit if we believed that such creatures actually existed. This leads us to the idea that adopting wide-scope rational requirements leaves the challenge against the narrow-scope strategy, largely unanswered. For even along the lines of the wide-scope strategy, it is also rational to adopt the otherwise unacceptable intention or belief. Whereas in the narrow-scope case we can detach attitudes that we consider unacceptable, the wide-scope conception gives us more leeway. Unfortunately though, if we combine it with the idea that rationality is normative, it implies that it is equally rational to reject the unacceptable attitude as it is to adopt it.

The apparent plausibility of the wide-scope treatment does not seem to withstand scrutiny. The reason why the wide-scope view seems initially to be more rational than its narrow-scope counterpart, in the first place, is that it allows for the idea that it is rational to drop the antecedent attitude, if it leads to the implausible conclusion. But that idea is forceful because it implies that we have *more reason* to reject the antecedent attitude than we have to adopt the consequent attitude which seems implausible. And although an improvement from its narrow-scope counterpart, it is hard to see how the wide-scope treatment of the principles of rationality can account for that idea. If the

[34] Scanlon (2014) holds that this is one of the objections that can be given against desire-based conceptions of rationality. He writes: "the problem with normative (actual) desire theory is not just that it leads to implausible conclusions about the reasons people have in particular cases but that it misdescribes the relation between desires and reasons from an agent's point of view." (2014: 87-88)

agent is rational in doing either, it seems that the reasons for adopting the consequent attitude are equally forceful as the reasons for rejecting the antecedent one. To illustrate a case in the epistemic realm, the wide-scope doxastic requirement, since purely formal, would imply that it is equally rational for an agent to believe that p, as it is to reject that there is sufficient evidence for believing that p.

B.2.2 ***Substantive and structural rationality***

Many thinkers believe that the solution to the bootstrapping problem lies in the rejection of the normativity of rationality. They hold that whether we think that rationality is to be codified in terms of narrow or wide scope requirements is irrelevant. The problem arises from the idea that we have reason to follow what rationality requires of us. They maintain that reasons and rationality are not essentially connected. Upon that basis, an objection arises to our account. Rationality isn't normative because it is not the case that rationality and reasons accompany each other, necessarily. After laying out the objection, I will claim that it is not clear exactly why this is considered to represent a significant challenge. We can allow that some instances of rationality are not accompanied by reasons, but we cannot maintain that all instances of rationality are unaccompanied by reasons, at pains of embracing a radical form of skepticism. Furthermore, whether normative or not, it seems implausible to maintain that rationality is an entirely internal matter, concerned only with coherence among attitudes. That would render conspiracy theorists the paradigmatic cases of rational thinkers.

It has become customary to draw a distinction between substantive rationality and structural rationality. In light of this distinction, Parfit (2001) claims that what we have most reason to do differs from what is the rational thing to do. According to structuralists about rationality, to be rational consists not in responding correctly to

reasons, but in achieving coherence among one's attitudes. To highlight this distinction, Kolodny writes:

> When we say 'you ought to' in the sense of 'you have reason to,' we usually seem to be saying something about the relation between your *situation* and your *attitudes*. When we say 'you ought to' in the sense of 'it would be irrational of you not to,' we usually seem to be saying something about the relation between your attitudes, viewed in abstraction from the reasons for them. We are saying something, for example, about whether your beliefs are logically consistent, or whether your intentions for ends cohere with your intentions for means – things that are true, if they are, quite independently of whether there is a reason for you to have any of those beliefs or intentions. (2005: 509-510)

Substantive rationality consists in thinking and acting in light of reasons. Structural rationality consists in achieving coherence among attitudes. At any moment we have various attitudes, some of which are epistemic. And some of our epistemic attitudes are attitudes that we do not have reasons for. Thinking has to begin from somewhere. If we take these attitudes as the starting point of our enquiry and we apply rational requirements on them, we are going to reach other attitudes for which we also do not have reasons. Eventually, we might encounter cases where these conflict with each other; cases where what is rational to do is what we have no reason to do, or worse, what we have reason not to do.

Certain philosophers, most notably Broome (2005) and Kolodny (2005) have used this distinction to deny the normativity of rationality. The central idea that is expressed by advocates of this objection is that rational requirements don't give us reasons to think and act rationally, as opposed to irrationally. Advocates of the normativity of rationality hold that: "To claim that rational requirements are normative is to claim that such requirements necessarily involve reasons (possibly decisive ones)

to conform to them." (Kiesewetter 2017: 149) Those who are sceptical of the normativity of rationality deny this very claim. They disconnect rational requirements from normative reasons.[35] In this sense, skeptics of the normativity of rationality can hold that there aren't always reasons for an agent to satisfy the requirements of rationality. There can be cases where the agent has no reason to satisfy the requirement, or even cases where the agent has a normative reason to violate such rational requirements. And if there are reasons *in some cases* to do what is the rational thing *in those cases*, they are not reasons that are generated by requirements of rationality.[36]

The bottom line of their objection is that rationality is not always accompanied by normative reasons. It is concerned with coherence of attitudes, irrespective of our reasons for and against each particular attitude. In some cases there are reasons to do that which from our perspective seems irrational to do (Broome 2007). Such normative reasons may even be independent of our perspective. Broome holds that this view entails that sometimes the rational thing to do is what we have decisive reasons against doing. He writes:

> The fish on the plate in front of you contains salmonella. This is a reason for you not to eat it. But you have no evidence that the fish contains salmonella. Then you might eat it even though your reasons require you not to, and nevertheless you might be rational. (2007: 352)

[35] This does not imply that they deny the existence of normative reasons. (E.g. Broome, 2007: 352)

[36] According to proponents of such a view, rational requirements should be seen just as standards of non-normative activities, which don't provide us with reasons to engage in them, in the prescriptive sense. There is no normative reason, in the categorical sense, to follow the rules of basketball, and no such reason not to. Therefore, due to this lack of a normative reason, which speaks in favour of following the requirements of rationality, violating such requirements isn't strongly criticisable.

For Broome, the fact that the fish is poisoned is a normative reason that requires us not to eat it.[37] If we are hungry and ignorant of the fact that the fish is poisoned, it is rational to eat it despite the fact that we have a normative reason not to do it. So, in this case, we have reason not to do that which is required by the application of the principles of rationality to our situation. Instead, we have a normative reason to be, in this case, irrational. We just don't know it.

Divorcing normative reasons from rationality leads to the distinction between substantive and structural rationality.[38] The former concerns what we have normative reason to do, whereas the latter concerns the structure of our attitudes, independently of whether we have normative reasons to have them. Kolodny writes:

> Normativity involves two kinds of relation. On the one hand, there is the relation of being a reason for. This is a relation between a fact and an attitude. On the other hand, there are relations specified by requirements of rationality. These are relations among a person's attitudes, viewed in abstraction from the reasons for them. (2005: 509)

Such a distinction between substantive and structural rationality rests on the rejection of a reason-responsiveness understanding of rationality.[39] For if we accept that reasons are normative and that rationality consists in responsiveness to reasons, we cannot maintain that rationality is not accompanied by reasons. According to a structuralist conception of rationality, rationality concerns coherence of attitudes,

[37] In this sense, Broome thinks that we may have normative reasons, without occupying the normative standpoint with respect to them [A.4].

[38] For a denial of the conception of reasons-responsiveness as distinct from mere coherence of attitudes, see Wedgwood (2017). An opposite view can be found in Lord (2017) and Kiesewetter (2017).

[39] Broome (2007) denies that rationality consists in responsiveness to reasons. He thinks that this violates our intuitions that in various cases we have reason to do that which is irrational from our perspective to do, as well as that it leads to bootstrapping. Broome (1999) maintains that *a part of* our reasoning concerns coherence among attitudes; not all of our reasoning.

whereas normative reasons concern relations between facts and attitudes. One is rational, in so far as her attitudes are structured in accordance with the requirements of rationality, independently of the normative reasons that might speak in favour of or against them. The bottom line is that rationality and reasons come apart. It might be rational to do something we have normative reasons against doing. Admitting the fact that such a gap might exist leads us to admit the fact that what is rational in some cases doesn't have to be accompanied by normative reasons, or most crucially, can be something we have normative reasons against doing. That means that it is not always the case that we should be rational. So, they conclude that there is no fact of the matter as to whether we should or shouldn't be rational.

B. 2.2.1 ***First reply***

There are lots of ways to understand the distinction between substantive and subjective rationality. Kolodny's formulation of the distinction is ambiguous. More specifically, the ambiguity lies between a conception of the gap as one that implies that we never have access to reasons and another which implies that we do have but we also have other attitudes that are not supported by reasons. Unless we are clearer about what we mean by the terms substantive/objective and structural/subjective rationality, we run the risk of conflating what I will call subjective descriptions of reasoning with descriptions of subjective reasoning.

The objection to the normativity of rationality will be examined in light of its strength in showing that in thinking we are not normative agents. In what follows, I will assume an internalist conception of justification, according to which we can be justified in thinking that p is the case, only if we have or can have access to a reason for thinking that p is the case. I will claim that certain notions need to be introduced, which can improve our understanding with respect to what is involved in this distinction. Finally, I will argue that the plausible ways of drawing the distinction

between substantive/objective and structural/subjective rationality are entirely consistent with the idea that in our thinking, we usually are under the scope of normative epistemic reasons.

The distinction between substantive rationality and structural rationality is ambiguous. It is not certain whether we appeal to it to make an ontological claim about the nature of reasoning or whether it refers to a description of a process of reasoning, which does not mention reasons for or against these attitudes. To illuminate the difference I have in mind, I will introduce and discuss the following terms:

i) metaphysically objective rationality
ii) metaphysically subjective rationality,
iii) objective description of rationality,
iv) subjective description of rationality,
v) Globally metaphysical objective rationality
vi) Globally metaphysical subjective rationality

By metaphysically objective rationality, I understand the kind of rationality that is exhibited by a thinker who engages in a process of thinking, in light of reasons. The thinker recognizes reasons and thinks in terms of what they support. To think in a metaphysically objective rational manner is to think in light of epistemic reasons.

By metaphysically subjective rationality, I understand the kind of rationality that is exhibited by a thinker who engages in a process of thinking, whose starting point consists in attitudes for which the thinker has no reasons, whether for or against. The thinker identifies an attitude, such as a belief or an intention and proceeds in thinking what is supported by that belief or intention. Such a process of reasoning does not require epistemic access to reasons for and against these attitudes. Still, one can engage in metaphysically subjective reasoning while having reasons for the attitudes that she or he will use as starting points, if they don't think of these reasons and go straightforwardly to see what follows from them.

By an objective description of rationality, I understand a description of a process of thinking which describes not only the attitudes that a thinker adopts and the successive transitions in thinking, but also the reasons that justify them or render them unjustified. Such a description can be a first-person description of one's own processes of thinking or description of someone else's processes of thinking.

By a subjective description of rationality, I understand a description of a process of thinking which describes only the attitudes that the thinker adopts and uses as starting points in her or his thoughts, without explicit reference to reasons for and/or against them. It is crucial to note that subjective descriptions can be given of processes of metaphysically objective reasoning.

I will use the term 'globally metaphysical objective rationality' to stand for the position according to which all instances of rationality are instances of metaphysically objective rationality. The advocate of such a view denies that there is any process of thinking that counts as reasoning which is not accompanied by epistemic reasons.

Finally, I will use the term 'globally metaphysical subjective rationality' to stand for the reverse position. Someone who embraces this position holds that all instances of reasoning are necessarily divorced from normative epistemic reasons. In the remaining of this section and chapter, I will claim that the only position that poses a threat to the idea of normative responsibility in our thinking is the position which holds all instances of reasoning to be metaphysically subjective, and which does not allow a thinker's access to such reasons. It is the position that we can have no access to reasons for belief that has to be combated if we are to make room for normative responsibility in our thinking; for being justified or unjustified in our thinking.

Suppose one claims that rationality is not normative but we have reasons for and against the belief that p. Then, we may say that we have no reason to be rational but we do have a reason that we can think in light of. Once this is accepted, the issue becomes largely terminological. And it is plausible to accept it if we claim that we can have knowledge and that knowledge requires justification; reasons for belief. We can

say that although we have no reason to be rational, we have reason to be rational*. Or, instead of claiming that rationality is normative, we may concede that it is not, but that reasoning is. Now, all this boils down to is the idea that although we may have no reason to be structurally rational, we may still have reason to be substantively rational. The question 'why be rational?' takes the more plausible form: 'why be structurally rational?' Surely, in some cases we have no reasons to believe or disbelieve a proposition and what follows from it. But inferring from that position that there is a global shortage of access to reasons is a huge leap, and an implausible one at that.

Another thing that is not clear about the distinction between substantive and structural rationality is whether it concerns a description or a claim about the metaphysical nature of the process. Very frequently, it is said that there is something wrong with incoherence. The bootstrapping objection to the narrow scope treatment of rational requirements, as well as its wide scope counterpart, is generated to a very large extent by the idea that such requirements are normative. Incoherence among attitudes is a sign of irrationality. Again, it is a very huge leap to say that because incoherence among attitudes is a sign of irrationality, and incoherence is an internal thing, that rationality is coherence. Incoherence can be a sign of irrationality for metaphysically objective reasoning as well. The point I am making is that when we describe processes of thinking and judge what goes right or wrong with them, we may focus on their formal aspects and conflate claims about the nature of our *description* of such processes with claims about their nature. That one can describe a process of reasoning in a way that doesn't refer to reasons for and/or against the attitudes involved in it, does not mean that there are no reasons involved in it. Hence, a subjective description of a process of reasoning isn't always a description of a metaphysically subjective process of reasoning. It could be the case that this specific piece of reasoning that is described is one of the metaphysically subjective kind –as defined above- but it could also be one of a metaphysically objective kind.

Let us recapitulate. It is frequently held that rationality is not normative because it is not always the case that being rational is accompanied by reasons. This is

frequently presented as a challenge to the normativity of rationality and consequently, as a challenge to the idea that we are frequently under the scope of normative reasons. In reply to this, we can maintain that the force of this challenge is exaggerated. Reasons that justify this claim can be provided by careful reflection on the distinction between substantive/objective and structural/subjective rationality. Its frequently ambiguous formulations make it easy to conflate a description of reasoning that excludes reference to possessed reasons with a description of structural/subjective reasoning. With the introduction of some useful notions, we can see that at pains of embracing a radical epistemic skepticism, according to which we can never have access to reasons, the challenge to the normativity of rationality amounts more plausibly to a challenge of the normativity of structural/subjective rationality. It is one thing to claim that structural/subjective rationality is not normative. It is a wholly different thing to claim that rationality is not normative. If we assume that knowledge requires reasons for belief, the price of making that leap seems to be a radical form of epistemic skepticism.

B.2.2.2 ***Coherence and conspiracy***

It is surprisingly easy to be coherent if you are a conspiracy theorist. In cases where we encounter incoherent patterns of attitudes, the wide scope treatment of rational requirements allows us to reject the idea that any sort of evidence against your theory is sufficient to falsify it. Not only is it possible for the conspiracy theorist to achieve coherence, it is also remarkably easy and economical to do so. All the conspiracy theorist has to do is to adopt the belief that any person who gives evidence against their theory is a member of this conspiracy and tries to hide the truth from reaching other people or to reject the belief that any criticism is sufficient to override his belief in conspiracy theories.

To motivate this idea, let us think of the following thought experiment. Assume that Archibald, a patient in a mental asylum has the belief that whenever he is not looking

at a television's screen or a mirror, there are faces there staring at him. His doctor, Lena, tries to convince him that this belief is mistaken and does so by means of offering evidence to the effect that this isn't the case. She asks him to look at the mirror and see whether someone is looking at her when she doesn't look at the mirror, with the hope of convincing him that if she isn't looked upon when she isn't looking, there is no reason for him to believe that he is looked upon when he doesn't look as well. Archibald can adopt a wide-scope treatment of the Modus Ponens principle, and hold that: he ought to [not to believe that Lena isn't watched *or* not believe that if Lena isn't watched then he isn't watched *or* believe that he isn't watched]. The problem with the wide-scope treatment of the Modus Ponens rational requirement in this case, is that it makes it equally rational for Archibald to believe either of the three options. All three of them are equally permissible. If all there is to rationality is achieving coherence of attitudes, then each of the three options is equally rational. It is as rational for Archibald to believe that he isn't watched, as it is for him to believe that the fact that Lena –or no one else for that matter- isn't watched doesn't give him a reason to believe that he isn't watched. Now, if we find it obvious that these options aren't equally rational, we will have to reject the wide-scope treatment of the Modus Ponens requirement of rationality. It makes an intellectual virtue out of what most would consider to be an intellectual vice.

To conclude, if we think that conspiracy theorists are not rational thinkers par excellence, we have reason to reject the wide-scope treatment of rational requirements. Furthermore, if we accept that structural requirements of rationality don't exhaust rationality, we should accept a reasons-responsiveness view of substantive/objective rationality.

B.3 ***The error theoretic objection***

Moral error theory is an umbrella term for various cognitivist, anti-realist theories of moral judgments. Unlike non-cognitivists who do not treat moral judgments

as truth-apt, moral error theorists hold that moral judgments are uniformly false. An error theorist about a domain of discourse holds that all judgements that qualify as that domain's distinctive judgments are mistaken. The main reason for which they embrace this position is a conceptual commitment to the position that moral judgments, in virtue of their reference to moral properties, are inherently prescriptive. And such an inherent prescriptivity of moral facts, reasons and properties does not seem to be compatible with a naturalistic worldview (Mackie 1977).

The insights of moral error theory can form the ground of an error theory in other domains. The denial of prescriptive properties creates problem not just for moral requirements, but for all irreducibly normative requirements. To the extent that epistemic reasons –reasons for belief- are normative reasons, an error theory about all normative judgments would imply that all judgments about reasons are uniformly false. Thus, an error theory about normativity in general, implies that it is wrong to think that we are ever under the scope of normative reasons, whether in thinking or acting. According to the terminology introduced in section B.2.2.1, error theorists about normativity embrace globally metaphysical subjective rationality; the position according to which, all instances of reasoning are instances of structural/subjective rationality.

One of the criticisms against the error theory is the companions-in-guilt criticism. According to companions-in-guilt arguments, prescriptive properties of the kind that characterize the domain of moral discourse and practice, are essential features of other domains as well, like the domain of epistemology. A fortiori, eliminating prescriptive properties has unwanted side effects not just in ethical discourse, but to epistemological discourse as well.[40]

[40] For an interesting discussion as to whether such arguments work, see Christopher Cowie (2016) who argues that they are inconsistent, and Ramon Das's (2016) response. In the text that follows, I will agree with Das, although for independent reasons.

Terrence Cuneo (2007) has put forward one of the most sophisticated companions-in-guilt arguments. According to Cuneo, consistent error theorists will have to deny epistemic reasons for belief, for precisely the same reason that they deny moral reasons; they entail facts about how and what one *should* think. This 'should-clause' represents a prescriptive fact which is queer naturalistically speaking, and therefore, has to be rejected in the precise manner that moral reasons are rejected. Cuneo offers three arguments/dilemmas against the error theory. First, the error theorist is faced with a choice between self-defeat and lack of reasons to believe anything. Second, he argues that error theorists are faced with a choice between self-defeat and the total elimination of epistemic merits or demerits. Third, the epistemic normativity error theorist, is faced with a choice between self-defeat and the position that there are no arguments for anything whatsoever.

Cuneo's first criticism against epistemic normativity error theory presents its advocates with the following dilemma. According to Cuneo: "Either epistemic nihilism is self-defeating and hence, we have no (sufficient) reason to believe it, or, it implies that there are no epistemic reasons and, a fortiori, that we have no reason to believe it." (2007: 118) According to this dilemma, believing in the error theory can never be justified. There can never be reasons to believe in the error theory. If error theorists hold that there are reasons to believe the error theory, then their position is self-defeating. For, they claim their position to be true on the basis of that whose truth is excluded by their theory. On the other hand, if there are no reasons to believe anything, there is no reason to believe in the error theory. Consequently, there is no reason to believe in the error theory about epistemic normativity.

Cuneo's second argument presents another dilemma to the error theorist, one of whose horns is also a charge of self-defeat. According to Cuneo: "Either epistemic nihilism is self-defeating or it implies a radical version of epistemological skepticism according to which no entity can display an epistemic merit or demerit." (ibid: 119) If there are no reasons to believe anything, it seems very hard to conceptualize discourse. Our thinking won't be able to be guided by anything and thus, there will be no standards

according to which, we can call thinking as justified or unjustified, warranted or unwarranted, right or wrong. Cuneo writes:

> For if there are no epistemic facts or reasons, then none of our propositional attitudes can exhibit epistemic merits or demerits; none of our propositional attitudes can be justified, warranted, entitled, irrational, a case of knowledge, based on reasons, or the like. (ibid: 119)

But although this is sufficiently unattractive, for Cuneo it isn't enough to establish the falsity of epistemic nihilism; the position that there are no facts of the matter as to what a person should think or not. For him, holding that this implies the falsity of epistemic nihilism, is question-begging. We can only hold that epistemic nihilism is false, if we presuppose that which is to be proven, namely, that there are reasons to believe or not.

Cuneo's third dilemma charges the error theorist with the unfortunate consequence of eliminating arguments. He writes: "Either epistemic nihilism is self-defeating or it implies that there could be no arguments for anything." (ibid: 121) The third dilemma follows naturally from the previous one. Our unfortunate predicament with respect to our inability to give a *coup de grace* to epistemic nihilism consists in the question-begging nature of any such attempt. Such attempts would have to presuppose the existence of reasons for belief without being able to counter the error theorist's objection that this is a dogma of epistemology. Cuneo writes:

> Were a person to press such an argument against nihilism, she would endeavour to establish that there are reasons by tacitly assuming that some statements provide evidential support to believe others. But –and this is the second implication- if nihilism were true, it would be impossible that there were premises of an argument that provide evidential support for its conclusion. (ibid: 121)

In response to these criticisms, some error theorists (Olson 2011, 2014, Streumer 2017) have seen it fit to bite the bullet and deny epistemic normativity altogether. According to Olson (2011) there is a distinction to be made between reducible and irreducible normativity. Reducible normativity concerns standards of correctness, which are not accompanied by reasons to comply[41] to them. The normativity of rationality is considered to be irreducible, by those who are not skeptical about substantive rationality, conceived of as responsiveness to categorical reasons. Olson (2014) claims that we can reject irreducible normativity and embrace the existence of instrumental reasons for belief.

Olson (2014: ch.8) undertakes the task of formulating a version of error theory about epistemic normativity which will answer Cuneo's three criticisms against error theoretical accounts of reasons for belief. The first horn of each dilemma that Cuneo presents to the error theorist, amounts to the charge of self-defeat. The error theorist defeats himself when he embraces error theory about epistemic normativity because if he believes in it in a principled way, he has reasons to believe in it. Olson thinks that those who think that "the fact that astronomers report observations of cosmological redshift is a reason to believe that the universe expands…" (2014: 155), make an illegitimate transition from descriptive facts to normative facts. Statements about facts are descriptive, whereas statements about epistemic reasons are normative. The transition from the former to the latter is mistaken and quite frequent.

This charge forms the core of Olson's defense of the error theory about epistemic normativity. It leads him to draw a distinction between arguments to the effect that there are reasons to believe something and arguments to the effect that something is true. He thinks that by pointing out the difference, the error theorist can respond to the charge of self-defeat. He writes:

[41] Complying with a standard is different than plainly conforming to it. The former implies an intention, whereas the latter can be an accident.

> But error theorists can distinguish between arguments to the effect that *p* and arguments to the effect that there are epistemic (irreducibly normative) reasons to believe that *p*. In particular, they can hold that their error-theoretical arguments are arguments to the effect that the error theory is true and not to the effect that there are epistemic reasons to believe the error theory. This also shows why there is no harm in conceding that epistemic error theory is polemically toothless in debates about what there are epistemic reasons to believe. (ibid: 159)

Olson holds that arguments are frequently thought of in normative terms, in that it is held that the propositions that form the argument's premises support a particular doxastic attitude towards the proposition that forms the argument's conclusion. He rejects this conception of arguments and claims that no such relation of 'favouring' has to hold between the premises and a doxastic attitude. He writes: "I take an argument to be a series of propositions such that the propositions expressed in the premises indicate (or purport to indicate) that the proposition expressed in the conclusion is true." (ibid: 157n) Arguments are frequently presented in normative terms because non-error theorists about epistemic normativity make the illegitimate move of drawing a normative conclusion from entirely descriptive premises. Once this is seen, arguments can be presented in an entirely descriptive manner. Arguments for error theory are not self-defeating because they do not rest on reasons to believe the error theory. They are arguments which move from descriptive premises to a descriptive conclusion. The error lies in those who aim to extract a normative conclusion from descriptive premises. And such an error is an error in light of a standard that is not strongly normative; a standard that there is no reason to follow.

Many have protested against such an error theory on the grounds that the notion of evidence is an irreducibly normative notion (Kim 1988; Kelly 2003, 2007; Shah

2011). [42] Eliminating epistemic normativity leads to eliminating evidence. Any conception of epistemology as a normative discipline seems to be straightforwardly incompatible with a view that rejects normative reasons to think or believe something to be the case. Expressing this standpoint, Jaegwon Kim claims that "a strictly nonnormative concept of evidence is not our concept of evidence: it is something we do not understand." (1988: 391)

At this point, it is proper to make several remarks about this error theory. If knowledge requires reasons for belief and such reasons are irreducibly normative, then an elimination of normativity would eliminate the possibility of knowledge. The error theorist could respond by claiming that knowledge can be secured by treating epistemic reasons as reducibly normative. This seems to amount to the project of naturalizing normativity, where all reasons are to be reduced to instrumental reasons, which at best, are to be conceived of as weakly normative.[43] Furthermore, the error theorist could claim that such an argument is question-begging and is not strong enough to show that we actually possess knowledge. A response to the first counter-reply requires an extensive defense of a realism about normative reasons which, although presupposed in this thesis, is not its main theme. For this reason, I fully accept the fact that the arguments offered in it are not likely to be convince the error theorist. Nevertheless, since a project is contextualized by its presuppositions, it is not my purpose to convince error theorists in this thesis. My purpose is to show that someone who is not an error theorist has a plausible case to make for the claim that in thinking and acting, we are frequently under the scope of irreducibly normative reasons. We can respond to the second counter-reply by claiming that convincing the epistemic skeptic about the truth of any claim cannot be a criterion of adequacy of any position. To claim that companions-in-guilt arguments cannot show that the error theory about epistemic

[42] Some responses are given by Adam Leite (2007) and Olson (2014).

[43] It doesn't do justice to the error theory about the normative to present it as being close to an expressivist conception of reasons because the former treats statements about reasons as truth-apt, whereas the latter conceives of them as non-truth-apt.

normativity is false because they cannot show that we ultimately have knowledge, seems like a hit below the belt. The strength of the error theory is seen to be the strength of skepticism against non-skeptical views. Arguably, this seems to imply that a debate can be had where it cannot. In this case, the error theorist's currency is not accepted in non-skeptic circles. If we occupy a realistic perspective about normative reasons, we may answer by finding the error theory to be unmotivated. We can admit that we cannot convince the error theorist but that does not mean that the error theorist can convince us. Particularly, the view that arguments to the effect that something is true are distinct from arguments to the effect that there are reasons to believe it seems hard to embrace. It is people who argue; not propositions. Even if one is a realist about propositions, one can still maintain that in a world with no intelligent beings, there are no arguments because there is no one who argues. In addition to that, those who embrace this distinction might conflate a description of an argument that omits reference to how the propositions that comprise the argument's premises and its conclusion feature in one's mental life –how they are thought of or entertained by a person- with a description of the relations between propositions, which is equivalent to the aforementioned conflation between subjective descriptions of reasoning and descriptions of subjective reasoning [B.2.2.1]. A corollary of this is that in drawing this distinction, the error theorist may also conflate the order of understanding with the order of being. The fact that in order to understand what a reason favours –if such favouring relations exist- we have to understand what it supports, does not imply that the relation of support can exist prior to and independently of the favouring relation. For, relations of support are *epistemic* relations. And a fully descriptive account of epistemic relations, one which doesn't refer to epistemic reasons, seems to lead to the claim that epistemic relations can exist independently of beings who think. It seems plausible for a realist about normative reasons to claim that this is an illegitimate move.

When it comes to rational agency and the normativity of rationality, this chapter is inconclusive. In the next chapter, I will argue in favour of a rationalist approach to a

priori knowledge and justification. Error theory rests on a naturalistic worldview that, if undercut, doesn't have to be dealt with, head on. If naturalism is false, then belief in the error theory about epistemic normativity becomes unmotivated.

Conclusion

This chapter features two sections. In section A, the notions of reason, reasons and rationality are discussed. Reason as a faculty is understood in a rationalist sense, as an intellectual faculty which enables us to grasp necessary facts and relations. In section A.2, various kinds of reasons are discussed in terms of the ways in which they feature in the relevant domains that we inhabit; epistemic, practical and evaluative. In section A.3, reasons are seen in their normative dimension. An epistemic reason is a reason that justifies the adoption of a particular epistemic attitude. It is a reason to think that something is the case. A practical reason is a reason to engage in a particular physical action, and an evaluative reason is a reason to value something. In section A.4, a general framework of normative responsibility is presented that sheds more light on the conditions that have to obtain in order for an agent to be normatively responsible. To be normatively responsible in light of a particular normative reason or obligation, the agent has to be able to act in light of it; not blindly. Section A.4.1 involves an argument, according to which, normative responsibility requires an agent's alternative possibilities. If we assume OIC, we will have to accept PAP because without the latter, the former loses its prescriptivity. Section A.4.2 involves an argument against causal conceptions of rationality and reasoning. If one defines 'engaging in reasoning' with 'engaging in reasoning correctly', no room for incorrect reasoning will be left. To respond to a reason is not to be efficiently caused by a reason; rather, it is to freely engage in thinking and acting in light of one's goals.

In section B, I discuss two major objections to the idea that reason is normative. In order to situate these objections, I make an introduction to the normative question [B.1.1] and I refer to recent attempts to treat rationality as normative and as codifiable

in terms of rational requirements, whether of wide or narrow scope [B.2.1]. The apparent implausibility of such attempts leads various thinkers to divorce rationality from reasons, and reject the normativity of the former, after several qualifications are made. Such qualifications involve a distinction between what is called 'substantive' and what is called 'structural' rationality [B.2.2]. In section B.2.2.1, a first reply is given to the attempted divorce between rationality and reasons. The distinction is ambiguous and efforts to qualify what it involves, lead us to claim that the only position that would pose a significant challenge to the idea that there are normative epistemic reasons would be one which would claim all instances of reasoning to be processes of transitions from some attitudes, for which we have no reasons for or against, to other attitudes. Such a view seems bizarre and seems to have skeptical conclusions in that it conflicts with the idea that knowledge requires justification; i.e. reasons to believe that something is the case. In section B.3, an error theoretical approach is discussed which targets epistemic normativity. Proponents of such an objection hold that although epistemic reasons possess the function of supporting propositions or facts, they have no 'favouring' function. Such views are discussed and found implausible for the same reason as the objection that reasons and rationality are utterly divorced. Nevertheless, error theorists are right about one thing. Normative reasons seem to fit uneasily to a naturalistic worldview. An argument that concludes in the implausibility of a criticism, doesn't show the plausibility of that which is targeted by the implausible criticism. To make my position more plausible, I will proceed in chapter II to claim that significant parts of our epistemic inventory that error theorists find 'queer' are necessary for the possibility of knowledge and justification.

Chapter II: The indispensability of a priori justification

In this chapter, I claim that a priori justification is indispensable for reasoning, our justified error ascriptions, and consequently, for our engagement in inquiry. The chapter features three sections. In section A, I will offer some conceptual clarifications as to what the concept of a priori justification means; epistemic justification that is independent of *sense*-experience. In section B, I proceed to give a brief account of deficiency arguments against non-rationalist epistemologies as well as the main strategies that philosophers adopt towards them: rationalism, Kantianism, logical empiricism and radical empiricism. In section C, I present an argument for a priori justification and defend it against objections. It will be argued that a priori justification is indispensable for inquiry because it is necessary for the justified and principled rejection of beliefs, at pains of embracing a radical form of epistemic skepticism. Rather than incoherent, radical empiricism will be presented as inferior to rationalism in combating epistemic skepticism. The discussion will be situated in a broader epistemological context which is characterized by its treatment of philosophical skepticism.

Section A: ***The concept of a priori justification***

In section A, I explore the concept of a priori justification. In A.1, the main difference between internalism and externalism about epistemic justification is briefly discussed and a broadly internalist conception of justification is assumed for the thesis. In section A.2, I discuss the difference between a priori and a posteriori justification and the fact that a priori justification is frequently defined in a negative way. Two accounts are given. The first defines a priori justification as justification independent of experience, whereas the second defines it as justification independent of *sense*-experience. I argue that the first one is a seriously misleading way to define a priori justification and engage in the relevant debate since it presupposes the truth of

empiricism's core position without answering the main challenges that rationalists direct against it. Defining a priori justification as independent of 'experience' is to stipulate –in a seemingly innocuous way- an empiricist ontology of mind and its account of what experience ultimately consists in. A priori justification will be conceived of as justification independent of sense-experience.

A.1 ***Internalism and externalism about epistemic justification***

It is perhaps safe to claim that at least since Plato, the overwhelming majority of philosophers hold that epistemic justification is a necessary requirement for knowledge.[44] One of the goals of Socrates in most of the Platonic dialogues is to question his interlocutors in order to show them that they suffer from double ignorance; they do not know that they don't know what they claim to know. They think they do but when pressured to back up their points, they eventually understand that they don't and that they cannot justify their knowledge claims. To know that *p*, tradition holds that among other conditions, we have to be epistemically *justified* in believing that *p*. Various reasons can be given for entertaining some beliefs but the relevant ones in the context of inquiry, are *epistemic* reasons; not moral or pragmatic ones. This distinction is the basis of the philosophical quest for truth, which holds ordinary claims of knowledge to be accountable to standards which to a non-philosopher seem extra-ordinary. That one engages successfully in actions that may be practically beneficial

[44] Bruce Russell (2014) writes: "Gettier examples have led most philosophers to think that having a justified true belief is not sufficient for knowledge, (…) but many still believe that it is necessary." (2014: 1) One of the chief representatives of externalism about justification, Alvin Goldman (1979) writes:

> In previous papers on knowledge, I have denied that justification is necessary for knowing, but there I had in mind 'Cartesian' accounts of justification. On the account of justified belief suggested here, it is necessary for knowing, and closely related to it. (1979:1)

may be sufficient for the attribution of *practical* knowledge. It seems that it is not sufficient for the attribution of *theoretical* knowledge.

Contemporary epistemology features a debate between proponents of two radically distinct concepts of epistemic justification, whose differences seem irreconcilable.[45] On the one hand, internalists about justification seem to hold that, broadly speaking, to be justified in believing that *p*, a thinker has to have –or has to be able to have- conscious access to a reason that justifies the belief that *p*. If one cannot consciously appeal to a reason that justifies one's belief, that person is not epistemically justified in endorsing it.[46] Laurence BonJour (2002) writes:

> The basic rationale is that what justifies a person's beliefs must be something that is accessible or available to him or her, that something to which I have no access cannot give me a reason for thinking that one of my beliefs is true (though it might conceivably provide such a reason for another person viewing me from the outside). (2002: 222)

On the other hand, externalists about epistemic justification seem to reject the idea that what justifies a belief has to be cognitively accessible to the thinker who entertains it. We do not have to be able to appeal to what justifies our beliefs, in order for them to count as justified. Reliabilism, perhaps the leading form of externalism about justification, amounts to the position that in order to be justified in believing that *p*, the

[45] Tyler Burge (1993) drops talk about justification and maintains that the central concept is what he calls 'entitlement'. This might initially give the impression that Burge rejects the idea of justification as important. Yet, as Albert Casullo claims: "Entitlement is alleged to be a distinctive type of positive epistemic support that is different from justification. Both are species of warrant. Justification is the internalist form of warrant; entitlement is externalist." (2015: 2716) The basic idea of *epistemic support* remains intact. The difference concerns whether access to such support is essential for justification or not.

[46] Accessibilist internalists hold that in order to be justified in believing that p, a thinker must have access, or must be able to access that which justifies the belief that p. Mentalist internalists hold that the nature of that which justifies a belief is internal to the mind, such as beliefs.

belief that *p* has to have arisen by a belief-forming mechanism that is reliable and whose reliability consists in its successful role in producing true beliefs (Goldman 1979). Whereas the internalist requires the ability to appeal to a reason that justifies a belief, the externalist does not; the nature of the origin of the belief matters more than the ability to back it up with reasons.[47]

A debate seems to presuppose some common ground. Assuming that we find some common ground in our idea of justification, it is rational to believe that an agreement may be reached as to what beliefs are justified or not. We find ourselves in agreement as to what constitutes a good tool for a task and there is reasonable hope that if we also agree in the formulation of the task to be achieved, as well as our starting points, we can meaningfully argue as to whether particular moves or patterns of actions can get us from where we are to where we want to go. But if there is no common ground in how we conceive of justification, it seems that our chances for a fruitful debate and meaningful exchange of ideas are significantly diminished, if not nullified. Not only do we not agree on what tool should be used for the task; we arguably conceive of the task of philosophy and the theory of knowledge in radically distinct ways.

In what follows, I assume a broadly internalist conception of epistemic justification. I will argue for a version of it, according to which, to be justified in believing that p, one must have or must be able to have access to reasons that justify belief in p, except for cases where p is apprehended directly; e.g. in the case of self-evident propositions like the proposition that contradictions are impossible. Consequently, to be a priori justified in believing that p, one must have access to a

[47] Here, I refer to the reliabilist version of externalism which seems to be the dominant version of that doctrine.

priori reasons that justify belief in p, or apprehend that p directly[48]/non-inferentially[49]. In the first case, access to the reasons that justify the belief that p via a priori –non-sensory- means, constitutes an a priori reason for believing that p. In the second case, the fact that p is rationally apprehended constitutes a direct a priori reason to accept it as true.[50] For instance, to be *a priori* justified in believing that a particular conclusion follows *given some specific premises*, one must be a priori justified in thinking that the principle in question is valid; given these premises, the conclusion *must* follow, independently of whether one is a priori justified in holding such premises to be true or not.

A.2 ***Experience, sense-experience and temporal priority***

Traditionally, the concept of a priori justification refers to justification that is independent of sense-experience.[51] It is essentially contrasted with the concept of a posteriori justification.[52] A posteriori justification is empirical justification that derives

[48] The claim that we can have direct/non-inferential justification in believing that p is true is entirely compatible with an inferentialist account of our understanding of the meaning of p. We can maintain that understanding what p means is a necessary but not sufficient condition for knowing that p is true.

[49] The idea of immediate justification is predominantly appealed to as a means of blocking an infinite regress of justifications of particular knowledge-claims. An account of circular a priori justification, focused on our knowledge of laws of logic, is given by Paul Boghossian (2000) who dismisses the idea of justification by rational intuition as mysterious.

[50] When referring to an extra-sensory faculty of pure reason, the notions 'apprehension', 'insight' and 'intuition' will be used interchangeably.

[51] A major point in this chapter is that describing a priori justification as justification independent of experience, rather than independent of sense-experience, is a seriously misleading move because it rests on the mistaken idea that participants in that debate agree on what is involved in experiencing the world.

[52] The a priori/a posteriori distinction is not to be conflated with the necessity/contingency distinction or the analytic/synthetic distinction. Most philosophers, particularly those of a logical empiricist persuasion tended to conflate these three distinctions and claim that what we can know a priori is necessary and analytic, and what we know a posteriori is contingent and synthetic. Kripke's *Naming and Necessity* (1980) is generally

entirely from what we have sense experience of.[53] To claim that a thinker is a posteriori justified in believing something to be the case, is to claim that she is justified in believing it on grounds of particular sense experiences. One does not know that it is raining outside prior to having sense experiences of one's surrounding environment. The proposition that 'it is raining outside' is a proposition which is verified or falsified, empirically. One's belief in it is justified by appeal to one's sense experience, which verifies or falsifies the proposition in question. The notion of a priori justification concerns propositions we are justified in believing *prior* to a particular experience; either in advance of particular empirical data –in the sense that we can be a priori justified in believing that we won't see a flower which is red all over and green all over in the future- or irrespective of them –when we understand that no sense experience of flowers could ground the necessity we associate to that proposition.

The notion of the 'priority' of a priori justification with respect to sense experience needs to be qualified. Priority is always priority in terms of an order, such as the order of time, the order of being or the order of justification. In the context of our present discussion, 'priority' is seen both in terms of location in the order of time and location in the order of justification. For purposes of illustration, it is helpful to draw a distinction between unrestricted temporal priority and restricted temporal priority. If interpreted unrestrictedly, it means that we can literally be a priori justified in believing things before we have *any* sense experiences. This seems to be close to Plato's myth, according to which, the soul has perfect knowledge of the Forms before it enters the sensible world. Upon birth, such knowledge is almost entirely veiled. Restricted versions of it, treat it as a 'priority' with respect to *particular* sense-experiences, but not independently of the actuality of sense-experience. Kant, a

treated as the text which convinced the majority of analytic philosophers that these three distinctions are not to be conflated.

[53] In this context, knowledge gained by introspection and memory is usually associated with empirical knowledge, rather than a priori knowledge.

proponent of such a view, writes: "In the order of time, therefore, we have no knowledge antecedent to experience, and with experience all our knowledge begins. But though all our knowledge begins with experience, it does not follow that it all arises from experience." (1929[1787], B1) The difference lies in the following: unrestricted views of temporal priority hold that no sense-experience of the world is needed for someone to be in a position to be a priori justified in believing something to be the case. Restricted views of temporal 'priority' of a priori justification with respect to sense-experience claim that once we have sense-experiences of the world, the particular sense-experiences we have are irrelevant in or insufficient for grounding – which concerns the order of justification- certain things we can know or be a priori justified in believing, a priori. It is in this sense, not the Platonist one, that Kant writes: "we shall understand by a priori knowledge, not knowledge which is independent of this or that experience, but knowledge absolutely independent of all experience."[54] (ibid: A2/B3) In a nutshell, advocates of unrestricted views of the temporal priority of a priori justification to sense experience hold that we can be a priori justified in thinking something to be the case, even before we have sense-experiences of the world. Advocates of the restricted view hold that we can be in the position to be a priori justified in believing something to be the case, independently of particular sense-experiences but not independently of the fact that we have sense-experiences.

The traditional formulation of a priori justification seems initially puzzling. The reason for this is that it is formulated in predominantly negative terms.[55] Such

[54] Any invocation of Kant in such a discussion requires mentioning his background premises. Kant's ultimate point is that a priori knowledge is knowledge we acquire by concepts that are innate in the mind, and which cannot be acquired by experience. What they bring into experience is independent of any particular experience, since they are the forms of all possible experience. Furthermore, his reaction against traditional dogmatic metaphysics rests on his view that extending the reach of such concepts beyond the domain of experience, is a seriously misguided attempt.

[55] A somewhat positive characterization of the notion is given by Putnam who identified a statement that is a priori justified with a statement "which has the maximum degree of confirmation in all circumstances". (1983: 90) For a criticism of such a characterization, see Casullo (2003: ch.1).

negativity, in conjunction with the widespread dissatisfaction with aprioristic philosophies in the twentieth century, leads to a misleading presentation of a priori justification by certain influential contemporary discussions of the concept. Such a tendency muddies the waters because it seems to obfuscate the fact that not all participants to the debate define 'experience' in the same way. Different conceptions of experience give rise to different meanings of a priori justification. In short, if we do not focus on *sense*-experience rather than experience in our accounts of a priori justification, we are faced with the twofold risk of portraying almost everyone as a detractor of it, on the one hand, or of stipulating the empiricist ontology of the mind, on the other.

After a painstaking exploration of the meaning of the concept of the 'a priori', Albert Casullo (2003) claims that: "My primary conclusion is that a priori knowledge is knowledge whose justification is nonexperiential." (2003: 81) According to Casullo, belief in a proposition is a priori justified if it is justified independently of *experience*; i.e. if its source of justification lies somewhere other than what is involved in experience. Then, he proceeds to conceptualize 'experience' as a natural kind term. He writes:

> My suggestion is that "experience" be viewed as a putative natural kind term whose reference is fixed by local paradigms. The local paradigms are the cognitive processes associated with the five senses, which are identified in terms of such characteristics as providing information about the actual world, involving a causal relation to physical objects, and perhaps having a distinctive phenomenology. (ibid: 159)

This account seems to obfuscate the fact that 'experience' means something totally different to each participant in the debate. If we focus on 'experience' as what a priori justification has to be independent of, we are going to present an age-old

philosophical debate as a senseless one. It is good to remind ourselves that rationalists, Kantians and radical empiricists conceive of 'experience' in entirely different ways. Describing the rationalists' position on a priori knowledge, Blanshard writes:

> They held that it was a mere misreporting of fact to reduce the elements that composed experience to greens and reds, hots and colds, and other such data open to the senses. These were there, of course, but they were neither the only elements of experience nor the most important of them. There were also concepts or universals. Besides this given green, there was greenness, which all green had in common, and colour, which all colours had in common; and none of these universals were open to the senses. (2013[1962]: 252)

Clearly, this notion of 'experience' is one that in the eyes of the empiricist, involves a regretfully inflated ontology. For the rationalist, to have experience of the world involves the activity of rational apprehension of properties and their relations, in addition to having particular sensations. To say that in this sense, there is no knowledge whose source is not to be found in experience, is to say nothing that the rationalist will disagree with. For Kant as well, experience does not boil down to sense-experience. Kant begins his critique with the following claim: "Experience is, beyond all doubt, the first product to which our understanding gives rise, in working up the raw material of sensible impressions." (A1) For Kant, as for the rationalist, experience is not reducible to sense-experience. In opposition to both, we can find the radical empiricism of John Stuart Mill. Mill seems to hold that deep down all knowledge is knowledge of particulars. Not only does he claim that "all experience begins with individual cases" (1884[1843]: Book. II, Ch. I, par.3), but also that "the individual cases are all the evidence we possess" (ibid: Book .II, Ch. II, par.3).

Focus on 'sense-experience' rather than 'experience' allows us to reframe our understanding of the debate. We can reconstruct the contemporary debate about a priori justification and knowledge into what it has traditionally been conceived to be about;

a debate that is ultimately concerned with the nature of the mind and its faculties, as much as it is concerned with what we can know. All participants state their conception of 'experience' by making a direct reference to what sort of capacities are involved in experiencing the world. Rationalists claim that experience involves rational apprehension of properties, relations and propositions. Kantians hold that in addition to sense experience, experience involves the activity of the mind in structuring the data provided by sense experience, in a spatio-temporal order, as well as in imposing upon them, innate categories such as categories of 'quantity', 'quality', 'relation' and 'modality'. It seems that it is only for the empiricist that experience boils down to sense-experience.

It is evident that one of the issues that we are debating in the context of inquiry about a priori justification is whether our knowledge is derived from particular sensations alone or not. How we eventually conceptualize 'experience' is a consequence of our answer to this question, instead of something to be decided in advance, as Casullo seems to imply. It is prudent then, to reject Casullo's definition of a priori justification as justification from a non-experiential source and embrace 'independence from sense-experience' as the criterion for a priori justification and knowledge, bearing in mind that epistemically speaking, sense-experience is always meant to be sense-experience of *particular* matters of fact.

We can end this section with a small summary. Traditionally, the concept of a priori justification refers to justification that is independent of sense-experience. Its priority to sense experience can be understood in a temporally restricted, as well as an unrestricted sense. If interpreted in an unrestricted sense, such priority is understood as temporal priority to all sense-experience. If interpreted in a restricted sense, such priority can be temporally prior to particular sense-experiences, although not independent of having sense-experiences. If we fail to focus on 'sense-experience' rather than 'experience' in our definition of a priori justification, we run the risk of misrepresenting what is at stake in this debate. It is a debate that concerns the nature of

the mind's faculties as much as it concerns their epistemic reach. If we define 'experience' and then proceed to inquire about whether we have justification independent of it, we effectively presuppose a particular ontology of mind before we enter a debate which concerns the ontology of mind and the reach of its faculties.

Section B: ***Deficiency arguments: rationale and possible reactions***

Deficiency arguments against a position challenge its ability to account for knowledge of something that we hold to be true. If a position P cannot account for our knowledge of X and we think we have to account for our knowledge of X, this constitutes a prima facie reason to reject P. In the debate between rationalists and empiricists, the former have accused the latter as incapable of accounting for knowledge of what *must be*. They hold that we know that some things hold of necessity, such as the law of non-contradiction, and they also hold that the empiricist machinery can, at best, confirm such truths without being able to establish them. In section B.1, I present a version of a traditional rationalist deficiency argument against empiricism and in section B.2, I present the most famous reactions to it. I will claim that the main conceptions of a priori knowledge and justification –rationalism, Kantianism, logical empiricism and radical empiricism- can be understood essentially as reactions to it.

B.1 ***Rationale of deficiency arguments***

In this section, I present the rationale behind deficiency arguments that are employed by rationalist proponents of a priori knowledge and justification against its detractors. After mentioning the fundamental insight that drives them, I will present a formalized version of such an argument and discuss its premises. Deficiency arguments point out a particular deficiency in non-rationalist epistemologies and are intended to lead to the conclusion that rationalism can account for a desideratum they cannot account for. Casullo writes: "Deficiency arguments provide a basis for preferring

theories that embrace the a priori over their radical empiricist counterparts only if the former theories avoid the undesirable consequences themselves."[56] (2003: 101) Once we allow that we have a particular type of knowledge, like mathematical or logical, we can inquire as to whether the faculties that each theory posits can account for it or not. If not, we either have to reject the notion that we possess such knowledge or trace the source of its possession to rational insight or intuition. In a passage discussing the birth of experience, Blanshard writes:

> We must so construe the world we first live in as to make escape from it conceivable. It is true that we must not read into the earlier what comes later, but it is also true that we must see it in light of the latter, if our account is ever to reach the latter at all. (1939: 57)

Although this passage is concerned with conceptual experience and its emergence, it can also shed light on the rationale behind deficiency arguments. It constitutes a deficiency argument against the attempt to understand conceptual thought as a complex bottom-up construct from particular sensations and their association. We can first understand ourselves as sense-experiencing the world, but if such first experiences cannot account for the kind of knowledge that we can possess in the future, we have to conceive of ourselves as capable of possessing means of knowing that do not derive from such first experiences; we have to accept non-sensory or extra-sensory faculties of mind. If we are to accommodate such knowledge, we have to appeal to something other than that which we hold as fundamentally incapable of providing us with it. At best, we will come to an understanding of what it is and what it involves. At worst, we

[56] It has to be mentioned though, that it is not only radical empiricism that rationalists target with these arguments. A position that employs deficiency arguments against logical and radical empiricism, as well as Kantianism about a priori justification is developed by Blanshard (2013[1962]: ch. VI).

are going to have to be satisfied with a severely limited understanding of what it involves; a negative account is better than no account at all.[57]

Such arguments are, in a way, an expression of one of the basic rationalist criticisms of empiricism. This criticism amounts to the claim that we cannot account for our knowledge of what must be, in terms of what we know contingently to be. The empiricist's identification of mentality with bundles of particular sensations makes the mind a prison from which we cannot escape to knowledge of what is necessary. Anything we know by sense experience, is known as contingent and particular. But no amount of particular sensations can establish anything more than a probable truth. It may confirm a universal claim but it cannot establish it.

Foremost among the non-empiricists, rationalists hold that some of our knowledge is not contingent in nature. According to this line of thinking, if we are not to abandon our claim to any knowledge of what must be, we have to allow for extra-sensory means of grasping necessity. Otherwise, we are left with a deficient account of knowledge which eventually leads to skepticism. A formalized version of the traditional deficiency argument is the following:

(I) We have some knowledge of what *must* be.

(II) Knowledge of what must be, cannot be based on knowledge of what contingently is; knowledge of particulars.

(III) Knowledge gained by sense-experience is knowledge of what contingently is; knowledge of particulars.

(IV) Therefore, knowledge gained by sense-experience cannot ground knowledge of what must be; no amount of knowledge of particulars can ground anything more than a probable truth.

[57] Such a position can be found in Jerrold Katz (1998: 32-34), who holds the operation of such faculties to be mysterious, but that this does not justify the charge that the mysterious is mystical.

(V) By (I) and (IV) we conclude that not all knowledge that we have is based on sense-experience.

Premise (I) states that we have some knowledge of what holds necessarily. The necessity of that which is known consists in the impossibility of it being otherwise. Such a notion of what *must* be, isn't only confined in knowledge of necessity of a single proposition; it could also be focused on *relations* between propositions. The rationalist who claims that given certain premises, we *must* infer the conclusion, doesn't imply that the conclusion is itself necessary. In this sense, she refers to the manner in which the conclusion follows from the premises; it is necessary *given these premises*. For instance, Descartes's *cogito* amounts to the claim that given the fact that he thinks, it *must* be the case that he also exists. But that doesn't mean that he takes it to be the case that he *necessarily* exists. That he exists, follows necessarily from the fact that he thinks. But it is not a metaphysical necessity that he thinks. The world could have taken a wrong turn by unfolding in a way in which Descartes wouldn't have been born.

Premise (II) amounts to the claim that the attempt to give an account of how we possess knowledge of what must be, requires us to appeal to something more than what we know to hold in particular cases. What is claimed to hold in *all* cases of a particular type, say deductive inferences, cannot be grounded on our knowledge of what holds in some specific cases of that very type. In *The Problems of Philosophy*, Bertrand Russell writes:

> In all our knowledge of general principles, what actually happens is that first of all we realize some particular application of the principle, and then we realize that the particularity is irrelevant, and that there is a generality which may equally be affirmed. (…) The same thing happens with logical principles. (1998[1912]: 39)

Although they may confirm it, individual cases in which the application of a principle is truth-preserving or truth-conducive are not sufficient to establish its validity. Our knowledge of what holds universally cannot be based on our knowledge of what holds in particular cases. What we claim to know as holding in *some* cases cannot be the ground upon which we claim that it holds in *all* cases, let alone what must hold in all cases. As Kant writes: "Experience teaches us that a thing is so and so, but not that it cannot be otherwise." (1929[1787], B3)

Premise (III) amounts to the claim that an empiricist epistemology and its corresponding ontology of mind allows us access only to particulars. If the mind consists in bundles of sensations, all we can access are the objects of our particular sensations. Such objects are not universals, but even if they were, we couldn't know that they were universals just by sensing them. The problem of induction rests precisely on this idea. A finite amount of observations about F's cannot ground the claim that all F's are G's, let alone the claim that F's are necessarily G's. That we have observed a thousand times that F's are G's doesn't suffice to ground the claim that it must be the case that they are so. Nothing secures us from encountering an F that isn't a G.

Premise (IV) follows by (II) and (III). If knowledge gained by sense experience is knowledge of what contingently is, then it cannot ground knowledge of what *must* be, whether in the case of facts or in the case of relations. To know that a fact must be the case cannot be grounded on the finite cases in which we have had sense experiences of it. The same applies for relations. To know that given certain premises a conclusion *must* follow, is to know that a principle of inference: i) yields a particular conclusion *given some premises*, and ii) is valid, in the sense that it is truth-preserving. To know that the conclusion must follow is to know that the principle is valid. If we try to ground our knowledge of the validity of an inferential principle on our knowledge of particular cases in which the use of it gave us true conclusions, we will never be able to claim that such a principle is valid. All we will be able to say is that *in specific cases*, the principle of inference we used was truth-preserving or truth-conducive. But this doesn't

allow us to infer anything about other cases, other than *probable* falsity, truth or validity.

Conclusion (V) follows from (I) and (IV). If we claim to have knowledge of what *must* be, whether this applies to facts or relations in all or some domains, then such knowledge cannot be based on knowledge that is gained from sensory experience; for such kind of experience can only give us knowledge of what holds contingently.

B.2 ***Possible reactions***

To understand these reactions better we need to understand ways in which the topic relates to epistemological skepticism. One can engage into the debate as someone who considers skepticism as a live option. The alternative is to enter the debate as a non-skeptic and hold that one of the main defects of a theory is that it implies skepticism. I will not address those who embrace the former attitude. I do not think this is possible. Someone who is a committed skeptic will not be convinced by any consideration for a priori knowledge. Perhaps they will not be convinced by any consideration against it, either. Equally, someone who thinks that knowledge may be possessed will find epistemological skepticism rather unconvincing.

By adopting a non-skeptic attitude to the topic, we can examine it in the context of its question. Both rationalism and empiricism, as well as other non-skeptic theories are broadly concerned with *how* we know what we know. This question is only available to the non-skeptic because it makes no sense to ask: "How can I know that which I cannot know?" or even "How can I know that which I don't know if I know?" The how-question, makes sense only after a positive answer is given to the question: "Can I know anything?" So, I will proceed to examine the theories below by presupposing that all their defenders agree that skepticism is to be avoided. I will discuss rationalism, Kantianism, logical empiricism and radical empiricism. The limited space of the thesis does not allow me to discuss all four views to a satisfactory

degree. So, in section C, I will focus on radical empiricism as the strongest theory against premise I) of our deficiency argument.

In many ways, the epistemological position one embraces in the debate concerning a priori knowledge and justification is a response to such deficiency arguments. The main strategies are:

A) Embrace (I) and (V) and hold that rationalism follows from that. (Rationalism)
B) Reject (I) (Radical empiricism)
C) Embrace (I) and (V) and reject that rationalism follows from that. (Kantianism or logical empiricism)
D) Reject (II) and (III) altogether.

The first strategy is the traditional rationalist one. We accept that we have knowledge of what must be and we also hold that such knowledge cannot be based on sense-experience. But in that case, what is it based on? The traditional rationalist holds that it is based on an extra-sensory faculty which is referred to as the faculty of rational apprehension, insight or intuition. The exercise of such a faculty results in a form of cognition which gives us access to aspects of reality that cannot be accessed by the senses.

The second strategy is the radical empiricist one. The radical empiricist denies that we can have any knowledge of what *must* be, whether in the case of facts or relations. The two main versions of radical empiricism are: a) Mill's inductivism, and b) Quine's holism. Mill holds that the rationalist commits the fallacy of conflating the limits of conceivability, which springs from the faculty of imagination, with the limits of possibility. Every statement, even statements of mathematical truths, gain their confirmation on the basis of sense-experience of particulars. What is not legitimate is to infer the necessity of something from its confirmation, however recurrent it is. Quine holds that: "Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of

the kind called logical laws." (1953: 43) There is nothing in principle that disallows us to revise a proposition which the rationalist holds to be necessarily true. Consequently, "no statement is immune to revision." (ibid: 43)

The third strategy is to allow for knowledge of what must be, but in a distinctively non-rationalist sense. It is the approach that Kantians and logical empiricists take. In their attempt to find a middle way between dogmatic rationalism and skepticism, Kant and advocates of his approach claim that in fact, we can and do have some knowledge of what must be. But unlike what the rationalists think, such knowledge is not knowledge of truths about the world. Statements of what must be do not reflect reality in itself but the phenomena; reality as our minds construct it.[58] Logical empiricists hold that such statements are reflections of features of our linguistic practices or empty tautologies. They are either expressions of certain syntactic and semantic rules we adopt or empty of all content.

The fourth strategy is the Kripkean one and it advocates the existence of necessary statements that are known a posteriori. According to Kripke (1980) some identity statements of proper names, like 'Hesperus = Phosphorus', are necessary truths which can only be known a posteriori. This line of thinking concerns identity statements. Kripke doesn't deny that we can have a priori knowledge. He denies the essence of logical empiricism; that the a priori/ a posteriori distinction is coextensive with the necessary/ contingent distinction and the analytic/ synthetic distinction. To the extent that such a strategy concerns identity statements, it poses a threat to the deficiency argument to the extent that its conclusion is itself a statement of identity. Since I do not intend to establish such a conclusion, Kripke's point will not be dealt with in this thesis. So I will proceed to describe the positions who adopt strategies A),

[58] This rests on Kant's distinction between the phenomena and the noumena, which roughly corresponds to the distinction between the world as we construct it and reality in-itself.

B) and C) and in section C, I will argue against radical empiricism about a priori knowledge and justification.

B.2.1 ***Rationalism***

Rationalism's core consists in the idea that pure reason can give us knowledge of aspects of reality that are not accessed by the senses or the imagination.[59] Pure reason is a faculty that allows us to grasp self-evident truths, necessary facts, as well as necessary relations. Its power consists in revealing to us something about the world's structure, which holds of necessity.

Despite its essential core, rationalism comes in many varieties. A sharp distinction can be drawn between traditional rationalists such as Descartes, Spinoza and Leibniz on the one hand and contemporary moderate ones on the other (e.g. BonJour 1985, 1998, 2002; Bealer 1996, 1999; Katz 1998; Peacocke 2004). Such a distinction is drawn in terms of their respective treatments of the reach of 'rational insights'. It is held that the former conceive of them as infallible and empirically indefeasible, whereas the latter conceive of them as fallible and empirically defeasible.[60] Such a distinction should not blind us to the fact that among the non-

[59] As Charlie Huenemann describes, among other things, rationalism involves the "belief in privileged cognitive machinery (logical processes) in human beings, which allows for a distinction between the intellect and the imagination." (2008: 7) Such a faculty has been seen as an impediment to naturalistic accounts of the mind. In describing the ramifications of Hume's conception of the mind, Don Garrett writes: "By rejecting a separate faculty of "intellect" and subsuming its functions within the inclusive imagination, he places metaphysical theorizing under substantial empiricist constraints and at the same time facilitates a naturalistic explanation of the mind's many cognitive operations." (2015: 113-114)

[60] If this is accurate, it seems to me that there is room for a third kind of rationalism which holds fallibility to be a matter of judgment; neither a matter of the senses, nor a matter of rational insight. Just like the senses are a window to the sensible aspect of the world, so rational insight can be a window to another aspect of it. Fallibility is more plausibly a matter of what one judges in light of what one sees through these windows. Nevertheless, this might also be the position that some traditional modern rationalists have taken in claiming that, rather than fallible, the sense are unreliable guides to judgment.

moderates, some are more moderate than others. What I mean by this is that traditional rationalists differ with respect to the extent that they take pure reason to give us knowledge of the world; the extent to which they think that there are necessary connections in the world, as well as the nature of the necessity of such connections, i.e. whether it is moral, causal, logical or mathematical.[61] For instance, Spinoza thinks that: "It is of the nature of reason to regard things as necessary, not as contingent." (1996[1677]: 59, Part II, prop.44) The extent to which reason is claimed to give us knowledge of the world reflects the extent to which necessary connections are admitted to exist in their respective domains, as well as the number of such domains. One can hold that every connection in every domain is a necessary one whereas another can hold that some connections, like logical connections are necessary, whereas causal ones are not or vice versa.

Rationalists essentially agree with the deficiency argument that is formulated in the beginning of section B. As mentioned above, they differ to the extent that they allow reason to give us knowledge of what must be. It comes as no surprise that the more moderate a rationalist one is, the less extensive she or he will conceive of the domain of our knowledge of what must be. Spinoza (1996[1677]) on the one hand, holds a strong rationalism which leads to necessitarianism, the idea that everything that is, must be as it is and not another way. On the other hand, BonJour (1985, 1998, 2002) is a moderate rationalist who thinks that what we know by rational insight is fallible, yet we can be a priori justified in thinking that some principles of inference *must* be valid, if radical forms of skepticism are to be avoided.

The metaphysically libertarian rationalism I advocate in this thesis is a form of rationalism that occupies a place between extreme rationalism and moderate

[61] Essentially, this reflects what Schopenhauer calls the 'fourfold root of the principle of sufficient reason' in the dissertation he names after it (1974[1813]). The most extreme rationalism would be a rationalism which would hold that the principle of sufficient reason is unrestrictedly valid in all four dimensions. For a view that claims Spinoza to be the most extreme modern rationalist, see Michael Della Rocca (2008: ch. 7).

rationalism. In contrast to the former, it holds necessary connections to exist in the logical and mathematical domains but not in the causal and moral domains. In contrast to the latter, it holds that some rational insights, such as the insight that contradictions cannot exist –ontologically speaking, not in the sense of contradictory descriptions[62]- are non-defeasible, leaving it an open question whether other rational insights are or not. Acceptance of causal relations as necessary relations would threaten libertarian free will because it would portray causal relations as relations of causal necessitation. It will be argued in Ch. IV/B.2.1 that free will is not compatible with causal necessitation of action. Unfortunately, a discussion of the viability of this position will have to be dealt with more sufficiently in further research.

B.2.2 ***Kantianism***

Kantians agree with the deficiency argument stated above but proceed to give a wholly different interpretation of why the argument is valid and sound. Such a different conception of the argument is based on Kant's radical distinction between the thing-in-itself and the phenomena and thus, avoids the rationalist conclusion. According to Kantians, we do possess knowledge of what must be, yet such knowledge is not knowledge of reality as it is. Rather, they hold it to be knowledge of the structures that our minds impose upon the raw material of sensation.[63] Unlike logical empiricists, whose position will be discussed in the section underneath, Kantians hold that a priori knowledge can be synthetic. Yet, Kant's examples of synthetic a priori propositions

[62] A recent treatment of Aristotle's law of non-contradiction as a metaphysical principle, is given by Tuomas Tahko (2009), who claims that a metaphysical treatment of the law of non-contradiction escapes the semantic paradoxes that dialetheists, like Graham Priest (2006) focus on.

[63] For a neo-Kantian position, according to which logic and linguistic competence are neither reducible to natural facts, nor to evolutionary facts, see Robert Hanna (2007: 19-21, 50)

are confined to the phenomena alone; not reality as it is in itself. In a passage where he criticizes Fichte's understanding of Kant, Arthur Schopenhauer[64] writes:

> If he had penetrated only to some extent the meaning of the *Critique of Pure Reason*, the book that made him a philosopher, he would have understood that its principal teaching was in spirit as follows. The principle of sufficient reason[65] is not, as all scholastic philosophers assert, a *veritas aeterna;* in other words, it does not possess an unconditional validity before, outside, and above the world, but only a relative and conditioned one, valid only in the phenomenon. (1969[1819]: 32, par.7)

For someone of a Kantian persuasion, knowledge of what must be is knowledge of the impositions of the mind on the data of sense-experience. Unlike the rationalist who believes that such knowledge can show us something about reality in itself, Kantians hold that reality as it is, the thing in itself, is strictly speaking unknowable. It is an

[64] Some interpretations hold that Schopenhauer believes that the blind, irrational will is Kant's thing-in-itself, whereas others hold that he holds will to be another aspect of how things appear to us. For an interpretation that holds Schopenhauer to have changed his mind from thinking initially that the blind irrational will is the thing-in-itself to thinking that the thing in-itself is unknowable, see Julian Young (2005).

[65] This quote by Schopenhauer is a good example of the difference between a strongly rationalist and a Kantian treatment of the principle of sufficient reason. Schopenhauer (ibid: 81) accepts that two things stand outside its province: the principle of sufficient reason itself, and the thing-in-itself. In contrast to a rationalist like Spinoza who arguably took the principle as being self-evident, Schopenhauer does not. He continues:

> Proofs cannot be its foundation, for these deduce unknown principles from others that are known; but to it everything is equally unknown and strange. There can be no principle in consequence of which the world with all its phenomena would first of all exist; therefore, it is not possible, as Spinoza wished, to deduce a philosophy that demonstrates *ex firmis principiis.* (ibid: 82)

instance of the fallacy which philosophers commit when they take the concepts of pure understanding and apply them outside the domain of sensible intuitions.

B.2.3 ***Logical Empiricism***

The logical empiricist strategy with respect to the deficiency argument against non-rationalist epistemologies is multi-faceted. Although its advocates may concede that we have some knowledge of what must be, they reconceptualise such knowledge as analytic and a priori or as pseudo-knowledge. In opposition to what the rationalist thinks, such statements do not reflect any substantive truths about reality. Logical empiricists conceive of necessary statements in at least two ways.[66] They can be conceived of as: i) reports of semantic and/or syntactic conventions that are true in virtue of meaning,[67] or ii) non-factual expressions of semantic and syntactic conventions.[68]

According to the first conception, necessary statements are statements about how language is used. To say that 'all houses have ceilings' is to make a claim about the conventions that are used in describing something as a house. To know that it is necessarily the case that something is a house and it also has a ceiling amounts to knowing something about how we use language; it does not amount to knowing

[66] A third can perhaps be found in Wittgenstein's *Tractatus* (1961[1922]) in his claim that necessary statements or statements of logic are tautologies that are empty of content. Peter Hacker (1996: 45-50) claims that this is a distinct view from logical conventionalism and that logical empiricists essentially misunderstood Wittgenstein's point who treated logic as transcendental.

[67] Ayer (1946). It is said that this approach is deeply defective because it makes necessary statements, which are conceived as analytic and a priori by logical empiricists, to boil down to empirical generalizations about language use (Pap,1958: 163-170, BonJour, 1998: 52) . Pap claims that Ayer reformulated his position "in the introduction to the second edition of his positivist manifesto" (ibid: 163) In that sense, Ayer seems to occupy a position closer to the second conception of necessary statements.

[68] Perhaps the most influential view of this sort can be found in Carnap (1937).

something about the world. Namely, if something has walls but no roof, we usually don't call it a house.

According to the second conception, necessary statements that pertain to reflect reality are pseudo-statements. They cannot literally be true or false because they do not have factual content. They are the expressions of conventions. In the *Logical Syntax of Language*, Carnap writes:

> *In logic, there are no morals*. Everyone is at liberty to build up his own logic, i.e., his own formal language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (1937: 52)

Carnap treats necessary statements as axioms of conventional symbolic systems or as their inferential rules. Axioms are statements that are true *within* a particular system of conventions as opposed to being true in a metaphysical sense. Metaphysical questions about what exists and what is true of the real world are to de dismissed as pseudo-statements. Questions make sense only when they are internal to a conventional system of signs; not when they are external to it. This represents Carnap's solution to the intractable problem of the clash of philosophical intuitions.[69] No one is right in claiming their alleged intuitions to be true because the conventional framework one adopts, in which they are axioms, cannot be called true of the world or false of the world without making a metaphysical claim, *external* to that system.[70] In an important sense, truth does not transcend one's arbitrarily chosen framework.

[69] For the view that the debate that Carnap had in mind was that between intuitionists, formalists and logicists in the philosophy of mathematics, see Cory Juhl & Eric Loomis (2010: 34-36).

[70] Juhl and Loomis claim that this represents one of the key disagreements between Carnap and Quine. They write: "Carnapian tolerance of linguistic forms is not an interesting possibility, Quine thought. By starting with our accepted theory of physical reality, we are starting with its ultimate categories of reality." (2010: 134)

Focusing on a framework's rules of inference, as opposed to its axioms or theorems, gives an interesting twist to this rich conception. It can be seen as parallel to the non-cognitivist treatment of ethical statements. Just like the non-cognitivist analyses the claim that 'murder is wrong' in terms of the expression of an imperative like: 'Don't murder', so the logical empiricist may analyse the claim that "Contradictions are false" to the exhortative function of the expression of an adopted rule of inquiry such as: 'Do not accept contradictory predicates in descriptions of things at each particular moment.'[71]

B.2.4 ***Radical Empiricism***

Radical empiricism comes in two main varieties: inductivism and holism. John Stuart Mill (1884[1843]) is the most known radical empiricist of the former type, whereas Quine (1953, 1969) seems to be the foremost representative of the latter. Mill holds all aspects of thinking to be modifications of induction, which for him, consists essentially in inference from particulars to particulars. Every "train of reasoning is but bringing many inductions to bear upon the same subject of inquiry, and drawing a case within one induction by means of another;" (1884[1843]: 147, Book II: Ch.V, par.1) Deductive reasoning is also a covert form of induction. The deductive inference that moves from the mortality of all human beings to the mortality of Socrates is deep down a form of induction; our acceptance of the premise "all human beings are mortal" is based on our finite observations of people. Mill's inductivism extends also to mathematical and logical laws and truths. We know that '6 + 2= 8' because we have repeatedly observed the sum of six and two units of what we count and find it to be eight. But aren't mathematical statements certain as well as necessary? Mill finds such

[71] It seems that unless qualified, imperatives of this sort are incompatible with Carnapian tolerance. Yet, it is not entirely unreasonable to conceive of Carnap's treatment of a system –or language as he preferred- as a system of hypothetical imperatives. It seems that his point would amount to the claim: 'If you accept language L with rule of inference R, proceed to draw inferences in light of R.'

certainty to be illusory. We also claim to know the law of non-contradiction because we haven't encountered any true contradiction. That we cannot conceive of them, doesn't imply that they *don't* exist, let alone that they *cannot* exist. Mill writes:

> Now I cannot but wonder that so much stress should be laid on the circumstances of inconceivableness, when there is ample experience that our capacity or incapacity of conceiving a thing has very little to do with the possibility of the thing in itself, but is in truth very much an affair of accident, and depends on the past history and habits of our own minds. (ibid: 157, Book II, Ch.V, par. 6)

An inductivist radical empiricist would definitely deny premise (I) of our deficiency argument. Any form of knowledge of what must be, is akin to absolute certainty, which cannot be possessed by fallible beings. Mill holds that any such affirmation of absolute certainty is inherently dogmatic. In *On Liberty*, he writes: "There is the greatest of difference between presuming an opinion to be true, because, with every opportunity for contesting it, it has not been refuted, and assuming its truth for the purpose of not permitting its refutation." (2015[1859]: 21, Ch. II) A statement that we think is necessarily true is just a statement that is confirmed by experience all the time for either of two reasons. First, it might be that it gets confirmed by experience because of the contingent make up of our minds and brains. Second, it might be that we become habituated in thinking that it is impossible that it could be otherwise and we constantly interpret our experiences by holding this statement as immune to refutation. But we cannot conflate the limits of our imagination with the limits of possibility. We cannot claim that the inconceivable or the unimaginable is in fact impossible. Hence, to claim that a contradiction is impossible is blind dogmatism. The world could literally be full of true contradictions. The fact that we cannot conceive of them or imagine them does not mean that they do not, let alone cannot, exist.

In contrast to Mill's inductivist radical empiricism, Quine's version is holistic. His holism is a direct outcome of his rejection of the analytic/synthetic distinction and of the dogma of reductionism, which consists in "the belief that each meaningful statement is equivalent to some logical construct upon terms which refer to immediate experience." (1953: 20) His rejection of the analytic/synthetic distinction and reductionism gives rise to a holistic approach to belief revision which dispenses with the notion of a foundational empirical belief. Our belief system forms an interconnected nexus, only a part of which is in direct contact with sense experience. Whereas Mill advocates a radical inductivism in which every large generalisation is in threat of empirical refutation in terms of the data of sense experience, Quine holds that there is no statement that cannot be revised. This seems to have two interesting ramifications. First, there can be no a priori known statements. Second, whenever we are faced with a conflict between some of our general statements and the data of sense experience, we can always choose to pronounce the data of sense-experience illusory, instead of revising our belief in the general statements. Due to this feature of Quine's holism, his version of radical empiricism is taken to fair better than Mill's inductivism, in explaining the felt certainty of logical and mathematical truths.

Holistic radical empiricists claim that no statement is immune to revision. So, there can be no knowledge of what must be because we can come to understand that what we held at one time to be necessarily true, can be revised.[72] And this leads also to a radical form of skepticism with respect to necessary truths. If all statements are revisable, then every statement that states an *alleged* necessary truth can be revised.

[72] The notion of 'revision' is used in at least two ways. The first is strictly psychological and concerns the fact that we may just reject something as being the case. The second is normative in that although it has a psychological component, it also makes essential reference to whether the revision is justified or not. In the first sense, to claim that every belief is revisable amounts to the claim that it is possible for someone to cease to endorse a belief and embrace its opposite. The 'revisability thesis' in this sense does not present a problem to the advocate of a priori justification. To present a problem, it has to be meant in a normative way, which implies that it amounts to the claim that we can be justified in revising any belief.

Even the law of non-contradiction can be revised; the fact that we find it inconceivable for it to be false does not imply that it is impossible for it not to hold.

Section C: ***The case for a priori justification***

In this section, I defend rationalism about a priori knowledge and justification. In section C.1, I argue that a priori justification is indispensable for reasoning, avoiding skepticism as well as for inquiry in general. These arguments are not intended as an argument against the skeptic. Rather, they are intended as arguments against non-skeptics who aim to dispense with a priori knowledge and justification, altogether. The main arguments against rationalism about the a priori deny: a) our knowledge of necessary facts, propositions and relations, b) the superiority of rationalism in accounting for our knowledge of necessary relations, c) the intelligibility of rational intuition, and d) that the necessary facts or relations we know of give us knowledge of the world, as opposed to knowledge of our minds or linguistic conventions. I will defend rationalism about the a priori by answering the first three criticisms. As mentioned above, the fourth criticism lies outside the scope of this thesis and will have to be dealt with in future research.

The central premise in the deficiency argument discussed in the previous section is the first one. According to premise (I), we have some knowledge of what *must* be. I will argue that if we accept premise (I) and defend it against criticism a), we can also answer criticisms b) and c), since the latter criticisms rest on the former for their strength. In order to defend premise (I), I will give an argument according to which a priori justification is indispensable for reasoning, maintaining the possibility of knowledge as well as for meaningful thought, in general.

C.1 *A priori justification and the law of non-contradiction*

In this section, I argue that a rationalist account of our knowledge of the metaphysical law of non-contradiction is preferable to a radical empiricist account of it. In this sense, we do have some knowledge of what must be. It must be the case that a contradiction is false and if a theory entails a contradiction, it must be the case that this theory is false. I will claim that radical empiricism leads us to a form of skepticism and thus, cannot attack premise I) of our deficiency argument in section B, without leading us to a radical form of doubt about much more than knowledge of necessity. Namely, if we do not allow a rationalist account of our knowledge of the metaphysical law of non-contradiction, we will be led to the unfortunate ramification of losing sight of truth. To see this more clearly, we have to think of how the fundamental functions of an epistemic theory would be distorted if we were not justified in thinking that contradictions are false, because impossible.

Every epistemic theory requires constitutive principles. Some of them notify us when a belief rejection or revision is required.[73] Others, such as the principle of inference to the best explanation, instruct us about what specific revision is to be made when a revision is called for. In both their inductivist and holistic versions, radical empiricists use the law of non-contradiction as the principle that instructs us when a rejection of one or more of our beliefs has to take place.[74] By rendering our knowledge of the law of non-contradiction contingent, they make it impossible for us to know whether that which we reject as false, because contradictory, is in fact false or not. Since the application of the principles that tell us which specific revision to make, rests on the principle that tells us when a rejection has to take place, we do not know whether

[73] For an argument according to which, the principle of non-contradiction plays the role of the principle that tells us when a belief revision has to take place, in Quine's epistemology, see Katz (1998: 72-74)

[74] This might become more evident in the case of Quine's defense of classical logic against deviant alternatives, where he writes: "Here, evidently, is the deviant logician's predicament: when he tries to deny the doctrine he only changes the subject." (1986: 81)

applying such principles in such cases, leads us closer towards truth or further away from it. And this amounts to a radical form of skepticism. Furthermore, radical empiricism makes our practice of error ascriptions entirely problematic. The very idea of fallibility seems to require a commitment to the truth of the law of non-contradiction because fallibility is fallibility in light of normative principles, which are internally connected with satisfaction conditions. As such, fallibility becomes possible when failing to meet a standard is treated as inconsistent with meeting it. By making our knowledge of the law of non-contradiction problematic, radical empiricism renders the former problematic as well. In section C.1.1, I will describe the first problem. In section C.1.2, I will describe the second and in C.1.3, I will answer some familiar objections against the rationalist commitment to rational insight.

C.1.1 ***Contradictions and radical empiricism***

Radical empiricism involves the idea that what we can conceive is not the criterion of what is possible. What we claim to be conceivable or inconceivable, is not an indication of what is possible or impossible. Rather, it represents a contingent fact about human cognition, whether structural, habitual or habitual because structural. We just possess this sort of mind with these sorts of faculties which allow us to think of anything within our cognitive limits. This can be called Mill's thesis:

Mill's thesis: The fact that we cannot conceive of X implies neither that X does not exist nor that it cannot exist.

Mill's thesis is essential for any epistemology that does not involve the claim that our minds have the power to grasp what *must* be the case. Our cognition involves the sensation of particulars and no universal truth may be established from them.

According to Mill's thesis, we do not know whether the metaphysical law of non-contradiction holds. To maintain that claims about contradictions existing are false because ontological contradictions are impossible and they are impossible because we cannot conceive of them, is to conflate the limits of our conceivability with the limits of possibility. Rather than revealing a truth about the world, our embracing the law of non-contradiction is a contingent feature of how our minds perceive the world. And because it is a feature of our minds, nothing prevents us from employing it in inquiry.[75] In fact, radical empiricism does employ the law of non-contradiction as a constitutive principle of inquiry. What it does not involve is a commitment to conceiving of it as a truth about reality, as opposed to just reflecting how our minds perceive the world.[76]

As mentioned in the beginning of this section, epistemic theories require constitutive principles. Their number and function varies according to the theory and its aspirations. Whether they are served by one or more principles, two functions stand out. First, a principle must specify when a belief rejection or revision has to take place. Second, given circumstances which call for a belief rejection or revision, a principle must specify which specific belief rejection or revision has to be made.

Inductivist and holistic radical empiricists agree on the principle that fulfils the first function but disagree on how to fulfil the second one. Both accept the law of non-contradiction as the principle that fulfils the first function. They accept that the circumstances which call for a belief revision or rejection to take place are circumstances in which our beliefs contradict one another. Their difference concerns the second function that is specified in the paragraph above. Inductivists employ a bottom-up approach in revising beliefs, where particular observations are strong enough to dislodge generalizations. The observation of an aircraft carrier is sufficient to overthrow the generalization that all ships are made of wood. Holistic radical

[75] Quine's (1953) account of the status of logical laws, in terms of their centrality to our web of beliefs, seems to be a metaphor that implies that they are central because our minds work that way, not because they are true.

[76] In this anti-rationalist claim, it seems that radical empiricists and Kantians are on the same boat.

empiricists allow for bottom-up revisions and for top-down revisions as well. If circumstances call for it, we might as well dismiss the observation of an aircraft carrier as an illusion. Of course, repeated exposure to aircraft carriers will make our lives significantly more difficult in upholding that statement, but in principle, there is nothing that stops us from doing so if we choose to follow holistic radical empiricism. But in both their respective cases, they agree that the circumstances which call for a rejection or revision of belief to take place are marked by the encountered contradiction between 'all ships are made of wood' and 'this ship is not made of wood'. Thus, they both agree that the principle which specifies when a belief revision or rejection has to take place is the law of non-contradiction. They disagree on what they consider a legitimate move in such cases.

Unfortunately, such a position leads to a form of skepticism. Assuming that we hold that when we claim to know that p is the case, we imply that we know p as opposed to not knowing it, Mill's thesis can cast radical doubt on most, if not all, of our knowledge claims. To see this, let us formalize the following argument:

I) Radical empiricists accept the law of non-contradiction as a principle that specifies when a belief revision has to take place, without being committed to its truth.

II) If the law of non-contradiction is not true of the world, then we cannot know whether that which we reject as false –because contradictory or because it entails a contradiction- is in fact false or not.

III) According to radical empiricism, we cannot know whether the law of non-contradiction is true of the world or not.

IV) Therefore, according to radical empiricism, we cannot know whether what we reject as false –because contradictory or because it entails a contradiction- is in fact false or not.

V) Epistemic principles that specify the particular belief revision that has to take place, are applied in circumstances which are themselves specified by the law of non-contradiction.

VI) Therefore, if we do not know whether the law of non-contradiction is true, we do not know whether applying such principles will lead us closer towards the truth or further away from it.

Premise I) claims that radical empiricists do accept the law of non-contradiction as the principle that specifies when a belief revision has to take place. Due to Mill's thesis, they think that they cannot establish its truth. It is this premise that the radical empiricist may contest. The radical empiricist might accept that as a matter of fact, she or he does employ the law of non-contradiction as a principle that fulfils this function, but this does not imply that they must use it. It may be the case that we use it, but in the future we might switch to another principle to fulfil the relevant function. This response will be countered in section C.1.2 with an argument that shows the centrality of the law of non-contradiction for error ascriptions, as well as its centrality in our conception of principles with satisfaction conditions.

Premise II) claims that the ramifications of Mill's thesis sever the ties between truth and the central function that every epistemic theory has to provide us with adequate means to fulfil. According to Mill's thesis, the fact that we cannot imagine X does not imply its falsity or its impossibility. Let us suppose that the structure of our minds is such that we cannot imagine a contradiction. Mill's thesis implies that it might be possible for such a contradiction to exist. The fact that we cannot imagine it does not imply its impossibility. For the sake of the argument, let us suppose that a contradiction is true. An object S has properties P and ~P. We cannot conceive of that, but a theory we have entails that S has properties P and ~P. Our employment of the law of non-contradiction will lead us to reject the belief that S has contradictory properties, as well as the theory that entails it. If that is the case, the central role that our epistemic principles have to play is rendered severely problematic. There is no connection

between truth and our main principle any longer. We may be mistaken about almost any proposition because of reasons that are not available to us, such as because they exceed the grasp of our cognitive capacities. For any claim that we make, even those whose falsity we cannot conceive of, we cannot claim that they are the case or not. And this leads us to a severe skepticism.

Premises III) and IV) represent what is entailed by Mill's thesis and an intermediate conclusion respectively. The first is a restatement of the radical empiricist position that we cannot know the law of non-contradiction to be true. The second is what follows from premises I), II) and III). It states that we do not in fact know whether the circumstances which we think of as meriting a belief revision, actually merit or not.

Premise V) and conclusion VI) involve an extension of the problem from the principles that fulfil the function of specifying us when a belief revision has to take place to the principles that specify what belief revision to make. The principles that instruct us which revision to make are applied in cases which are deemed appropriate for that by the principle that specifies when a belief revision has to take place. If the central principles of an epistemic theory that fulfil the first function, are not in any way connected to truth, then employing principles that fulfil the latter function will be disconnected from truth as well. There is no guarantee that employing them in our thinking leads us closer to truth or further away from it.

I consider this to be a strong defect of radical empiricism. It does deny premise I) of our argument in section B of this chapter, according to which, we have some knowledge of what must be. But the way it undermines it, leads us to a radical form of skepticism about almost everything. It entails that all beliefs, however self-evident we may take them to be, could be false. We may think it inconceivable that they may be false, but that does not imply that they are false.

C.1.2 ***Contradiction and fallibility***

Radical empiricists may defend their position be focusing on premise I). They may agree that they do in fact employ the law of non-contradiction as a principle that fulfils the function of specifying when a belief revision has to take place, but they may protest that they may not. They may claim that another principle can be used to fulfil that very function. I will argue that this cannot happen. Principles are normative, in the sense that there is an internal connection between a principle and its satisfaction conditions. A principle that has no satisfaction conditions is a contradiction in terms. The very idea of satisfaction conditions, involves the inconsistency of holding that a particular engagement in a normative task, may satisfy the principle and violate it, in the same sense and at the same time. This amounts to the normative version of the law of non-contradiction. If an action A is at once in accordance with a principle and not in accordance with it, serious doubts arise as to whether what we call a principle is in fact a principle or not. One of the issues that arise is that if we are not committed to the truth of the principle of non-contradiction, we face trouble in understanding fallibility itself. To motivate this claim, I give the following argument:

I) Due to their normativity, principles can be violated as *opposed to satisfied,* or satisfied *as opposed to violated*.

II) If the law of non-contradiction is not known to be true, then we cannot know whether the very same action satisfies and violates the very same putative principle, in the same sense and at the same time.

III) If we cannot know whether a putative principle may be violated *as opposed to satisfied* or satisfied *as opposed to violated*, we cannot know whether that putative principle is a normative one; i.e. a principle with satisfaction conditions.

IV) Fallibility consists in violating a principle as opposed to satisfying it.

V) Therefore, if we cannot know whether a principle may be violated as opposed to satisfied, or satisfied as opposed to violated, we cannot know whether we are fallible in light of it.

VI) It follows from Mill's thesis that we cannot know whether an action satisfies and violates a putative principle, in the same sense and at the same time.

VII) Therefore, it follows from Mill's thesis that we cannot know whether we are or can be fallible in light of any principle.

Premise I) unpacks the normativity of a principle. By 'normativity', in this context, we do not have to refer to prescriptivity. 'Weak' normativity is sufficient to get our point across. Principles are those that can fail to be satisfied by those under their scope. And to fail to be satisfied is to be violated.[77] Premise II) claims that the law of non-contradiction is involved in our very own idea of a principle, via being presupposed in understanding whether the principle is satisfied or violated. Premise III) claims that if we cannot know when something violates a putative principle and when it satisfies it, then serious doubt can be casted as to whether the putative principle is in fact, a principle or not. If everything satisfies it, then it is not a principle that we may use to fulfil the first function. And if we cannot know whether contradictions are true, we also cannot know whether that which seems to us to violate the principle doesn't satisfy it as well.

Premise IV) claims that fallibility consists in violating a principle, as opposed to satisfying it. To make a mistake is to make a mistake *in light of a principle*. Fallibility and epistemic conflict are inextricable interwoven. And epistemic conflict is always epistemic conflict in light of a principle. Whenever we claim someone or something to be mistaken, we hold that persons, beliefs or inferences are in conflict with something else, whether a principle or something entailed by one.

[77] In this context, the violation of a principle means that that an agent has not acted in accordance with it.

Statement V) is an intermediate conclusion which follows from premises III) and IV). If we cannot know whether actions violate putative principles as opposed to satisfying them, or satisfy them as opposed to violating them, we cannot know whether the putative principles we refer to can serve as principles, because serious doubts are casted with respect to their normativity. And precisely because fallibility consists in violating a principle, i.e. violating it as opposed to satisfying it, we cannot know whether we are fallible. Absent such principles, there can be no mistakes. Premise VI) links Mill's thesis with the previous discussion. If we cannot know whether true contradictions exist, we cannot know whether true contradictions exist in the case of actions. That implies that we cannot know whether a putative principle, is in fact a principle, because we cannot know whether it is normative. We cannot know whether we can violate it *as opposed to satisfying* it or vice versa, and in that, we cannot know whether there are errors; violations of principles. It might seem to us that there are but from Mill's thesis it does not follow that we can know whether this is the case or not. Statement VII) is the argument's final conclusion which states that if we accept Mill's thesis, we cannot know whether we are fallible. It might seem to us that it is impossible that this is the case, but Mill's thesis claims precisely that the fact that we cannot conceive of something implies neither that it does not exist nor that it cannot exist. It seems that due to its commitment to Mill's thesis, radical empiricism ends up denying much more than the idea that we have some knowledge of what must be.

The radical empiricist might protest that the standards for what counts as a principle, as well as the standards for what counts as fallibility, are too strict. We may adopt a normative principle, in a deflated sense. According to this deflated reading, a putative principle can be normative if it *seems* to us that it can be violated *as opposed to satisfied* or satisfied as *opposed to violated*. It may be a fact that true contradictions exist and that a principle cannot be violated as opposed to satisfied, or satisfied as opposed to violated, but this doesn't matter. If it *seems* to us that it cannot be violated and satisfied in the same sense and at the same time, then we may use it. So, although we may use it as our central principle that specifies the conditions under which a belief

rejection or revision has to take place, it does not mean that we have to use it. The law of non-contradiction is involved in our understanding of the satisfaction conditions of the new principle, but not in a way that commits us to its truth.

This response shows exactly how radical empiricism loses sight of truth. To claim that it doesn't matter whether it is true that contradictions exist, but we can rejoice in the fact that we can coordinate by being committed to rejecting them, is precisely not to care about whether the fundamental principles of our theories are connected to the world as opposed to how we think of it. The question is not just whether we can use something as a principle, but whether we can use it as a principle that provides us with a connection with truth. To say that we can accept something as a principle is one thing. Whether this principle is true or valid is quite another. If the central principle of our inquiry has no connection to truth whatsoever, then we cannot know whether employing further principles that specify what belief to revise, will lead us closer to or further away from it. Surely, coordination is achieved if we mutually agree that we have to reject contradictions, but the way it is achieved severs the ties between the fundamental epistemic principles and truth. There can be no bridging the gap between appearance and reality if one claims that our fundamental principles do not reveal to us anything about the world, besides the workings of our own minds. On the other hand, a rationalist who rejects Mill's thesis can maintain that the workings of our own minds do show us something about reality. Even if entirely negative, our knowledge that contradictions do not exist amounts to knowledge of reality. Thus, it seems that rationalism can close the gap between mind and world that the radical empiricism opens and avoid the radical skepticism that it implies.

If the considerations in sections C.1.1 and C.1.2 are correct, then radical empiricism does lead to a radical form of skepticism. In denying that we have some knowledge of what must be, it ends us denying far more. It leads to disentanglement between truth and epistemic principles. A final word is due. The metaphysical law of non-contradiction claims that there are no true contradictions. In this sense, it is not the

law of non-contradiction so much that is used as a central principle that specifies conditions in which a belief rejection has to take place. Rather, it is the command: 'reject contradictions'. This allows us to see more clearly why epistemologies that do not incorporate a commitment to the truth of the metaphysical law of non-contradiction end up losing truth from sight. A question arises as to why we have reason to reject contradictions. Those who are committed to the truth of the law of non-contradiction may claim that we have reason to reject contradictions because contradictions are false. Those who are not committed to its truth cannot say this. All they may say is: as a matter of fact we reject contradictions without knowing whether what we reject as contradictory or as entailing a contradiction is false or not. We just reject it. And this is how truth is lost from sight. We can think of Socrates asking: do we reject something as false because it is *false* or do we call something false because we *reject* it? If we follow the first horn, truth is not lost from sight and its importance is not downplayed. If we follow the second horn, truth becomes something else. It seems that if we do not hold that we know the metaphysical law of non-contradiction, we will lose sight of truth in our inquiries. Since the law of non-contradiction is a claim about what must be, without some knowledge of what must be, we end up losing sight of truth.[78] We cannot answer anymore that we reject something because it is false. We end up claiming that something is false because we *reject* it. Even worse, we may end up claiming that we don't know whether it is false or not.

[78] In this thesis, I presuppose that truth does not boil down to justification. This is both for purposes of exposition and due to the limited space provided by the thesis. Nevertheless, in positions that truth is reducible to justification, the transcendental arguments against determinism, in virtue of the a-rationalism it implies, would straightforwardly establish the existence of free will. The reason is the following: if truth boils down to justification, then falsity reduces to lack of the possibility of justification. As it will be pointed out in chapter III, according to transcendental arguments for libertarian free will, it is never rational to claim that one lacks free will. So, if truth boils down to justification, these arguments would work straightforwardly, without requiring any of the work that I discuss in chapters I and II.

C.1.3 ***Inadequacy and occultness***

Two objections remain to be discussed.[79] First, I will address Casullo's objection against the adequacy of moderate rationalism in accounting for knowledge of what must be. Second, I will address the objection that rational insight is inherently mysterious and does not bode well with a naturalistic picture of the world.

Casullo (2003: 100-104) challenges rationalism by challenging the idea that rational insight is adequate to give us knowledge of what must be. He attacks BonJour's 'generality' argument and claims that it is not strong enough to show us why rationalism is not deficient in accounting for knowledge of what must be, just like radical empiricism. He writes:

> Moderate rationalism can avoid the skeptical consequences of the Generality Argument only if it can sustain the claim that we can apprehend properties. Moderate rationalism can sustain this claim only if it can provide a nonmetaphorical account of the apprehension of properties that does not involve some kind of quasi-perceptual relation to those properties. Since he does not provide an account of how a thought can have as its content some property, … BonJour's moderate rationalism does not avoid the skeptical consequences of the Generality Argument. (2003: 103-104)

[79] For various reasons, I will not discuss Paul Benacerraf's (1973) objection against the existence of abstract objects. His claim that our best epistemic theories require causal contact between the belief that a fact obtains and that very fact, seems to me to rest on physicalism and the rejection of rationalism. Thus, it cannot be an argument against rationalism. Furthermore, a causal account of modal knowledge –knowledge of what must be- seems to me to be inherently inadequate. First, finite causal chains cannot establish knowledge of something that must be. Second, assuming a causal link between a fact and a belief that it obtains, the modal status of the fact, seems to be irrelevant in the causal link, however reliable the link is and continues to be. The specified causal links that a reliabilist has to identify seem to be independent of the modal status of the facts that cause beliefs about them.

Casullo points out that the same kind of argument that leads us to hold that radical empiricism is deficient in accounting for knowledge of forms of necessity, can be used against rationalism itself. The main reason why we do hold that sense experience cannot give us knowledge of what must be is that it allows us access to particulars. No amount of observation of what holds in particular cases may establish what must hold in all cases. And we cannot ground knowledge of what must be in terms of observation of particular cases. Furthermore, we cannot ground the validity of a principle, which involves a form of necessity, upon intuiting a number of particular instances in which employing that principle in our thinking leads us from true premises to a true conclusion. The rationalist move is to appeal to an extra-sensory faculty of rational intuition which provides the source for the desired knowledge. But Casullo points out that if the content of what is grasped by rational intuition is also particular, it will also be deficient in accounting for knowledge of what must be. At the end of the day, rational intuition seems to be giving us little more than sense experience does.

Essentially, Casullo's argument can be answered in the following ways. Casullo targets moderate rationalism, rather than traditional rationalism. An account of our knowledge of the metaphysical law of non-contradiction, along the lines discussed in the previous two sections [C.1.1-C.1.2] of this chapter seems stronger than what moderate rationalism may be able to give us. But even if this is not the case, it is plausible to claim that what is involved in grasping the law of non-contradiction involves the apprehension of properties.[80] It seems plausible to claim that to grasp that objects cannot have contradictory properties is to grasp among other things, something about properties.[81] So there doesn't seem to be a reason to think that what rational

[80] For a defense of a qualified version of BonJour's 'generality argument' along these lines, against this argument by Casullo, see Joshua C. Thurow (2008).

[81] An alternative would be to claim that no such thing is needed. What is grasped is something that involves the concept of being, or Being itself.

insight grasps is particular in the metaphysical sense. It can be described as a particular necessity that is grasped about the world's structure.

Those who embrace Casullo's criticism against rationalism may respond in two ways. If they consider skepticism to be a live option, they can claim that rationalism does not answer skeptical worries and claim that it is not better than radical empiricism. If they do not consider skepticism to be a live option, they may proceed to claim that accepting rationalism is ontologically costly because it commits us to entities and faculties like properties and rational insight, which do not fit well in a naturalistic worldview.

Against the first reply, rationalists may claim that if one enters the debate as someone who entertains skepticism as a live option, nothing could ever answer her worries. For the skeptic, no theory is *ultimately* better than another in countering skepticism. Seen through the perspective of a skeptic, we cannot ever produce justification for any knowledge-claim. So, this point, is not so much a point against rationalism. It is a point against every epistemology that aims to give an account of *how* we know what we know.

Against the second reply, rationalists may respond that ontological simplicity is not a virtue in itself. It may only be a virtue, if combined with explanatory power. Assuming one is a non-skeptic, one accepts that we do possess knowledge and one of the tasks of epistemology is to uncover how we can know what we know. The arguments is sections C.1.1 and C.1.2 motivate the claim that without some knowledge of what must be, the connection between our epistemic principles and truth is lost entirely. If after this, one remains a non-skeptic, then one accepts that a fundamental desideratum is left unaccounted for unless we inflate our ontology by positing the faculty in virtue of which we grasp that which is required to account for it; something that must be the case. There is no reason to prefer a theory with a simpler ontology to one with a more complex one, if the former does not explain something that needs explaining, while the latter can, precisely because it has a more complex ontology.

Still, an entirely justified worry might remain as to the nature of one's engagement with this debate. Do rationalists embrace an ontology of the mind's faculties because they want to explain something or do they first commit themselves to it for reasons of a different kind and then try to defend it no matter what? This seems to me to be a serious worry because how one engages in a debate indicates its fruitfulness. And just as one may enter the debate as a skeptic and claim that no theory can ultimately answer skeptical worries, one can also enter the debate as someone who accepts an inflated ontology and just reject any claim against it, simply on the grounds of it being against her or his preferred ontological commitments. This may well be the case. But we have to make a distinction between a thinker's psychological motivations and a thinker's arguments. Clearly, some thinkers do engage in debates in precisely that manner. But they don't have to. Furthermore, this is not something unique to rationalism. In fact, it seems that those who argue against rational insight on grounds of its irreducibility to naturalistic terms and faculties are precisely presupposing a particular ontology of the mind, its faculties, as well as the world. As mentioned above, the topic of a priori justification is concerned both with the ontology of the mind's faculties and their epistemic reach.

A lot of the subsequent debate seems structured around the intelligibility or unintelligibility of the attempts to give a positive account of a priori justification. As we have seen in section A, most accounts of a priori justification are thoroughly negative in that they equate a priori justification with justification independent of sense-experience. We are told what a priori justification is not, not what it consists in. It is maintained by opponents of rationalism that all positive accounts of the source of rational insight are mysterious and occult. These criticisms lack force when it comes to convincing the rationalist to abandon the idea of a priori justification for the following two reasons: i) a positive account doesn't have to be given, and ii) the positive account of the source of a priori justification can only be mysterious and occult if we have already conceded that the standard of what is ontologically acceptable or not is potential reducibility to what can be known by sense-experience.

With respect to the first, a positive account of the source of a priori justification might not be possible to be had. But that is not a reason to abandon the idea. The very reason for positing such a faculty is the deficiency of elements of sense-experience in accounting for knowledge that we do accept we have. If we allow that we do possess knowledge that cannot be accounted for by sense-experience, we will have to allow room for a non-sensory source of epistemic justification.[82]

With respect to the second point, the argument begs the question against the rationalist. For, if we allow that we have to posit such a faculty, the fact that it is pronounced to be occult when examined in light of how it reduces to what can be known by sense experience, is not a reason to disbelieve in it. On the contrary, it is the exact reason why we have to accept it. To be more precise, the thinker who holds that the deficiency argument in section B is valid, sound and is to be interpreted along rationalist lines rather than Kantian or moderate empiricist ones, is committed to the idea that there is a form of knowledge that sense-experience cannot account for. What would be mysterious would be for the rationalist not to posit such a faculty. It would mean that we have knowledge without having the relevant faculty which would allow us to have it. Hence, the rationalist has to posit a faculty of mind that is extrasensory. If such a faculty is reduced to a sensory one, the knowledge that we claim to have, specified by premise (I) of the deficiency argument in the beginning of section B, will be unaccounted for. Therefore, if we are to account for such knowledge, the faculty of mind which will allow us to grasp its objects has got to be a faculty that defies reduction in sensory faculties. But apart from the conceptual irreducibility, its operation has to be irreducible to the operation of sensory faculties as well. The demand that it be reducible to sensory faculties is illegitimate and misunderstands the reason why

[82] It has to be mentioned that sense-experience is equally mysterious. Naturalistic accounts do not seem to explain it; they seem to be trying to give a plausible account for it, with the elements available to us. It is hard to think that lack of an ultimately satisfactory account of how sense-experience operates would ever lead anyone to deny its existence.

someone is led to believe in such a faculty. And it is on such a demand that this protest against rational insight rests. Ontological simplicity is not a theoretical virtue in itself. It can only be a theoretical virtue if combined with explanatory power. The whole point of introducing rational insight into our ontology of mind is that without it, we will not be able to account for a fundamental desideratum of an epistemic theory.

This point seems to be rather important for showing where and how metaphysical considerations about the nature of the mind infiltrate one's thought about the topic of a priori knowledge and justification. Traditionally, the notion of a priori knowledge and justification is associated with a non-sensory faculty of mind which can be called 'rational insight' or 'rational intuition'. To someone who starts by holding as a premise the truth of a naturalistic or empiricist theory of mind, there will be no room whatsoever for such a non-sensible faculty of mind. Phillip Kitcher (1980) writes:

> A person's experience at a particular time will be identified with his sensory state at the time. (Such states are best regarded physicalistically in terms of stimulation of sensory receptors, but we should recognize that there are both "outer" and "inner" receptors.) The total sequence of experiences X has had up to time t is *X's life at t*. (1980: 5)

In this passage, Kitcher construes 'experience' in such broad terms on the one hand, as to cover one's whole life up until a time t, and such narrow terms on the other hand, by equating one's life up until time t, with that person's series of sensory states up until that time. In a nutshell, he gives a broad definition of experience combined with a narrow account of the nature of the mind's faculties. It is clear that this begs the question against the rationalist. Kitcher starts with the empiricist theory of mind as a given. It is clear that nothing can be epistemically independent of sense-experience if sense-experience exhausts all aspects of one's life. But this masks a terminological shift that leaves the issues unaddressed. Once a thinker adopts a Humean or physicalist ontology of mind, they can constantly maintain that recalcitrant examples that the

rationalist offers of what we can know self-evidently, are to be treated as a mistake on our part rather than actual cases of a priori knowledge (Salmon 1967; Harman 2001). Wesley Salmon writes:

> Even if a recalcitrant example were given –one that seemed to defy all analysis as either analytic or a posteriori – it might still be reasonable to suppose that we had not exercised sufficient penetration in dealing with it. If we are left with a total epistemological mystery on the question of how synthetic a priori propositions are possible, it might be wise to suppose it more likely that our analytic acumen is deficient than that an epistemological miracle has occurred. (1967: 40)

The problem with this move is that it comes one crucial step too late. The whole debate is one which starts with the question: can an empiricist theory of mind accommodate all instances of our knowledge? To start by presupposing that it does is not to engage on the debate; it is to reject a priori knowledge from the very beginning. Once again, we see that the alleged mysteriousness of rational insight is a charge against rationalism on the very ground that it posits epistemic capacities that do not seem to be reducible to those that can be accommodated by an empiricist ontology of mind. But this charge is question-begging from the very beginning. It is little more than a stubborn reluctance to engage with the issues properly.

If we accept the deficiency argument we don't have to give a positive account of a priori justification. We can accept its existence as well as the metaphorical nature of our accounts of it. That is why the counterintuitive nature of accounts of the a priori, in terms of 'seeing with the mind's eye', 'or quasiperceptual acquaintance', and so on, don't have to be taken literally. Their exotic nature doesn't constitute a reason to reject them. As mentioned before, the whole point for introducing such a faculty is that it is

irreducible to empirical faculties which are deficient for accounting for a segment of our knowledge that is necessary in order to avoid a radical form of skepticism.

Conclusion

This chapter constitutes a defense of a priori knowledge and justification. A priori justification is defined as justification that is independent of sense-experience. Such independence can be considered in ways which broadly involve unrestricted or restricted temporal priority. The former seems to capture the Platonic doctrine according to which, the soul has perfect knowledge of the Forms before birth. The latter seems to require sense-experience, as an enabling a priori justification without grounding it. Furthermore, it was said that accounts of a priori justification that define it in terms of independence from experience rather than from sense-experience, obfuscate the issue that the debate about a priori justification is concerned with the nature of the mind's faculties as much as it is concerned with epistemology.

The position that is defended in this chapter and in the thesis at large is a rationalist one. Rationalists frequently motivate their position by employing deficiency arguments against non-rationalist views. The main insight that drives such arguments is that we have some knowledge of what must be the case, which cannot be based on what we know to be the case, contingently. Due to the limited space that a PhD thesis allows, four views are mentioned but two are discussed. The views that are mentioned are: i) rationalism, ii) Kantianism, iii) logical empiricism and iv) radical empiricism. The views that are discussed are rationalism and radical empiricism, as the perceived stronger opponent to rationalism.

Section C involves an argument for the centrality of the law of non-contradiction in our thinking. It motivates the claim that without a commitment to its truth, our epistemic theories lose sight of truth. Each epistemic theory requires principles in order to perform at least two functions. The first function concerns the specification of conditions in which a belief rejection or revision has to take place. The second function

concerns the specification of particular revisions that have to be made, once it is established that a revision has to take place. A rationalist account of our knowledge of the law of non-contradiction allows us to claim that our inquiry is not entirely disconnected from truth. If we accept the law of non-contradiction as true, upon encountering contradictions, we can reject them and the theories that entail them, as false. A radical empiricist epistemology seems inadequate to the task. Radical empiricists accept Mill's thesis, according to which, the fact that something is inconceivable implies neither that it does not exist, nor that it cannot exist. This leads to a skepticism about whether contradictions are false or not. Eventually, such doubt leads us to skepticism about whether the epistemic principles we may employ are connected with truth. We end up in the position of not being able to know whether the circumstances that are specified by the principle that fulfils the first function, actually call for a rejection to take place or not, as well as whether employing the principles that fulfil the second function, leads us closer to or further away from truth. Furthermore, a commitment to the law of non-contradiction is involved in the very idea of the satisfaction conditions of a principle. Absent a commitment to its truth, our very own fallibility becomes doubtful. These considerations render radical empiricism vulnerable to a radical form of skepticism. Finally, two objections against rationalism are discussed. The first, by Albert Casullo (2003: 103-104) claims that rationalism does not ultimately perform better than radical empiricism in avoiding skepticism. The second claims that the faculty of rational insight is not reducible to a naturalistic worldview and therefore, should be rejected. Against the first objection it is claimed that there are two ways of engaging in the debate. One can engage in the debate as someone who holds skepticism as a live option or as a non-skeptic. If one holds skepticism as a live option, no theory will ultimately be able to answer her. If one does not, then one may accept that rational insight can grasp properties. Doubts about the plausibility of such an option rest ultimately on the second objection; that of the claiming that rational insight is inherently mysterious because it does not fit in a

naturalistic worldview. Against the second objection, it is argued that it has no force because it ultimately rests on a stubborn presupposition of a particular ontology of the faculties of mind. As mentioned above, the debate about a priori justification is a debate that is concerned with the ontology of the mind and its faculties, as much as it is concerned with justification and what we can know. Therefore, to presuppose a particular ontology of the mind's faculties and reject views on that very basis is not to engage in the debate properly.

By making this move, we may also close the discussion that we opened in chapter I about the normativity of reasons, by countering the error theoretic objection against normative facts. 'Queerness' arguments against claims about normative reasons hold that the existence of such reasons should be rejected because they do not fit easily in a naturalistic picture of the world. If we allow that we may have knowledge of what must be, we have to posit faculties that also do not fit in a naturalistic worldview. No such account holds force anymore. The fact that irreducibly normative reasons are not naturalistically reducible is no objection against embracing their existence. It is time now to discuss transcendental arguments against determinism and show how incorporating them into a rationalist framework, extends their reach and can turn them into positive argument for libertarian free will.

Chapter III: The libertarian rationalist argument

In this chapter, I present a rationalist version of the argument that thinking and acting in light of reasons requires free will. In conjunction with several presuppositions of the thesis, such as the position that knowledge requires justification of an internalist sort, as well as the "'Ought' implies 'can'" principle, the argument concludes that to the extent that we can think and act in ways that we may call justified or unjustified, we possess the powers we naturally associate with libertarian free will. This argument resembles several transcendental arguments against the rationality of determinism, although it differs from many of them, in virtue of the fact that it aims to establish a positive ontological claim and is explicitly not Kantian in nature. Along with the aforementioned presuppositions of the thesis, the rationalist framework that was developed in chapter II serves as the argument's scaffolding which allows for a more secure understanding of its ontological claims and implications. In a nutshell, it provides the ground for libertarians about free will to make an ontological claim, as opposed to a claim about how the mind must view the world it constructs, in order for the latter to be intelligible.

In section A, I present the main argument of the thesis. Substantive rationality is essentially connected with normative reasons. To be under the scope of normative reasons, an agent has to have alternative possibilities with respect to meeting or violating the demand constituted or generated by them [Ch. I/A.4.1-A.4.2]. Alternative possibilities can be possessed by a situated person, only if that person has powers of control over her actions that are associated with libertarian free will; the ability to do otherwise. Lack of alternative possibilities renders the person as being outside the scope of normative epistemic reasons and hence, places that person outside the scope of substantive rationality. Thus, if epistemic reasons or substantive rationality are required for knowledge, lack of alternative possibilities implies lack of the possibility

of knowledge. And this is construed as a claim about the world; not about how the mind must see it.

In section B, I discuss various objections against the argument and respond to them. The second objection [B.2] concerns the possibility of a compatibilist understanding of the ability to do otherwise and will receive its fullest answer in chapter IV/A.2.3. Unfortunately, due to the limited space allowed by the PhD project, some objections will have to be left out for further research. An important objection which won't be dealt with, yet has to be acknowledged, concerns the ultimate feasibility of the combination between rationalism and libertarian free will, due to the frequent association of the former with the principle of sufficient reason.[83]

Section A: ***The rationalist argument for libertarian free will***

It is customary to think that we are normatively responsible for what we can control and not responsible for what we cannot. Throughout our lives, we are presented with circumstances that may either enhance or diminish the control we can exert over our actions. Awareness of the fluctuation of the impact of changing conditions upon the ability to control what we do, is reflected in our judgements of normative responsibility. As Haji holds, it is an "uncontentious view that factors beyond our control can effect changes in moral obligation over time or imperil obligation." (2018: 1) It is also customary to think that rationality has a normative nature which is

[83] In chapter II, I have hinted that this boils down to the question of the extent to which a rationalist treats the principle of sufficient reason as applying to the world. Traditionally, this concerns the following four domains: the logical, the mathematical, the moral –concerned with the motivation of action- and the causal/physical (Schopenhauer 1974[1813]). Libertarian free will cannot be accommodated by a rationalism that holds all four domains to be under the scope of the principle of sufficient reason. The requirement for libertarian rationalism to be coherent is that it treats the causal/physical domain, as well as the moral domain, as domains in which the principle of sufficient reason does not hold; i.e. not all actions have causes which explain why they had to happen as they happened and not in any other way. But, this is a much larger project which cannot be dealt with here in its entirety. It will have to wait for further research.

interlinked with normative responsibility. After all, to think rationally is to think in light of reasons that justify our thoughts, inferences and beliefs. Without a normative dimension, such reasons could never justify us in thinking what they support or make it the case that we are unjustified in thinking what they don't. If the considerations laid out in the first two chapters are correct, then reasons to deny the normativity of substantive rationality are unmotivated. We can allow that *some* instances of rationality –such as instances of structural rationality- may not be accompanied by reasons but we cannot hold that *all* of them are, without also denying that we can possess knowledge.

Despite the customary nature of such beliefs, dissenters exist. Some hold that we are neither rational nor responsible in a normative sense. Others claim that we are both, but that alternative possibilities are required neither for responsible action nor for rational one. One kind of argument that is raised against views that hold rationality to be possible for agents that inhabit deterministic universes, ends up concluding that determinism is rationally self-undermining or self-defeating[84] (e.g. Wick 1964; Jordan 1969; Hasker 1973; Boyle, et. al. 1976; Lockie 2018).[85] To oversimplify this position, its advocates maintain that if determinism were true, we could never affirm anything rationally, including the truth of determinism. Ultimately, thought would be determined by non-rational forces such as the causal force of brain events, which is independent of their mental content. Causal force and rational force would come apart.

Such arguments are predominantly negative. They frequently claim that rationality is not compatible with determinism, without proceeding to make the claim

[84] A thorough exposition of the history of such arguments, as well as of more positive arguments for free will, from the middle of the nineteenth century onwards, is given by Jim Slagle (2016).

[85] I do not claim that all of these authors embrace Kantianism. My whole point is that frequently, although not always, advocates of such arguments make a negative claim and do not proceed into making a positive one. To break with this negativity, a positive argument has to be made which seems to be of an a priori nature. To make a positive case, an a priori argument for free will, will have to be a synthetic a priori argument in a distinctively traditional pre-Kantian sense.

that it is a matter of fact that we are rational. Determinism could be true as far as advocates of such arguments are concerned. It is just the case that if we are determined, we cannot be rational in thinking and acting. The argument I will put forward results in a positive ontological claim which is arrived at in an *a priori* manner. That means that the explicit rationalist framework of this argument for free will, allows us to claim that if it works, it shows that we are actually free in the libertarian sense, as opposed to the Kantian strategy which at most, can show that we have to think of ourselves as free, if we are to treat ourselves as rational, which could be the case even if we were not free.[86] It is a synthetic a priori argument, in a pre-Kantian rationalist sense. Let us formalize the argument:

I. Substantive rationality essentially involves normative demands, which are expressed in terms of rational 'oughts', or in terms of what we have 'reason to do'.
II. Normative demands cannot form the ground of one's normative criticism if one cannot meet them (OIC), and if one can, if one cannot also fail to meet them (PAP), non-accidentally.
III. Demands that are directed towards situated beings are indexed at a time t and in circumstances C (F n); circumstances which have a specific form (n) rather than any other form (~n).
IV. Given (II) and (III), the applicability of substantively rational demands to an agent requires that the agent at a time t and in circumstances C (F n) can meet the demand or fail to meet it, non-accidentally.
V. If one lacks libertarian free will, one cannot satisfy the requirement specified in premise (IV).

[86] As such, it is closer in nature to the argument of John Lucas (1970), who appeals to Gödel's incompleteness theorems to argue that a being who is not free cannot understand Gödel sentences.

VI. Therefore, by (IV) and (V), if one lacks libertarian free will, one cannot be substantively rational or irrational.

VII. We cannot deny that we are substantively rational without denying the possibility of knowledge.

VIII. But the acquisition of knowledge is possible.

IX. Therefore, we have libertarian free will.

The first premise states the normativity of substantive rationality. If we conceive of substantive rationality as a particular kind of responsiveness to normative reasons, then being substantively rational is inextricably interconnected with them. One cannot be substantively rational or irrational without being under the scope of normative reasons. In this sense of rationality, to be rational is to think and act in light of normative reasons. Depending on what we take the central normative concept to be, the normativity of rationality will be expressed in terms of rational 'oughts', 'shoulds' or in terms of what we have 'reason to do'. Whether we take each of these notions as fundamental or derivative is a matter of conceptual strategy which ultimately depends upon whether we adopt conceptual normative monism or pluralism (Robertson 2009). As mentioned in the first chapter, I will not take part in this discussion and I will use the notions interchangeably. The bottom line is that due to their normativity, reasons and obligations generate demands with which, the person who is under their scope should comply. As Nicholas Rescher writes: "Rationality makes demands upon us. It speaks in didactic tones: this or that is what you *should* do" (1988: 9).

Premise (II) concerns the conditions that have to obtain for an agent to be under the scope of normative reasons. To be under the scope of normative reasons the agent must have alternative possibilities available to her; it must be possible for her to meet the demand that is generated by reasons or violate it, non-accidentally [Ch. I/A.4.1–A.4.2]. It is a very convoluted premise which concerns the proper conditions for normative responsibility and is composed of three distinct elements. First, it involves

Kant's law that an agent lies within the scope of an obligation to do X, when the agent can do X. According to Kant's law, "'ought' implies 'can'" and if the agent cannot do X, it is not proper to hold her normatively responsible for not doing X. Second, it involves an alternative possibilities requirement for being under the scope of a normative reason to do X. Normative responsibility requires the possibility of violation of such an obligation or reason.[87] To be under the scope of a normative reason, an agent has to be in a position where she has the ability to respond to it correctly or incorrectly. For an agent to 'have a reason to X' or to 'ought to X', it is required that she can X or refrain from X-ing. The third element is that of non-accidentality. To be properly assessed in light of normative reasons, an agent has to be able to act in light of normative reasons. Merely having reasons differs from thinking and acting in light of them (Davidson 1963). From the constituent elements of premise (II), we get the following condition of normative responsibility:

CNR: For an agent to 'have a reason R to X', that agent has to be physically able to X or refrain from X-ing, in light of R.

Premise (III) sheds light on the spatio-temporal element of the conditions of normative responsibility. Human beings are situated beings; we occupy concrete circumstances which have a specific form rather than another one, at a given moment in time. Such circumstances present us both with opportunities and limitations. To speak of a situated agent as being under an obligation to do anything is to say that an agent has an obligation in circumstances C (F n) –circumstances which have a particular form (n) rather than another form (~n) - at a time t to engage in a particular act, whether mental or physical, at some point in the immediate or far future. Such

[87] I contrast normative responsibility –the kind of responsibility we inquire about in matters of morality and rationality- with causal responsibility. In contrast to the former, the latter does not imply that one has violated any normative reason in doing something or letting something happen. An agent that is just causally responsible for her behaviour cannot be and isn't under the scope of normative reasons.

circumstances involve a multitude of factors which might make it easier for us to do the right thing, harder or even make it impossible; sometimes permanently, other times temporarily. Because we are situated beings who occupy concrete spatio-temporal circumstances, factors involved in such circumstances may impose constraints for action which render a person outside the scope of particular normative reasons. Once we incorporate the constituent elements of the second premise with the elements of the third, we can take a more sophisticated approach to the conditions of normative responsibility which includes reference to the agent's spatiotemporal circumstances.[88] We can account for conditions of situated normative responsibility (CSNR) in the following way:

CSNR: For an agent in circumstances C (F n) at a time t, to 'have a reason R to X', that agent has to be physically able in C (F n) at a time t, to X or refrain from X-ing, in light of R.[89]

CSNR involves reasons to do something at that very time. In that sense, demands generated by reasons are conceived of as synchronic.[90] And this presents us with a problem: if reasoning takes time, then reasoning from the identification of reason R to a final verdict takes some time as well. It cannot plausibly refer to something that has

[88] Unfortunately, many authors who speak of agents, speak of agents at a time t, without also mentioning that the agent is also situated in concrete circumstances. In Ch. IV/A.2.3, I will argue that conditional analyses of ability, which are an integral part of classical compatibilism about free will and determinism, are ultimately untenable and that compatibilists end up treating situatedness as an accidental characteristic of an agent, rather than an essential one.

[89] It might seem weird to say that the agent can refrain from X-ing in light of R, given the fact that R favours X-ing. But, the 'in light of' clause is not interpreted in a causal sense [Ch. I/A.4.2]. In that sense, refraining from X-ing in light of R is more like the phenomenon of weakness of will.

[90] A very interesting discussion about the difference between synchronic and diachronic requirements of rationality can be found in John Broome (2013: ch.9, 10)

to take place at that time irrespective of our temporally extended reasoning process. That would violate the 'in light of' clause. Thus, to cover such cases, we can make the following modification to CSNR:

*CSNR**: For an agent in circumstances C (F n) at a time t, to 'have a reason R to X at t+n', the agent in C (F n) at t has to be physically able to X at t+n or refrain from X-ing at t+n, in light of R.[91]

Changing factors might make it temporarily or even permanently impossible for us to act in light of such an obligation or reason. The presence of such factors which make it impossible to either meet or violate the demand non-accidentally transform the normative 'ought' to a non-normative one. Such an axiological statement places us in another domain of the axiological realm. A normative demand that can be directed to an agent at one time can be inapplicable at another –say for reasons of manipulation- and kick back in afterwards (Vranas 2007).

Premise (IV) is an intermediate conclusion that follows from premises (II) and (III). Since premise (II) holds that an assessment of the agent in light of demands of normative reasons is senseless if the agent cannot non-accidentally meet or violate the demand generated by them, and premise (III) states that demands directed to situated beings are always indexed at a time t and circumstances C (F n), it follows that in order to be properly assessed in light of a normative reason, an agent has to be able to non-accidentally meet the demand generated by it or violate it, at a time t and in circumstances C (F n). Absent such conditions, the agent does not occupy the normative standpoint with respect to the normative reason in question; she is not within the range of its authority. Thus, since substantive rationality is essentially connected

[91] CSNR* involves diachronic demands whereas CSNR involves synchronic ones.

with normative reasons, not being under their scope implies that one cannot be substantively rational.[92]

Premise (V) states that the relevant powers that are required of an agent to be able to non-accidentally meet or violate the demands generated by a reason at a time *t* and in circumstances C (F n), are powers that are associated with libertarian free will. This is the main point of the whole thesis and its defence will be completed in ch. IV where it will be argued that a compatibilist understanding of such powers is not a viable option.

Premise (VI) is an intermediate conclusion that follows from premises (IV) and (V). If one lacks libertarian free will, one cannot be properly held accountable to normative reasons because one is not under their scope. To be substantively rational or irrational, one has to be under the scope of normative reasons. Thus, if one lacks libertarian free will, one is not substantively rational; at least, with respect to the cases where the agent lacks libertarian free will.

Premise (VII) states that we cannot deny that we are rational without embracing skepticism. The very process of denying the possibility of rationality or irrationality presupposes such a rational status for the denial itself. Therefore, denying rationality on the basis of reasons, is rationally self-undermining. An objection will be considered in this chapter [B.6], according to which a-rationalism might be true, albeit rationally unaffirmable. In order to counter this objection, I will draw a distinction between two understandings of the claim that something can be true, yet rationally unaffirmable. It is one thing to interpret this claim as implying that although we can rationally affirm some things to be the case, others we cannot. It is a wholly different to interpret this claim as implying that we cannot rationally affirm X –in this case, the truth of a-rationalism- because we cannot affirm anything. I will claim that a-rationalism cannot

[92] A distinction has to be made between not being under the scope of normative reasons in some cases and not being under their scope in all cases. In defending substantive rationality, I do not defend the claim that we are always under the scope of normative reasons. It suffices to claim that we are sometimes under their scope.

be a position which can be maintained as true in the context of the former interpretation. And the second interpretation is just a disguise of radical epistemological skepticism.

Premise (VIII) concerns the falsity of skepticism. If skepticism is false, then knowledge is possible for us. That does not mean that knowledge is possible at each and every moment of our existence. The previous emphasis on the importance of our situatedness prevents us from making such a claim because it leads us to recognize that it is frequently the case that we are not under the scope of normative reasons. But to claim that we are always outside their scope seems a wholly different claim. This premise is an explicit presupposition of the thesis. I do not give any argument for it and it seems to me that it is very difficult to engage in a genuine debate with a skeptic. There doesn't seem to be any common ground. Nevertheless, I fully accept the fact that this argument will not convince a skeptic. To maintain the possibility that a skeptic can be convinced by any argument seems to be hinting towards the fact that the person we are talking about is not an actual skeptic. It is also stated explicitly that this thesis does not aim to address the skeptic of this kind. Instead, it is an attempt to engage in a debate with those who believe that epistemic skepticism is false, yet are skeptics about libertarian free will.

Proposition (IX) is the final conclusion. If we cannot deny our potential to be substantively rational as well as the possibility of the acquisition of knowledge, any ontological picture that cannot accommodate libertarian free will leads to a conception of human beings as a-rational. Thus, we have to conclude that libertarian free will is a necessary condition for substantive rationality and consequently, for knowledge.

Section B: ***Objections***

In this section, I will discuss six objections to our argument. According to the first objection, premise II) can be denied because normative facts are entirely reducible to evaluative facts. Thus, no appeal to alternative possibilities is needed to render it 'justified' for an agent to do something. The fact that it is good suffices to ground the

rightness of the action, regardless of other considerations about the agent's abilities or knowledge. According to the second objection, premise V) can be denied because it ultimately rests on a libertarian understanding of premise II). Advocates of this objection may agree with premise II) but give a compatibilist reading of it. This is the main objection I will deal with in this thesis and it will be ultimately answered in Ch. IV. According to the third objection, indeterminism is worse for freedom and control over action due to the randomness implied by non-deterministic worldviews. According to the fourth objection, premise II) can be denied because alternative possibilities are either not required for normative responsibility [B.4.1-B.4.2] or not required for praiseworthiness [B.4.3]. According to the fifth objection, we can never be substantively rational, in the way that the argument implies, because we lack control over our beliefs. Finally, according to the sixth objection, the argument's conclusion doesn't follow from premise VII) because the fact that a position is rationally unaffirmable does not imply that it is false. It may still be true.

B.1 ***Reducing normative facts to evaluative facts***

A reductionist account of the normativity of rationality would explain the normative/deontic terms appealed to in accounts of substantive rationality, in evaluative terms. Such an account would explain 'justified' and 'unjustified' in terms of 'goodness-conducive' and 'badness-conducive' respectively. Selim Berker (2013a, 2013b) argues that we can draw a useful parallel between ethical consequentialism and epistemic consequentialism.[93] Just like in the former, the 'right' is held to be derivative from the 'good', –facts about value- so in epistemology the epistemically 'right' can be derived from facts about the epistemic 'good'. Berker (2013a) holds that:

[93] Berker prefers the term 'epistemic teleology' in (2013a) and the term 'epistemic consequentialism' in (2013b).

> Most teleological theories have three basic components: *a theory of final value*, *a theory of overall value* and *a deontic theory* (…) When all three are present, each successive theory crucially depends on the preceding one: the theory of overall value depends on the theory of final value, and the deontic theory depends on the theory of overall value. (2013a: 344)[94]

The deontic theory prescribes what ought to be done in terms of what is the overall best state of affairs that can be brought about by our actions, which in turn rests on what kinds of states of affairs we ultimately consider as intrinsically valuable. Translated into the epistemic realm, according to Berker, an epistemic consequentialist/teleologist would essentially be a truth-conducivist. She would have to assign states of intrinsic epistemic value and disvalue, such as the possession of true beliefs and false beliefs, respectively. Such specification of final value would serve as the ground for judging the overall value of alternative thought processes, which in turn would assign deontic properties to them.[95]

This viewpoint does not automatically translate to a criticism of our argument. We can concede that claims about reasons to believe are or can be derived from claims about what is epistemically good or bad, such as the possession of true or false beliefs.

[94] Berker (2013a) holds scalar versions of teleology to be an exception because they lack a deontic theory. Nevertheless, he accepts that the ethics/epistemology analogy is an analogy that can be stressed far enough, without there being an absolute identity between them. (2013a: 339) Since they do not involve a deontic theory, I will not deal with them in this thesis, because they are not intended as grounds for normative notions.

[95] Berker (2013a, 2013b) holds that just like ethical consequentialism faces the objection that it obfuscates the separateness of persons, truth-conducivism is guilty of not taking seriously the separateness of propositions and demands implausible trade-offs between alternatives, such as disbelieving doctor's positive diagnoses of a deadly virus for the sake of achieving states of true beliefs by carrying on research, which would not be carried out for psychological reasons if the patient believed the doctor's diagnosis. Although I am sympathetic to his points, the argument I will give does not require the truth of Berker's criticism of epistemic teleology/consequentialism.

An objection against our argument arises only if the person who concedes this, does so in an unqualified way; i.e. if she or he think that facts about epistemic reasons are based on facts about epistemic value, regardless of facts about the agent's knowledge and abilities. Such a reductionist about the normativity of rationality would maintain that in order for us to be held accountable in light of reasons, it does not have to be up to us to meet or violate the demand constituted or generated by that reason, at a given time t. Since in that sense, normative/deontic notions are reduced to evaluative notions, which are not compromised by determinism or any other position which entails the falsity of libertarian free will, it is emphatically wrong to think that we need such powers and abilities in order to be properly called rational or irrational. Whenever an agent deliberates between A and B, we can judge which outcome is better and our evaluation will form the basis of our judgment of what rationally speaking, the agent 'ought' to do, as well as what the agent 'ought not' to do.

Our reductionist critic may challenge premise II) of our argument if she does not treat claims about the agent's knowledge or ability as relevant with respect to the truth value of claims about what we ought to think. The epistemic consequentialist norm of reasoning (ECNR) may be encapsulated in the following way:

ECNR: Agent S, in circumstances C (F n) at a time t, *should* (ought to/has reason to) think/infer p because thinking/inferring p would lead S to achieve epistemic value.

This is a very crude position. A lot more has to be understood with respect to what is implicitly built into it. ECNR does not have much to say as to whether the person who embraces it, thinks that there is nothing about the agent that would make this statement false, or inapplicable to the agent. Consider the following modification:

ECNR*: Agent S, in circumstances C (F n) at a time t, should (ought to/has reason to) think/infer p because thinking/inferring p would lead S to achieve epistemic value, provided S in C (F n) at t can think p in light of possessed reasons (i.e. non-accidentally).[96]

A crude reading of ECNR is hostile to our argument, whereas ECNR* is not.

A proponent of ECNR has various choices to pick from. She can pick ECNR or modify it so as to be responsibility-absolving with respect to *some* responsibility-relevant conditions about the agent, or even *all*.[97] This would open up the possibility that a thinker can be non-normatively responsible for achieving an epistemically valuable effect, just like in Ethics, we may concede that good effects might come from someone who was not under the scope of moral reasons. Thus, apart from ECNR, she could also pick the following:

ECNR-K: Agent S, in circumstances C (F n) at a time t, *should* (ought to/has reason to) think/infer p, because thinking/inferring p would lead S to achieve epistemic value, provided S in C (F n) at t knows the reasons that *would* justify thinking that p, despite the fact that S in C (F n) at t is physically incapable of believing that p.

ECNR-A: Agent S, in circumstances C (F n) at a time t, *should* (ought to/has reason to) think/infer p, because thinking/inferring p would lead S to achieve epistemic value, provided S in C (F n) at t is able to think/infer that p, despite the fact that S in C (F n) at t is not aware of the reasons that would justify thinking that p.

[96] For reasons of space and simplicity, I only refer to OIC here and omit reference to PAP.

[97] Doing so, would collapse any proposed version with ECNR* which does not represent a threat to our argument.

Both ECNR-K and ECNR-A move away from the crude ECNR. Their proponents cannot include both lack of knowledge and ability as responsibility absolving conditions without advocating ECNR*, which does not pose a threat to our argument.

All aforementioned expressions of the epistemically consequentialist norm of reasoning, except ECNR* face the same problem. They are unfair as norms of judging an agent as normatively responsible. If the considerations laid out in chapter I/A.3-A.4.1 are right, then it seems that these epistemically consequentialist principles cannot ground fair normative criticism. In all three, the agent is not judged in terms of doing the best she or he can, in light of their circumstances, physical, biological, cultural, etc. ECNR is completely blind as to whether an agent has epistemic capacities that allow one to think what it requires non-accidentally, and also completely blind with respect to the agent's situatedness. ECNR-K cannot be a norm of reasoning because it violates the OIC principle.[98] It literally claims that we have normative reasons to do what is impossible for us to do. And this is counterintuitive because it locates normative responsibility in what lies outside the normative standpoint [Ch. I/A.4]. ECNR-A cannot be a norm of reasoning because it is a norm for thinking something, independently of the reasons that justify it.[99] Assuming reasoning to be thinking in light of reasons, lack of awareness of such reasons makes any transition to thinking that p, an a-rational one.

To combat this reply, proponents of ECNR, ENCR-K and ECNR-A have to reject some really intuitive positions. They have to reject the idea that fair normative criticism is a criticism of an agent in light of whether they are doing the best they can, in light of the circumstances they are situated in. Such circumstances include all sorts

[98] Again, since OIC is an explicit presupposition of the thesis, I am fully aware that this is not likely to convince someone who rejects it.

[99] It has to be stated again that this thesis presupposes an internalist conception of epistemic justification. Perhaps the externalist about justification would find these norms of reasoning more plausible.

of factors, but most importantly, facts about the agent's physical and cognitive limitations. Not to judge an agent in light of them, is in principle equivalent to judging someone who was born in 1970 guilty for not preventing both World Wars.

B.2 ***A compatibilist account of OIC and PAP***

According to this objection, 'ought' does imply the power to do as one ought to do, as well as the power to do what one ought not to do. But such a power can be understood in terms of a reductionist/compatibilist account of ability. It does not require the ability to do otherwise in a libertarian sense. Even better, our critic will maintain, such an ability is compatible with determinism and the inference to libertarianism is unmotivated.

A compatibilist account of OIC is developed by Ralph Wedgwood (2013). He maintains that rational demands are expressed in deontic terms like 'rational ought' and that all kinds of obligations presuppose an ability to conform to them, as well as to violate them. Such an ability is not just the ability conferred to an agent in the sense of it being merely logically possible that the agent meets or violates the demand (Lavin, 2004). For ought-statements that entail *only* the logical possibility of their realization, Wedgwood (2013) writes: "There may be some 'ought'-concepts that do not entail any stronger kind of possibility than bare logical possibility. This may be the case with what I have elsewhere called the "'ought' of general desirability"." (2013: 71) But Wedgwood does not believe that this is an account of 'ought' that is normative in any interesting sense. As such, these claims cannot do justice to rational demands. He writes:

> However, many other kinds of 'ought' are different from this. In particular, some kinds of 'ought' seem to express a concept that is indexed to a situation of a particular agent *x* at a particular time *t*. (…) Here, it seems, it is necessary for the truth of the proposition '*O* <*x*, *t*> *(p)*' that this particular agent *x* should

have the *power* or *ability*, at this particular time *t*, to realize the embedded proposition *p*. (ibid: 72)

Broadly speaking Wedgwood agrees that if we are to make sense of rational demands we have to hold that they are indexed at a time t, and that mere logical possibility to meet them or violate them is not sufficient to place an agent under their scope. What is vitally needed to do that is the physical ability of meeting the demands of rationality or violating them, construed along compatibilist lines. In that sense, the compatibilist agrees with premise II) of the rationalist argument for libertarian free will and thinks that a conditional analysis of the ability to do otherwise suffices to capture both OIC and PAP.

To answer this objection, I will evoke certain considerations which will be exposed in a more detailed manner in the next chapter [Ch. IV/A.2.3]. Wedgwood's compatibilist treatment of OIC and PAP seems to require the plausibility of conditionalist compatibilism. Conditionalist compatibilism aims to account for the compatibility of free will and determinism in terms of the potential truth of qualified conditional statements about the agent's abilities. Provisionally, let us understand free will as the ability to do otherwise. Let us also employ Van Inwagen's definition of determinism, according to whom, "it is the thesis that there is at any instant exactly one physically possible future." (1983: 3) Conditionalist compatibilists hold that determinism does not represent an essential threat to free will because it is not necessarily the case that several conditional statements about the agent's abilities are necessarily false in deterministic universes. An agent in a deterministic universe possesses free will with respect to the cases where such conditional statements are true of her. She does not possess free will with respect to the cases where such conditionals are false. Since determinism does not threaten the truth of such conditional statements about the agent's abilities, it does not represent an essential threat to free will. Finally,

they hold that the truth of such conditional statements is sufficient to render free will compatible with determinism.

I will argue that this project is inherently flawed. A conditional analysis of the ability to do otherwise cannot ground *a situated agent's* ability to do otherwise. Agents are situated beings. We constantly occupy circumstances which have one form rather than another at any given time. At pains of talking about a worldless agent, reference to an agent is always reference to an agent in C (F n) at t [Ch. I/A.4.1]. Thus, asking whether we have several abilities or not, amounts to asking whether we have several abilities in specific circumstances in which we are situated and which assume one form rather than another at a given time. This directly impacts question of free will as well as the question of its compatibility with determinism. Assume an agent performed action A at t+1. Asking whether that agent could have done otherwise –say, B- than what she did, amounts to asking whether that agent in C (F n) at t could have done B at t+1. Assume that the agent who performed action A at t+1 inhabits a deterministic universe. Furthermore, if the universe is deterministic, asking the same question about that agent, amounts to asking whether that agent, who in deterministic world W, performed action A at t+1, had the ability in C (F n) at t to perform action B at t+1.

To see exactly how the conditionalist compatibilist loses sight of the situatedness of agents, it is helpful to inspect the answers they give to the aforementioned questions. A conditionalist answer to both questions will assume the following form:

CA (conditionalist answer): The agent could have done B at t+1 if and only if it is the case that: if she intended at t to do B at t+1, she would perform action B at t+1.

This answer is inherently flawed. Upon closer inspection, we will see that at pains of denying the situatedness of agents, CA will have to be modified in the following way:

CA*: The agent in C (F n) at t could have done B at t+1 if and only if it is the case that: if she intended in C (F n) at t to do B at t+1, she would perform action B at t+1.

Now, it is clear that the agent of whom we speak did not intend at t to do B at t+1. For, if she intended at t to perform action B at t+1 and proceeded to perform action A at t+1, the conditional would be false and the compatibilist would agree that this is not a case with respect to which the agent has free will. Since the agent at t did not intend to perform action B at t+1, the necessary condition for performing action B at t+1 is that the agent were situated in C (F ~n) at t.[100] But the situated agent, of whom we speak – the agent in C (F n) at t- cannot ever be in C (F ~n) at t. Thus, the necessary condition –specified by the antecedent of the conditional statement the conditionalist compatibilist proposes- is necessarily non-actual for a *situated* agent. Along conditionalist lines, the necessary condition for an agent who in C (F n) at t proceeded to perform action A at t+1, to have performed action B at t+1, is that she were situated in C (F ~n) at t. Thus, the necessary condition for *doing* otherwise is necessarily non-actual for a situated agent; an agent in C (F n) at t, of whom the question is asked.[101] Since the necessary condition for any situated agent to do otherwise is necessarily non-actual at all times, no situated agent can ever do otherwise at any time. Agents in deterministic universes are also situated agents. Therefore, along conditionalist lines, no agent can ever do otherwise in a deterministic universe. The truth of such

[100] This renders the antecedent of the conditional contradictory, at pains of neglecting situatedness. Unless 'C (F n) at t' is treated as an empty symbol, the locution 'if she intended in C (F n) at t to B at t+1' is identical with 'if she intended in circumstances in which she did not intend to B at t+1, to B at t+1'. In a sense, the situatedness arguments shows that whenever we refer to agents, we refer to flesh and blood human beings who inhabit concrete circumstances. Thus, the symbol cannot be empty of content, waiting to be filled by hypothesizing about the agent's situatedness. This implies that antecedent of the conditional is contradictory and therefore, impossible.

[101] For someone to be able to do something, the necessary conditions for doing it have to obtain. If such conditions do not obtain, then the agent cannot do that thing.

conditional statements is plainly insufficient to ground a situated agent's ability to do otherwise in a deterministic universe.[102]

This case can be directed against the compatibilist understanding of OIC as well as PAP. For the question here is whether the agent has the power to follow the principle or to violate it. The compatibilist says that the agent can have such a power in a deterministic universe. But we can respond by pointing out that since situatedness is essential to human beings, the question whether one has the ability to do otherwise is elliptical for whether one has the ability to do otherwise in given circumstances having one form rather than another, at a given time. In other words: does the agent in circumstances C (F n) at t, have the power to follow a rational demand R by performing (A) or to violate it by performing (B) at t+1? Let us assume that a thinker in Wedgwood's compatibilist universe in circumstances C (F n) at t, proceeded to perform (A). Could that person have performed (B) instead, in C (F n) at t? It seems that the answer is negative. For the human agent who in a deterministic universe is situated in C (F n) at t and performs A at t+1, cannot in the exact same circumstances freely perform (B) at t+1, because along the lines of the conditional analysis of the ability to do otherwise, the necessary condition for performing B at t+1 is that the agent were situated in C (F ~n) at t. To conclude, since assessment in light of normative demands presupposes the power to meet them or violate them, the agent who inhabits a deterministic universe and is situated in circumstances C (F n) at t cannot meet them or violate them at t+1. Therefore, since we are always in circumstances C (F n), we can never fulfil the condition of being able to meet or violate a demand generated by a normative reason if our powers are understood conditionally. Thus, along the lines of conditionalist analyses of our abilities, we can never be under the scope of a normative reason or obligation, whether moral or rational. And deterministic universes are no exception to that.

[102] In order to avoid misunderstandings of my point, it is not determinism that creates problem for the conditional analysis of the ability to do otherwise. The latter is its own worst enemy.

To recapitulate, in order to be properly assessed in light of a rational demand, we must be able to meet the demand or violate it. Since we are always situated in given circumstances, we can be properly assessed in light of rational demands in circumstances C having form (F n) at t, if in circumstances C (F n) at t, we can meet the demand or violate it, at t+1. If we follow the logical conclusion of the compatibilist understanding of OIC, we are led to deny that in circumstances C (F n) at t, we have the ability to meet a rational demand or to violate it, at t+1. If this argument works, it shows that we cannot square compatibilism with substantive rationality because the powers that are required for the latter cannot be accounted for in terms of the powers that are associated with the former.[103] Therefore, a compatibilist understanding of OIC and PAP does not seem to hold water.

B.3 ***The objection from chance***

Another objection that can be raised against our argument is the objection from chance. It is frequently protested against libertarian conceptions of free will that in their effort to run away from determinism they end up embracing something far more inimical to human freedom; randomness. Helen Steward describes the proponents of this objection as claiming that "the denial of determinism merely introduces an unhelpful randomness into the causal chains that underlie our intentional activity, and that such randomness could never help us to understand how free agency is possible." (2012: 125)[104] They frequently follow the Hobartian position according to which

[103] If on the other hand it doesn't work, then perhaps we can revise the argument in order to ground responsibility along compatibilist lines by holding that the possession of powers, of a compatibilist sort, is after all sufficient to place a person under the scope of normative reasons.

[104] In her response to this objection (2012: ch.6, 7) Steward makes certain concessions to the compatibilist that I personally do not find necessary. But one theme that arises from her discussion of such concessions, such as her discussion of Susan Wolf's asymmetrical position on the conditions of praiseworthiness and

deterministic causal chains are constitutive of action and it is only a piece of behaviour that is deterministically caused by an intention to so act, that we may meaningfully call an action. In other words, freedom is inconceivable without determinism (Hobart 1934). Indeterminism seems to lead to the idea that randomness reigns supreme and that if we look close enough to what is involved in an agent's behaviour in an indeterministic universe, we will see that it ends up being a matter of blind luck. If in an indeterministic universe, behaviour is ultimately a matter of chance, we have trouble thinking how we can be responsible for anything, let alone be responsible in a normative way. Therefore, libertarian free will and normative responsibility do not seem to be compatible.

This objection rests on a whole nexus of presuppositions whose denial is one of the main themes of this thesis. It seems that even those who are sympathetic to libertarianism, (e.g. Van Inwagen 2000, Mele 2006, Steward 2012: ch. 6, 7) frequently embrace such presuppositions. But doing so, seems to me to be inimical to the position of libertarianism about free will, as well as any position that holds persons as moral agents. It seems to me that the way to respond to it lies in the dissolution of the dilemma that its advocates present metaphysical libertarians with, between embracing determinism as a necessary requirement for normatively responsible action on the one hand, and rejecting normatively responsible actions on the other.

---

blameworthiness (ibid: 142) is that there are two importantly distinct ways to understand the definition of free will as the ability to do otherwise. It is one thing to say that the agent possesses libertarian free will if she can do A or refrain from it, and quite another to say that she can do A or B or C, etc. Although B and C can be described as non-A, they can be seen as actions in their own right, rather than omissions. If I understand her point correctly, this implies that certain agents can have free will and be normatively responsible, even if they cannot literally violate an obligation. It is one thing to be able to fail to do the right thing because you are able to abstain from doing it, and quite another to be able to fail to do the right thing because you are able to do the wrong thing. This seems to help the libertarian to counter the objection from chance by saying that it doesn't have to be up to the agent to perform crazy actions –such as hitting someone who is kind to you, for absolutely no reason- to be morally responsible.

The first horn of the dilemma can be dissolved with a rejection of its presuppositions. The demands for the intelligibility and the individuation of actions that are implicit in formulations of this objection are so strict that they end up ruling out the possibility of error in action. This point needs qualification. The obvious objection is that frequently we do the best we can and it isn't good enough. We act for good reasons but doing so leads us to bad results. This is not what I mean by error. Doubtlessly, the best we can do isn't always good enough. But this does not seem to be an error on the agent's part. It may be due to limitations of knowledge and abilities, which are not under the agent's control. In cases where the person is not responsible for such limitations, these factors are sufficient to place her outside the normative standpoint with respect to the bad effect that resulted from her actions that were justified in light of her point of view [Ch. I/A.4]. The kind of error, the possibility of which is required to place someone inside the normative standpoint is that of performing unjustified actions; something akin to incontinence; weakness of will. Since normativity and the possibility of error –in that sense- go hand in hand, a view that defines 'action' as 'correct action' –in the sense of 'justified' action- [Ch. I/A.4.1-A.4.2] leaves no room for the possibility of mistakes in action. If they are not 'correct actions', they do not count as 'actions' at all. Lack of the possibility of a mistake, in the sense of an unjustified act, leaves no room for normative responsibility because normative responsibility requires a person to be able to violate a moral reason/obligation under whose scope she is. Therefore, the demand that action should be constituted by deterministic causal chains in order to ground an agent's normative responsibility seems to contradict the claim that in order to be under the scope of a normative reason/obligation one has to be able to violate it. Therefore, it is not the case

that determinism and freedom are a match made in Heaven and this dissolves the dilemma.[105]

At this point, the advocate of the objection from chance faces a dilemma. Assuming she will concede the point of our reply, she will either allow for the possibility of normatively responsible action that isn't determined or she will press the second horn of the dilemma above, as an independent claim. The first case is a concession to our point and nothing further has to be said about it. It is the second point that has to be dealt with. It seems to me that if we accept the claim that knowledge requires internalist justification, which I explicitly assume for the thesis without arguing for it, its biggest defect of this reply is that it leads to skepticism. The reason it does is that if we do not act, whether mentally or physically, we cannot be justified in thinking or inferring anything. To be justified in thinking something to be the case requires that one is able to perform a series of mental actions, such as engaging in processes of reasoning. It seems that if we presuppose this, we cannot hold that we do not act without embracing a radical epistemic skepticism.

A final word has to be said about the dialectical nature of this debate. It seems to me that upon engaging in it, libertarians about free will frequently make implicit assumptions that are inimical to their view. I propose that it is best to expose this implicit assumption and reject it, as opposed to accept it and try to find a way to make it consistent with libertarianism. The objection from chance boils down to the point that if no account can be given about why an action happened: i) as it happened as opposed to anything else happening, and ii) at the particular time it happened rather than at any other time, then the libertarian cannot account for the existence of actions as distinct from series of random and disconnected events. To accept that one has to square libertarianism with such high demands of contrastive explanation is to accept nothing less than the extension of the principle of sufficient reason to the

[105] Positions similar to the one exposed by the second horn of the dilemma will be discussed in sections B.5 and B.6 of this chapter.

causal/physical domain, at least with respect to human action. Accepting this demand seems to me to be inherently inconsistent with metaphysical libertarianism.

B.4 ***Problems with alternative possibilities***

Harry Frankfurt (1988) argues that the widespread belief that the principle of alternative possibilities is entailed by OIC is false.[106] The ability that is required for being under the scope of an obligation to perform a particular action is the ability to perform that very action; no ability to do any other action is required for the agent to be under the scope of an obligation to perform an action or refrain from performing it. Frankfurt writes:

> With respect to any action Kant's doctrine has to do with the ability to perform *that action*. PAP, on the other hand, concerns his ability to do *something else*. Moreover, the Kantian view leaves open the possibility that a person for whom only one course of action is available fulfils an obligation when he pursues that course of action and is morally praiseworthy for doing so. On the other hand, PAP implies that such a person cannot earn moral credit for what he does. This makes it clear that renouncing PAP does not require denying that 'ought' implies 'can' and that PAP is not entailed by the Kantian view. (1988: 95-96)

Frankfurt's insight has influenced many semi-compatibilists as well as advocates of asymmetrical views about moral responsibility. In their own ways, advocates of such theories reject the principle of alternative possibilities as a requirement for responsibility in the normative sense or modify it significantly. Semicompatibilists

[106] David Widerker (1991) claims that Frankfurt's claim might be true when it comes to the case of praise, but not when it comes to blame. In Ch. I/A.4.1 I argued that it is irrelevant whether it does or it doesn't because without PAP, OIC loses its prescriptive mantle.

hold that an agent can be morally responsible for performing an action or for omitting to perform an action, even if she cannot do otherwise (Frankfurt 1969, 1988; Fischer and Ravizza 1998). Asymmetrists (Susan Wolf 1980, 1990; Dana Nelkin 2008, 2011) may hold that an agent requires alternative possibilities for blameworthiness but not for praiseworthiness. In sections B.4.1-B.4.3, I will discuss various views that dissociate OIC from PAP whether entirely or in part. I will argue that this dissociation has unwanted ramifications that are far less plausible than the idea that: "'ought' implies 'can' and 'may not'"; i.e. that PAP and OIC go together.

B.4.1 *Yaffe's objection to the derivation of PAP from OIC*

Gideon Yaffe (2005) maintains that alternative possibilities are not required for being under the scope of a moral obligation.[107] He protests against Widerker (1991) and Copp (1997) that their arguments work only if they hold that for every obligation not to do something, there is a corresponding obligation to do something positive. But this, he claims, does not seem plausible. Assume one has an obligation not to kill people. One can fulfil that obligation by doing nothing. This does not imply that in doing nothing, one discharges a further obligation to do something else. Rather, it is that the agent omits action altogether and in doing so, discharges a negative obligation. Discharging negative obligations does not require the ability to do otherwise. Consequently, being under their scope does not require them either. For instance, we have an obligation to return a book we borrowed, which doesn't go out of existence when we are asleep. Being asleep –in which case, unable to act- and being under an obligation shows, according to Yaffe, that being under an obligation's scope does not require alternative possibilities. Furthermore, Yaffe thinks that we can discharge

[107] Copp (2008) holds that Yaffe's argument is better understood as an objection against the OIC principle, rather than PAP.

negative duties while being asleep. While asleep, we have no alternative possibilities open to us, yet we can discharge such negative duties.

Copp (2008) holds that this is not plausible. Obligations demand of agents, specific courses of action. The agent has to know how to meet these demands, otherwise it is unfair to hold that agent under their scope. Assume person S to have grown up in an island with just her parents, in complete isolation from society. It seems implausible to say that S discharges her obligation not to kill strangers. She has never met any stranger. It is one thing to say that she has not killed strangers and quite another to say that she has discharged her obligation not to kill strangers.

Copp thinks that Yaffe's point can be resisted if we pay attention to the temporal aspect of the relevant moral requirements. He writes:

> It would be unfair to impose an all-in requirement on a person that was not time-specific, for such a requirement would not imply anything specific about what an agent must do at a given time to avoid violating it. When fully spelled out, a fair all-in requirement would give sufficiently clear guidance to enable an informed agent to avoid violating it; it would be time-specific. (2008: 72-73)

Copp's argument that PAP can be derived from OIC seems strong. But the point of this thesis does not rest on it. As was argued in ch. I/A.4.1 the importance lies not in whether PAP is entailed by OIC. It lies in the fact that without PAP, OIC ends up not being relevant to considerations of moral responsibility because it loses its prescriptivity. So, against Yaffe's point, even if we concede that PAP is not entailed by OIC, we may still say that without it, OIC loses its significance as a condition for moral responsibility.

In addition to Copp's answer to Yaffe, the idea that one can discharge an obligation while not having alternative possibilities open to her, distorts a significant

aspect of morality; that of moral signalling. Engaging in moral assessment of each other functions as a way to signal pro tanto reliability, which is among other things, valuable and crucial for social coordination. Consistent and persistent wrongdoing is a pro tanto sign of unreliability. Consistent and persistent moral action is a pro tanto sign of reliability. Maintaining that a person can and does discharge a negative moral duty while situated in circumstances that deprive her from alternative possibilities, distorts moral signalling and hence, social coordination. Why? Because attributing 'morality' to a person in virtue of consistent discharge of negative duties while not being able to do otherwise, does not signal, prima facie, that this person is going to act morally when they have the opportunity to act. So, maintaining that someone can be moral in omitting to do the wrong thing, when they have no ability to do the wrong thing, is not a prima facie signal that this person will act morally when the possibility of doing the wrong thing is present.[108]

B.4.2 ***Semi-compatibilism: Fischer's and Ravizza's semi-compatibilism***

Another objection to PAP comes from semi-compatibilism. Semi-compatibilists maintain that moral responsibility does not require the ability to do otherwise. They develop an understanding of moral responsibility that does not require of the agent to have alternative possibilities at any moment prior to action. Their account is motivated by dissatisfaction with classical compatibilism and is broadly based on Harry Frankfurt's (1969) thought experiments –known as Frankfurt cases- which are meant to show that our intuitions tell us that human agents can be morally responsible in cases where they lack alternative possibilities. Frankfurt's counterexamples typically involve a counterfactual intervener who has ensured that the agent in question will not act against what the intervener wants the agent to do. The presence of such an intervener

[108] This idea is still in its infancy. A lot more has to be said about it to become more convincing, but it seems to me to be prima facie intuitive.

is held to deprive the agent of alternative possibilities. Assuming that the agent lacks alternative possibilities –for doing otherwise is impossible due to the counterfactual intervener- the agent goes on to do willingly what the intervener would force him to do, if the agent tried to do otherwise. Frankfurt holds that in such cases we think that the agent is morally responsible despite the lack of alternative possibilities. Thus, moral responsibility does not require alternative possibilities. And because moral responsibility is a kind of normative responsibility, to be responsible in a normative sense does not require the presence of alternative possibilities.

One of the most sophisticated versions of semi-compatibilism is the one developed by John Martin Fischer and Michael Ravizza (1998). Influenced by Frankfurt's approach, they develop a reasons-responsiveness model of normative responsibility. They maintain that despite lacking alternative possibilities, the agent can be responsible if the agent has a moderately reasons-responsive mechanism, characterized by receptivity and reactivity to reasons. The receptivity of a reasons-responsive mechanism is what enables the agent to identify reasons for actions, whereas reactivity is what enables the agent to react to recognized reasons. Receptivity to reasons, as well as reactivity to reasons may be strong, moderate or weak. According to Fischer and Ravizza, an agent is morally responsible when their reasons-responsive mechanism displays at least moderate receptivity to reasons, and at least, weak reactivity to reasons. The challenge that their view poses to our argument for libertarian free will consists in the fact that if their view of what grounds normative responsibility is correct, then it can work as a rough analogue for how one would ground normative responsibility in the case of theoretical reasoning as well. If their account works in providing us with necessary and sufficient conditions for holding a physical action as justified or unjustified, it can provide the groundwork for understanding how thought processes can be justified or unjustified.

Their denial of alternative possibilities as necessary for moral responsibility creates a tension in their account, which is evident in their treatment of action in light

of reasons. According to their account, to act is to perform a movement that is non-deviantly caused by reason-states, which in turn are conceived of as efficient causes. With respect to its reactivity to reasons, the operation of the agent's reasons-responsiveness mechanism consists in a movement being non-deviantly caused by a reason-state. And it is this reasons-responsiveness mechanism that is supposed to ground the agent's normative responsibility when its reactivity is at least, moderate and its reactivity is at least, weak.

The tension in their account stems from the fact that the way they define action leaves no room for anything less than strong reactivity to reasons. And in turn, this leaves no room for the possibility of irrationality. Strong, moderate and weak responsiveness to reasons are descriptions of how the reasons-mechanism works across time; not in particular circumstances. To characterize a reasons-responsiveness mechanism as moderate or weak is to claim that across a particular time-span, this mechanism functions mostly well or well in several cases. But, the way they define action, i.e. without entailing the ability to do otherwise, does not allow for irrational action.

To see this, let us look more closely on their causal view of action. Fischer and Ravizza conceive of reasons as efficient causes. To the extent that action is non-deviantly caused by such reasons, the agent reacts to them. Reactivity to reasons is conceived of as a passive outcome of the exertion of physical force by the reason-states. If the considerations in ch. I/A.4.2 are correct, then such responsiveness to reasons leaves no room for error. The reason is that due to their denial of alternative possibilities, their account faces structurally the same problem with the one discussed in ch. I/A.4.1 with respect to the treatment of the criteria for being under the scope of an obligation and the criteria for meeting the demand that is generated by that obligation. In Fischer's and Ravizza's account, the criteria for what counts as engaging in rational action, in the sense that is contrasted to a-rationality, are identical with the criteria of what counts as rational action, in the sense contrasted to irrationality. If reason-states cause an agent to act, then the agent is reacting to reasons. If they don't,

then the agent does not react to reasons at all, and consequently, lies outside the scope of rational assessment. So, we need to distinguish between criteria for when one is under the scope of normative rational considerations and criteria for when states of affairs obtain, that if the agent were under the scope of normative rational consideration that would favour bringing them about, the agent would discharge the normative demand generated by them in realizing these states of affairs. In a nutshell, we need criteria for when one is under the scope of normative reasons and criteria for when someone is rational as opposed to irrational.

The bottom line is that the way they define action, leaves no room for anything but strong reactivity to reasons to count as rational action. There cannot be moderate or weak reactivity to reasons across time, because any particular instance that does not consist in a movement that is non-deviantly caused by reason-states, does not count as a reaction to reasons on the first place. Furthermore, if the criteria for strong reasons reactivity are identical to the criteria of rational action, no irrationality seems possible. If there is no possibility of violating an obligation of a demand generated by a reason, it seems that both that obligation and that reason are not normative. Thus, it seems that the semi-compatibilist account of moral responsibility cannot ground responsibility of a normative sort and a fortiori, cannot provide a rough analogue for understanding epistemic normativity. If it cannot give us a satisfactory account of what counts as a morally justified or unjustified action, then it cannot provide the structure for a theory of what counts as a justified or unjustified thought process.

B.4.3 ***Asymmetrical views of moral responsibility***

The last objection of this kind that will be discussed stems from asymmetrical views of moral responsibility. Asymmetrists about moral responsibility claim that the conditions for praising a person are not symmetrical with the conditions for blaming her. Susan Wolf (1980, 1990) and Dana Nelkin (2008, 2011) hold that blameworthiness

requires alternative possibilities, whereas praiseworthiness does not. According to Nelkin, the asymmetrical view is "the view that one is responsible for an action if and only if one acts with the ability to recognize and act for good reasons." (2011: 3) The twist comes in her claim that such an ability to recognize good reasons and act for them is compatible with determinism.

This view may represent an objection to our argument if it is employed to claim that the conditions for being justified in one's thinking are not symmetrical with the conditions for being unjustified in one's thinking. In a nutshell, one can be rational in a normative sense –which implies meeting prescriptive demands, whether in thought or physical action- despite being incapable of being irrational. This can be seen directly in Wolf's Hobartian treatment of the notion of an undetermined action. Discussing an agent who is not determined to act by his best reasons, she writes:

> One might think such pieces of behavior should not be classified as actions at all –that they are rather more, like spasms that the agent cannot control. If they are actions, at least, they are very bizarre, and an agent who performed them would have to be insane. Indeed, one might think he would have to be insane if he had even the ability to perform them. For the rationality of an agent who could perform such irrational actions as these must hang by a dangerously thin thread. (1980: 153)

Roughly speaking, this amounts to the familiar by now claim that if an action is not determined, it cannot be classified as an action; at least a sane one. Translated into the epistemic domain, this amounts to claiming either that if a process of thinking isn't rational –in the sense of rationality as a success term- it is not a process of thinking at all, or that it is not a process of sane thinking. The first horn denies the possibility of irrationality whereas the second reduces all instances of irrationality to instances of insanity. It is the first that represents a challenge to our argument and it has been discussed in Ch. I/A.4.2. If the possibility of mistakes in reasoning is to be upheld,

views that end up defining ‘engaging in reasoning’ as ‘engaging in reasoning correctly’ must be rejected. The second horn of the dilemma does not create any problem to our argument because ‘sanity’ does seem to be a success term. Nevertheless, to maintain that all instances of irrationality are instances of insanity does not seem to be plausible. Conceived of along these lines, someone who believes that we have the ability to do otherwise has to be committed to the possibility of insane actions.

This asymmetric account implies that it is sensible to claim that we can win in a game that we cannot lose in. Strictly speaking, this amounts to the denial of the claim that in order to be under the scope of a prescriptive reason –whether practical or epistemic- one has to be able to violate the demand constituted or generated by that reason. This seems wrong and entirely ad hoc. If the considerations laid out in Ch. I/A.4.1 are correct, then there is no sense in claiming that we can be under the scope of a prescriptive consideration if we cannot violate it. But perhaps the asymmetrist won’t be convinced. To engage in a debate with the asymmetrist, I will proceed to discuss two objections to it that do not rest on the aforementioned considerations.

The asymmetrist position seems entirely ad hoc. First, it is not entirely certain why we should accept this intuition as opposed to reject it in light of the asymmetry. Second, it is not entirely certain that the intuition, as presented by Wolf, withstands scrutiny. In fact, it seems to me that an entirely different interpretation is much more plausible and does not lead to the unfortunate effect of treating moral praiseworthiness and moral blameworthiness, asymmetrically. To substantiate the first claim, I will discuss Wolf’s intuition and to motivate the second, I will reconstruct her argument and claim that the asymmetrical interpretation rests on an ambiguous treatment of the notion of ‘appraisal’. In a nutshell, the fact that two things can be appraised, does not straightforwardly imply that they can be appraised in the same way. Merely evaluative appraisal is not the same as normative appraisal.

Wolf holds that the view that praiseworthiness requires alternative possibilities does not do justice to certain intuitions about it. She presents a thought experiment to

motivate the view that we generally think that agents whose actions are determined can still be morally praiseworthy. She writes:

> Two persons, of equal swimming ability, stand on equally uncrowded beaches. Each sees an unknown child struggling in the water in the distance. Each thinks "The child needs my help" and directly swims out to save him. In each case, we assume that the agent reasons correctly –that child does need her help–and that, in swimming out to save him, the agent does the right thing. We further assume that in one of these cases, the agent has the ability to do otherwise, and in the other case not. (1990: 81-82)

Wolf claims that both agents are equally praiseworthy despite the fact that one could have chosen not to try to save the drowning child, whereas the other was necessitated to try to do so. Given the fact that we would praise them both, it follows that one's lack of alternative possibilities is not relevant to praising her. Both are equally praiseworthy because both acted in a way that displays sensitivity to good reasons.

The first objection against this view is quite straightforward. We have lots of intuitions and some of them are inconsistent. Why should we choose to embrace this asymmetrical intuition as opposed to reject it? The obvious response would be that as a matter of fact, the asymmetrical position is the one that allows us to achieve greater coherence among our moral intuitions. In this sense, if we employ the method of reflective equilibrium, we may find out that parting with the principle of symmetry of the conditions of praiseworthiness and blameworthiness, leads us to achieve a wider coherence among our moral beliefs in general. The response to this seems to me to lie on the second objection that can be raised against the asymmetrist position. In effect, if this is the defense against rejecting Wolf's intuition, the point may be raised as to whether a modified version of it, allows us to achieve even greater coherence among our moral beliefs.

It seems that the asymmetrical interpretation of the intuition we have about the praiseworthiness of both swimmers is neither the only nor the best one. An alternative interpretation is also available, according to which both swimmers are praiseworthy but in a different way. To show this, we will have to sketch an outline of Wolf's argument and show that her point does not follow. We can formalize her argument in the following way:

i) Both S1 and S2 save a drowning child in their respective worlds W1 and W2.
ii) S1 in W1 lacks the ability to do otherwise, whereas S2 in W2 possesses it.
iii) Both S1 in W1 and S2 in W2 are praiseworthy for saving the drowning child after intentionally trying to.
iv) Therefore, the ability to do otherwise is not essential for praiseworthiness.

Unfortunately, iv) does not necessarily follow from i), ii) and iii). The reason is that premise iii) is ambiguous. Those who do think there are symmetrical conditions for both praiseworthiness and blameworthiness –for acting in the right way, as well as the wrong way- may concede that both swimmers are praiseworthy. What we can plausibly reject is that they are praiseworthy *in the same way*. And in doing so, we may say that such praise that we express towards the swimmer who lacks alternatives is not a form of praise that is best understood in a normative context, but rather a kind of praise that is best understood in a merely evaluative one. As Derk Pereboom writes:

> Praise can at times simply be an expression of approbation or delight about the goodness of the actions or accomplishments of another, and then even if it is not fundamentally deserved, it would still be right to praise. (2014a: 102)

Once we maintain that normative praise is distinct from merely evaluative praise, the asymmetry between moral praiseworthiness and blameworthiness that asymmetrists

appeal to turns out to be the asymmetry between moral praiseworthiness in the evaluative sense and moral blameworthiness in the normative sense.

The asymmetrist could bring forth certain terminological considerations that speak in favour of the view that we do not 'appraise' what is not a result of action, but we use other terms such as 'fitting to value'. For instance, it is not customary to speak of our lungs as praiseworthy if they function well. This point is correct, but it is largely terminological. In this sense, we may reject premise iii) again, on different grounds; on grounds that the swimmer who has no ability to do otherwise, is not 'praiseworthy'. Yet again, this modification can be incorporated in a broader context that is sensitive to the difference between the merely evaluative and the normative. It seems that with this move, we can achieve an even greater coherence among our moral beliefs.

The force of the asymmetrist interpretation of the intuition arises from the conflation of two standpoints that we employ in ascribing abilities to ourselves and others. The first standpoint consists in the ordinary standpoint we employ in our everyday ascriptions of abilities, which arises from the practical necessities, demands and uncertain conditions of an everyday setting. Intuitions that stem from this standpoint are intuitions that are based on a framework that is neutral with respect to the potential actuality of determinism or even hostile to it, if we assume that most people have the belief that they are free in a libertarian sense. The second standpoint consists in the standpoint that we may employ in thinking about the ramifications of determinism, or lack of alternative possibilities of action, which operates given the assumption of the truth of determinism. Intuitions that stem from this standpoint do not face the uncertainty that is inherent in the other standpoint. It seems that the asymmetrist, in discussing the ramifications of determinism –via discussing the ramifications of a lack of alternatives- bases her claim upon intuitions that stem from another standpoint, without noticing the difference. This point will be developed more in Ch. IV/A.2.3 where it will be argued that compatibilists about free will and determinism fall into the same trap in giving a conditional analysis of the ability to do otherwise.

B.5 *Lack of voluntary control over belief*

Another objection to our argument can be found in doxastic involuntarism. Its core insight seems to be captured best in Hume's quote that: "belief is more properly an act of the sensitive, than of the cogitative part of our natures." (1985[1739]: 234, Book I, part I, section I) It has been argued by doxastic involuntarists, such as Bernard Williams (1973: 136-151), William Alston (1988) and David Owens (2000) that we do not have any sort of control over our beliefs that is appropriate to ground epistemic responsibility.[109] According to Alston (1988: 260), several of our expressions, like: 'make up your mind' seem to commit us to the idea that we can control what we believe. But this is illusory. This posits a problem to those who aim to understand inquiry in terms that are normative. Epistemic deontology and especially the deontological conception of epistemic justification is misguided in that, although 'ought' does imply 'can', there is insufficient voluntary control over our beliefs. Therefore, since we lack voluntary control over belief, there are no epistemic 'oughts'.

The main claim that our critic is committed to is that we cannot believe at will.[110] No matter how much we try to convince ourselves that we have three eyes, or that we were born four centuries ago, we just cannot. So our critic generalizes from these cases and draws the conclusion that we cannot ever believe anything at will. If we cannot believe at will, we cannot be assessed in terms of whether we think something correctly or incorrectly, at least in the deontic sense. Therefore, believing cannot be assessed in

[109] Some prominent defences of doxastic voluntarism are given by Carl Ginet (2001) and Matthias Steup (2008).

[110] Ironically, many doxastic involuntarists are empiricists. The irony consists not so much in the fact that they move from observations of finite cases to a generalization. This is understandable. Instead, it consists in the fact that they proceed into claiming that the contrary is impossible. Seen through the perspective of Hume's criterion of what constitutes a 'matter of fact', one starts wondering whether the position of doxastic involuntarism is anything more than the expression of an attitude towards the mind and its faculties.

terms of what we should and shouldn't believe, or in terms of what we have reasons to or not to believe. The implication of this view for our argument consists in the fact that if it were true, whether the person's thoughts move in such a way as to reach the content which would be the one demanded according to the epistemic deontologist, or not, would not be up to the agent, and therefore, the agent would not be normatively responsible for it. Since we are not free to choose our beliefs, there are no epistemic 'oughts'.

This presents us with a serious challenge. To answer it sufficiently would require a much lengthier treatment than the one this thesis allows for. But two things can be said. The first reply focuses on the suppressed quasi-associationist presupposition of the doxastic involuntarist, according to which, cognition reduces to a series of acts of believing. The second reply focuses on the way in which the doxastic voluntarist enters the debate.

The first reply against the objection from doxastic voluntarism is that it does not seem to be the case that the entirety of our cognition reduces to a number of distinct episodes of believing. The phenomenology of belief involves a characteristic sense of endorsement, which is absent from other epistemic endeavours such as engaging in hypothetical reasoning. Surely, part of drawing inferences from premises whose truth you do not accept or reject, at least for that time, involve a belief in that a particular conclusion follows from some premises. But it does not exhaust everything that is going on, epistemically speaking. Conceived of along these lines, we may concede the criticism that belief is not subject to our will and deny that what follows from that is the impossibility of epistemic agency. A lot more would have to be seen to be involuntary, that does not seem to be, in the same sense that our believing obvious falsehoods[111] does seem impossible to the doxastic involuntarist.

[111] Robert Lockie argues that the examples invoked by the doxastic involuntarist have an intuitive pull because they are frequently centred on cases of obviously false beliefs that someone is asked to believe immediately prior to engaging in a belief-forming process (2018: 49-52). He writes: "the literature repeatedly argues for

The second reply targets the methodology of some doxastic involuntarists. Just like we mentioned that one can engage in the topic of a priori justification as either holding skepticism as a live option, or as a non-skeptic, so a doxastic involuntarist may enter this discussion as one who won't part with doxastic involuntarism, no matter what considerations may be given against it, or as someone who maintains the possibility of parting with it.

The first kind of engagement is dialectically ineffectual and unpersuasive. Everyone can enter into a debate while holding that a particular position may be guarded against any consideration because it may be used as a ground for rejecting all considerations that can be levelled against it. Different thinkers have lots of different intuitions. For instance, many have the intuition that human beings have free will, whereas others have the intuition that the universe is rigidly necessitated. Entering the debate as someone who conceives of doxastic involuntarism as a position which may be used to counter all possible criticisms against it, is the equivalent of a fatalist entering the respective debate while claiming that fatalism is strong enough to reject all considerations against it, or even the way in which someone who believes in libertarian free will may claim that belief in libertarian free will is epistemically sufficient to ground the rejection of all criticisms that may be raised against it. A doxastic involuntarist would definitely find these positions unsophisticated. But if that is essentially the way she treats doxastic involuntarism, there seems to be no reason to treat doxastic involuntarism in a privileged fashion.

The second kind of engagement seems to be more apt for genuine philosophical discussions. If a doxastic voluntarist enters the discussion in such a spirit, then we may reply that the considerations laid above, especially those in the first two chapters, may suffice to give reasons for turning the doxastic involuntarist's modus ponens against

involuntarism on the basis that we cannot choose to believe unjustified and obviously untrue things." (ibid: 51)

epistemic ‘oughts’ into a modus tollens against doxastic involuntarism. The doxastic involuntarist’s modus ponens consists in arguing for the lack of epistemic ‘oughts’ on the basis of accepting the OIC principle, as well as doxastic involuntarism. If considerations about the importance of normativity in practical and theoretical reasoning have any merit, then we may use them as an argument against doxastic involuntarism.

A final word is due before we proceed to discuss the last objection. The doxastic involuntarist position does present us with a challenge. It does not seem to me to be plausible to deny epistemic normativity altogether. Nevertheless, an interesting question lies on how to incorporate the claim that right now that I am writing, I cannot choose to believe there is a purple book flying around. But to proceed to claim that there is no epistemic agency whatsoever, seems to invite a radical form of epistemic skepticism. Doxastic involuntarists do not seem to be skeptics. An epistemic skeptic who is a doxastic involuntarist seems to me to be a contradiction in terms. Thus, it seems that if an unrestricted doxastic involuntarism –one which would apply to the entirety of our mental life- leads to epistemic skepticism, it is in the interest of doxastic involuntarists who are not skeptics, to try to modify their position in a way that blocks this implication. Prima facie, it seems there is room for both claiming that epistemic agency exists and that some aspects of it, perhaps most, are not voluntary.

B.6 ***The possibility of a-rationalism***

The final objection concerns the possibility that a-rationalism is true, yet rationally unaffirmable. In this context, a-rationalism consists in the position that all aspects of our lives are a-rational. This objection has led most advocates of transcendental arguments against determinism (e.g. Jordan 1969; Boyle, Grisez & Tollefsen 1976; Lockie 2018) to claim that the argument’s reach does not suffice to ground the existence of libertarian free will; all it can do is to show that we have to see the world through the prism of a relation between rationality and free will. But this

does not allow us to move one step further to the claim that we, in fact, possess libertarian free will. James N. Jordan writes:

> That every thought is so conditioned one might entertain as a guiding principle of scientific inquiry or as a generalization from controlled experiments, and nothing in the argument that determinism is self-defeating could possibly show that he is ill-advised or wrong. The argument does not assume or imply that determinism is false. All the argument does is to raise, and answer in the negative, the question whether, if every thought were so conditioned, we could have recognizably good reasons to entertain anything, including the methodological principle or empirical generalization that every thought is so conditioned. (1969: 63-64)

According to our critic: a-rationalism could be true, without us having any reason to believe it. So, the fact that it would never be rational to give up libertarian free will would not show that determinism is false.[112] All that our argument does is to show that lack of free will is inconsistent with substantive rationality.

As hinted above, there are two ways to interpret the claim that something may be true yet rationally unaffirmable. The first interpretation holds that X may be the case and we cannot rationally affirm it, although there are other things which are the case and which we can rationally affirm. The second interpretation holds that X may be the case and we cannot rationally affirm it because we can never rationally affirm anything. The first interpretation is not consistent with radical epistemic skepticism because it implies the possibility of knowledge and rational affirmation. The second interpretation seems to be the very expression of radical epistemic skepticism. So long as we focus on the first interpretation, complete a-rationalism cannot be what is signified by X. The

[112] This seems to rest on a conception of truth as verification-transcendent.

reason is that it is a contradiction to hold that although we can rationally affirm certain things to be the case, it may still be true that we are entirely a-rational and that we cannot rationally affirm the latter. The former rules out the latter. The latter interpretation involves a radical form of epistemic skepticism. If we maintain that radical epistemic skepticism is to be avoided, then we can reject the second interpretation.

Conclusion

In this chapter, a rationalist argument for a libertarian conception of free will is presented. Its chief difference from transcendental arguments against determinism is that it rests not with claiming free will to be a presupposition of thought, but due to its rationalistic framework, promises to go beyond that. It begins with the claim that substantive rationality, which is essentially connected to normative reasons generates rational demands. Normative reasons and the demands generated by them raise the question of their scope. It is argued that ultimately, the kind of power that enables us to think and act in light of reasons is that associated with libertarian free will. Without libertarian free will, we cannot be under the scope of normative reasons, whether epistemic or practical. Thus, without free will we cannot be rational. In conjunction with the presuppositions of the thesis, which involve commitment to the falsity of radical epistemic skepticism as well as to the claim that knowledge requires internalist justification, we conclude that to the extent that we can be rational or irrational, as well as acquire knowledge, we possess libertarian free will.

In section B, six objections are raised and discussed. According to the first objection, premise II) can be denied because normative facts are entirely reducible to evaluative facts. Thus, no appeal to alternative possibilities is needed to render it 'justified' for an agent to do something. The fact that it is good suffices to ground the rightness of the action, regardless of other considerations about the agent's abilities or knowledge. According to the second objection, premise V) can be denied because it

ultimately rests on a libertarian understanding of premise II). Advocates of this objection may agree with premise II) but give a compatibilist reading of it. This is the main objection I deal with in this thesis and it will be ultimately answered in Ch. IV. According to the third objection, indeterminism is worse for freedom and control over action due to the randomness implied an indeterministic setting. According to the fourth objection, premise II) can be denied because alternative possibilities are either not required for normative responsibility [B.4.1-B.4.2] or not required for praiseworthiness [B.4.3]. According to the fifth objection, we can never be substantively rational, in the way that the argument implies, due to our lack of control over our beliefs. Finally, according to the sixth objection, the argument's conclusion doesn't follow from premise VII) because the fact that a position is rationally unaffirmable does not imply that it is false. It may still be true.

Against the first objection, it is maintained that the only way in which it can present an objection to our rationalist argument is by commitment to certain questionable positions about norms of reasoning. According to such positions, it is fair to treat someone as being under the scope of normative reasons, regardless of facts about her knowledge and abilities. Against the second objection, it is maintained that no compatibilist account of OIC and PAP is tenable. Such a project rests on the plausibility of a conditional analysis of the ability to do otherwise, the tenability of which is questionable. The point will be developed more in Ch. IV/A.2.3. Against the third objection, it is maintained that the dilemma its advocates present us with can be dissolved. Furthermore, to accept its terms is to implicitly accept the untenability of libertarianism about free will. Against the positions that question PAP's relevance to one's normatively responsibility, whether entirely or in part, it is maintained that their accounts are ultimately incompatible with the position that in order to be under the scope of a normative reason, one has to be able to violate it. Otherwise, this reason loses its normativity. In one way or another, thinkers who downsize the importance of alternative possibilities end up defining 'acting' with 'acting correctly' and thus, face

problems in accounting for the possibility of error. Against the asymmetrists, it is further argued that their position is unmotivated. Any merit this position has with respect to aiding us to achieve coherence among our moral beliefs, can be surpassed by embracing the distinction between merely evaluative and normative appraisal. Against the fifth objection, it is argued that it can be understood in an unrestricted sense, if it applies to the totality of our cognitive attitudes, as well as a restricted one, if applied to some of them but not all. The first one invites radical epistemic skepticism, whereas the latter presents no ultimate threat to our argument. Against the sixth objection, it is claimed that two interpretations can be given against the claim that something that is rationally unaffirmable can be the case. One interpretation involves the claim that something, X, can be rationally unaffirmable as opposed to other things which are rationally affirmable. The other interpretation involves the claim that X is rationally unaffirmable because everything is rationally unaffirmable. If interpreted along the lines of the first interpretation, claim X cannot be the claim that complete a-rationality is possible, at pains of contradiction. It is inconsistent to maintain that we are entirely a-rational and that we can rationally affirm some things to be the case. If interpreted along the lines of the second interpretation, the claim amounts to an expression of a radical epistemic skepticism. Although this will not convince the radical epistemic skeptic, it is not intended to. This thesis is addressed to those who do not embrace the possibility of radical epistemic skepticism.

Chapter IV: Can there be a compatibilist solution?

In this chapter, I argue that the prospects for compatibilism between free will and determinism are dim. In section A, I discuss several conceptions of free will and argue that the most adequate definition of free will holds it to be the ontologically fundamental/irreducible specific power to actively do otherwise. Active powers are conceived of as two-way powers, whereas passive powers are conceived of as one-way powers. Two rival definitions that conceive of free will as a passive power will be discussed [A.2-A.2.2] and countered [A.2.3] by means of a criticism of conditional analyses of the ability to do otherwise. An alternative conception remains that understands free will as the reducible specific power to actively do otherwise. It is maintained that this concept secures alternative possibilities, but in a way which is difficult to see how they are or can be alternative possibilities of action. In section B, I discuss two versions of determinism and claim that each of them is incompatible with free will, although for different reasons. Robustly causal determinism presents problem to free will because it is committed to robustly causal necessitation of action, whereas neo-Humean determinism seems to either eliminate agency or just change the subject.

Section A: ***What would make free will impossible?***

In this section, I argue that the most adequate definition of free will holds it to be the irreducible specific power to actively do otherwise. When referring to the past, one had free will with respect to action A, if one had an irreducible power to actively do other than A, in those exact circumstances prior to her engaging in A-ing. When referring to the present or future, one has free will if one has an irreducible power to actively perform a particular action, or refrain from performing it, whether that involves intentionally omitting action altogether or performing a different action. In section A.1, I will explore the notions of ‘ability’, ‘doing’, as well as the notion

'otherwise'. In section A.2, I will discuss and criticize conditional analyses of ability in both their reformed and classical form. I will consider a new objection against such attempts, according to which, their advocates neglect human situatedness and end up treating the circumstances in which an agent is situated as inessential. This will constitute a rebuttal to the objection from the compatibilist conception of both the OIC principle and PAP [Ch. III/B.2] that can be raised against the main argument of the thesis [Ch. III/A]. The section concludes that the criterion for judging whether a position is inconsistent with free will consists in whether it is committed to the denial of at least one of its constituent elements. Whether determinism is incompatible with free will is going to be the subject of this chapter's section B.

A.1 ***The concept's constituent notions***

Free will is usually defined as the ability to do otherwise.[113] Despite its initial straightforward appearance, this notion is quite complex. It involves three constituent concepts which are quite complex in themselves. In that sense, the notion of free will is like a tree with three main branches, each of which grows its own twigs and leaves. The main three branches are represented by the notions of 'ability', 'doing', as well as the notion of 'otherwise'. Abilities can be seen as fundamental or reducible elements of the universe, as well as general capacities or specific powers.[114] What we do can be

[113] Not everyone conceives of free will in this way. Source views of free will conceive of it as a power that the agent has over action, if the action is initiated by the agent as opposed to being caused by factors of the agent's environment. For instance, those who reject the importance of alternative possibilities for agency and moral responsibility, frequently end up defining human freedom as a state in which there is a right sort of connection between the agent's states (e.g. Frankfurt 1971, Watson 1975). These are called 'mesh views' and they "account for free will in terms of a well-functioning harmony between different elements within an agent's psychic structure." (McKenna and Pereboom, 2016: 207)

[114] Frequently, participants in the debate charge each other with focusing on the wrong kind of ability to ground claims about the ability to do otherwise. For instance, Whittle (2010) argues that new dispositionalists have failed to show how free will can be reconciled with determinism because their account works for global

interpreted in both an active and a passive sense.[115] Depending on whether we have control over the exercise of such capacities, what we do can represent the exercise of one-way or two-way powers. The notion of 'otherwise' is an interesting notion, because it is frequently neglected. It is fundamentally relational in that 'otherwise' is elliptical for 'otherwise than'. Understanding what stands after 'than' in that clause, sheds more light on our understanding of free will. Each interpretation of each constituent notion of 'the ability to do otherwise' can be combined with each interpretation of the rest into forming a unique meaning of the concept of free will. It will be my purpose in this section to argue that not all ways of understanding the notion of free will are of equal value and that a particular one captures more adequately the kind of power we refer to in discussions of free will, determinism and moral responsibility. I will say more about each constituent concept and then proceed to settle with four candidate definitions of free will.

Abilities are always abilities to *do* something. They can be described as modal properties (Vetter and Jaster 2017).The active or passive nature of abilities is a direct outcome of the active or passive nature of the 'doings' that we are capable of. To ascribe an ability to a person is to ascribe a power to her. Such powers can be

abilities instead of local abilities, which are the relevant abilities in the debate concerning free will and determinism. Some other writers who point out that the notions of ability that are used by incompatibilists and compatibilists, are Mark Schlosser (2017), Alex Grzankowski (2014) and Michael Fara (2008) who writes: "So, the incompatibilist might contend, even if determinism is compatible with the ability to act otherwise in my sense of 'ability' it is not compatible with the kind of ability to act otherwise that is required for free action." (2008: 863) Fara's comment is misleading since it implies that there is no disagreement over the notion of action. Free action isn't necessarily meant in the same way as free will. Compatibilists like Hobbes (1982[1651]) would hold that although the latter is senseless, the former is always understood negatively, as the absence of constraints.

[115] Frequently, 'doing' and 'thinking' are contrasted with each other. This has the unfortunate effect of making it impossible to view thought as active in a significance sense.

understood in terms of: i) the place they occupy in the ontological hierarchy, as well as, ii) whether they are said to be possessed in a specific or a general sense.

The ontological status of powers refers to the place they occupy in the world as well as in our theories. If they are among the fundamental entities that exist, entities that cannot be reduced to more fundamental ones, then powers are fundamental themselves. The universe contains powers which cannot be understood in terms of something more basic. In that case, the statement "This person has the power to X" cannot be analyzed in terms of a sentence that does not directly or indirectly contain the notion of 'power'. To claim that we analyze it in terms of the sentence "This person has the ability to X" is to miss the point. For how do we understand 'ability'? If we understand abilities in terms of 'powers' and 'powers' in terms of 'abilities' we are not giving an analysis of statements about powers, but a non-reductive account of them.[116] Those who claim that powers do not belong to the fundamental entities of the universe, claim that statements about powers can be reductively analyzed in terms that refer to more fundamental entities, which according to David Lewis, a proponent of such a view, are the:

> vast mosaic of local matters of particular fact, just one little thing and then another. (...) We have geometry: a system of external relations of spatiotemporal distances between points. Maybe points of space-time itself, maybe point-sized bits of matter or aether or fields, maybe both. And at those points we have local qualities: perfectly natural intrinsic properties which need nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. There is no difference without difference in the arrangement of qualities. All else supervenes on that. (Lewis, 1986: ix)

[116] An influential statement of this view can be found in Thomas Reid (1969[1788]: Book I)

The other important question about powers and abilities concerns the manner of their conception, evident in various distinctions [117] that feature in contemporary discussions of abilities. For this discussion, I will refer to this distinction as that between general and specific abilities. General abilities can remain with the agent even when they cannot be exercised. A world class chef retains the general ability to cook, even in circumstances where she lacks the specific ability to cook. For instance, a world class chef has the general ability to cook, even in circumstances where she doesn't have access to any cooking ingredients. The criterion for an attribution of a 'general' ability to an agent is the presence of a capacity, irrespective of the relation between that capacity and the circumstances in which its bearer is situated in. The criterion for the attribution of a 'specific' ability involves reference not just to the presence of a capacity, but also to opportunities, as well as the conditions that enable its exercise. To recapitulate, to ascribe the general ability to X to a person requires the possession of a capacity to do X, irrespective of whether the agent is situated in circumstances which give her an opportunity to exercise it or not. To ascribe the specific ability to X to a person requires the possibility of its exercise, in that it requires both the possession of the capacity to do X, and the actuality of conditions that enable its exercise in those very circumstances in which the agent is situated.

The notion of 'doing' can be understood in many ways, but an adequate exposition of the topic has to do justice to the categorical difference between active doing and passive doing. Since abilities are abilities to do something, the distinction between active and passive doing informs directly our understanding of the distinction between active and passive abilities. We use the term 'doing' to cover things like engaging in overt physical action, as well as sleeping. The former seems, at least prima

[117] The terminology of the distinction varies according to the author. Mele (2003) distinguishes between general and specific abilities, whereas Lewis (1981) distinguishes between 'weak' and 'strong' abilities. Clarke (2009, 2015) and Vihvelin (2013) distinguishes between 'narrow' and 'wide' abilities, whereas Ann Whittle (2010) distinguishes between 'global' and 'local' abilities.

facie, a paradigmatic case of action whereas the latter seems as a paradigmatic case of inaction. A traditional understanding of such a distinction understands active doing as effecting a change in the environment, and passive doing as the ability to be acted upon in certain ways. Koons and Pickavance (2015) hold that:

> One thing has an active power when it is disposed to bring about a certain kind of change in other things. (…) A thing has a passive causal power when it is disposed to undergo change of a certain kind under certain circumstances. (2015: 63)

This conception of the distinction between active and passive powers does not seem to be the relevant distinction when it comes to the question of free will. The reason is that under such an understanding of 'active' powers, it is hard to imagine any concrete entity that fails to satisfy it. In this sense, the hammer that descends upon the sword that is held on the anvil, has active powers; the power to turn a molten mass of steel into a weapon, as well as the power to bend, break and reshape swords, shields and spears. Yet, it would take the most animistic of blacksmiths to hold the hammer as endowed with free will.

This notion of activity does not secure the desired element of control that is supposed to go along with the notion of free and responsible action. If we understand human beings as active, only in the sense of being endowed with the power to effect a change to objects of their environment and fellow human beings, the only notion of control that would be meaningfully ascribed to us would be what Fischer and Ravizza (1998) call 'guidance control'. Yet, it would not be 'regulative control', which is the kind of control they hold to be involved in the notion of free will. Throughout the remaining discussion, I will use the notion of 'control' to refer to what Fischer and Ravizza refer to as 'regulative control', which involves the ability to do otherwise in the exact same circumstances; not different ones.

A more adequate conception of the distinction between passivity and activity in the exercise of powers is that between one-way and two-way powers. One-way powers are those whose manifestation is guaranteed, given specific stimulus conditions. Given specific conditions, the power is exercised. In discussing one-way powers, Helen Steward writes:

> All sorts of objects have powers, e.g. magnesium has the power to dissolve in acid, my printer has the power to print pages of text, my heart has the power to pump blood around my body. But none of these things –magnesium, printer, heart–has, at the same time, the power *not* to exercise these other powers, once those conditions for their realization are present (for this reason, indeed, it is much more natural to speak of these one-way powers being realized than it is to speak of them being exercised). (2012: 155-156)

Two-way powers are those whose manifestation is not guaranteed by anything other than the agent's choice or act; in contrast to the agent who possesses only one-way powers, the agent who possesses two-way powers has control over whether she will exercise these powers or not. As Maria Alvarez writes:

> for my ability to cook omelettes to be a two-way power it must be up to me whether I exercise the ability when I have the opportunity to do so – that is, it must be up to me whether or not I cook omelettes then. (2013: 109)

To put everything back together, the traditional distinction between active and passive powers captures the intuitive difference between that which is affected by something and that which affects it. But activity in this sense, is not sufficient to capture the element of control over action that is a desideratum of an adequate conception of free will. To capture such control, it is not enough to focus on the presence of powers to

effect a change to their environment. Instead, we should focus on the manner in which these capacities are exercised in the specific circumstances the agent is situated in. If someone has powers but lacks control over whether they are exercised or not, that person cannot be said to possess free will.[118] The reason is that if the agent has no control over whether such powers will be exercised or not, their exercise or lack of exercise is a passive outcome which is either causally necessitated to take place or probabilified by factors which are ultimately beyond the agent's control. To possess free will, a situated agent must have some control over whether her powers will be exercised or not.

As mentioned above, 'otherwise' in the 'ability to do otherwise' is elliptical for 'otherwise than'. Whenever someone claims that one has the ability to do otherwise, we may ask: otherwise than what? Otherwise than what one did or will do in the future, or otherwise than what is impossible? If the latter, everything that happens is an instantiation of a power of things to happen otherwise than what is impossible to happen. But such a conception of free will would be obviously vacuous. We would be trivially free, to the extent that there would be no cases in which we could lack the ability to do otherwise. That is why, it is important to treat past, present and future actual actions as the answer to the question: 'otherwise than what?' Thus, so long as the notion 'otherwise' is concerned, to claim that one had the ability to do otherwise with respect to an action A, is to claim that this person could have done otherwise than A in those specific circumstances in which she was situated at the time before action. When we talk about actions that haven't been performed yet, to have the ability to do otherwise is to have the ability to perform more than one action, whether in the sense of being able to perform an action or refrain from it, on in the sense of being able to perform either of distinct actions.

[118] To the semi-compatibilist this may seem to be a quick move. It rests in the criticism of semi-compatibilism in Ch. III/B.4.1-B.4.2.

The meaning of each constituent notion of the complex notion of 'the ability to do otherwise' in combination with particular meanings of the rest, creates a unique conception of free will. For clarity of exposition, I will just mention the following four conceptions of free will. These are:

FW (c1): The fundamental/irreducible specific power to actively do otherwise

FW (c2): The fundamental/irreducible specific power to passively do otherwise

FW (c3): The non-fundamental/reducible specific power to actively do otherwise

FW (c4): The non-fundamental/reducible specific power to passively do otherwise

As defined above, the notion of a passive doing, or an ability to passively do something, is to be understood as an ability whose exercise consists in a passive reaction to a stimulus, regardless of whether it results in a change that is internal or external to the agent; it consists in the necessary or probabilified outcome of an agent's being acted upon by something. According to this view, a being is acted upon in such a way by the forces of its environment and reacts in a way which is determined or made more likely by these forces. This ability is trivially attributed to almost any concrete entity. In this sense, action is the movement of a billiard ball which is hit by the billiard stick. The billiard ball can be moved towards one direction, as well as towards the opposite direction, yet whether it will do either, is not up to the billiard ball. It is ultimately dependent on how the ball will be hit by whatever hits it. It is hard to

imagine anything concrete that doesn't satisfy this condition. A puppet might fall on one's head, causing that person to be hit, yet effecting such a change is not up to the puppet. Such abilities are passive because their bearers have absolutely no control over whether they will be exercised or not. Although such agency has a causal element, it has nothing more. Bearers of passive abilities have the ability to effect a change in their environment –which in that sense represents an activity- but they do not have any saying whatsoever as to whether such abilities will be exercised or not. Since free will is supposed to capture a kind of agency that is distinct from the agency of beings such as stones, leaves and puppets, I shall hold that FW (c2) and FW (c4) are straightforwardly inadequate as a characterization of free will. The locus of action is not to be found in what is done to us, but in what we do; hence, we talk of the ability to *do* otherwise, not the ability to be *acted upon* otherwise.

Advocates of such conceptions of free will –FW (c2) and FW (c4) - will protest that such conceptual clarifications are not enough to discredit their position. They will hold that the distinction between activity and passivity in terms of one-way and two-way powers is influenced by incompatibilist intuitions that they do not share. They will hold that activity and passivity are to be understood in a different way and that one can secure activity with one-way powers without being a semi-compatibilist, if one provides a conditional analysis of the ability to do otherwise. To counter them, I will give an argument against conditionalist compatibilism which targets the very idea that an agent can do otherwise in a deterministic universe if a conditional statement of a specific kind, is true of the agent [A.2.3].

Apart from FW (c2) and FW (c4), someone can embrace FW (c3). According to this conception of free will, the ability to do otherwise is a reducible power, whose exercise is, nevertheless, up to the agent. This is a neo-Humean understanding of powers, which has recently been appealed to in the position that Helen Beebee and Alfred Mele (2002) and Bernard Berofsky (2012) call 'Humean compatibilism'. According to this view, the universe can be deterministic, in a qualified sense, without that creating a problem for our ability to do otherwise. As the universe unfolds, there

is nothing that causally necessitates an action to take place because laws of nature do not govern the universe (Beebee 2000). They rest upon what takes place. In this sense, laws of nature are partly up to us. Apart from considerations that concern the objection from chance[119] with which this chapter will not be occupied, I will claim that FW (c3) seems to eliminate agency altogether or to simply change the subject. Furthermore, it seems to be vulnerable to the disappearing agent objection.[120] This view seems to secure alternative possibilities at a very high price. It purchases lack of causal necessitation at the price of denying the productive relation between causes and effects. For this reason, it is hard to understand what the *exercise* of powers might be, if everything boils down to the Humean mosaic of events. Agency seems to be eliminated or changed beyond all recognition. This view will be briefly discussed in the next section [B.2.2]. If it makes sense, then defenders of free will have cause to celebrate. To the extent that this view doesn't make sense to them, they have reasons to embrace one of the other definitions of free will and try to work along the lines of another framework.

For the remainder of section A, I will argue that FW (c1) is the most adequate characterization of free will. I will do so by arguing that FW (c2) and FW (c4) are

[119] Beebee and Mele (2002) explicitly accept that Humean compatibilism faces the objection from luck in the same way that libertarianism about free will faces it, and that this might be a reason why traditional compatibilists will not find it appealing. It has to be added though, that they are neutral with respect to whether each position has the ability to overcome this objection.

[120] According to the disappearing agent objection, rather than a being who acts, the agent becomes the locus in which events take place. She gets ultimately lost in the unfolding of events. In his discussion of the disappearing agent objection to event-causal and non-causal forms of libertarianism, Pereboom (2014b) claims that: "it may be that the libertarian needs to appeal to agent-causation to answer the disappearing agent objection." (2014b: 68) This represents one of the reasons that agent causalists think that besides causal relations between events, another kind of causal relation has to be posited, in order to explain how an agent can control an action (O'Connor 2000, 2011, Clarke 2005). An argument according to which, agent-causal conceptions of free will are incompatible with the reducibility of powers is given by Erasmus Mayr (2011: 170).

untenable. The most famous attempt to develop a conception of such respective powers rests on the project of providing a conditional analysis of the ability to do otherwise. Such projects take two forms: i) the classical, and ii) the reformed conditional analysis. In section A.2.3, I will give an argument against both and conclude that the best characterization of free will is FW (c1): the fundamental/irreducible specific power to actively do, otherwise. Neo-Humean compatibilism, which involves an appeal to FW (c3), will be further discussed in section B [B.2.2].

A.2 *The conditional analysis of ability*

Compatibilists who give a conditional analysis of the ability to do otherwise, may conceive of such abilities as ontologically fundamental or reducible.[121] In the first case, they are arguing for FW (c2) whereas in the second case, they are arguing for FW (c4). In both cases, they think that the key to accounting for the ability to do otherwise, consists in grounding it on the truth of a conditional statement about the agent, of a sort to be specified by their respective accounts. Such an analysis is followed by many traditional compatibilists [122] such as Hume (2007[1748]), Moore (1969[1912]), Hobart[123] (1934) and Ayer (1954). According to these thinkers:

[121] The notion of an analysis of power which treats it as an ontologically fundamental characteristic of the universe, seems paradoxical. Nevertheless, not all conditionalist compatibilists seems to have embraced Hume's skepticism of the notion of power. Before Hume, Hobbes 1982[1651] seems to be a realist about powers and a conditionalist compatibilist. This raises an interesting question as to whether the notion of an ontologically fundamental power is consistent with determinism.

[122] Berofsky (2012: 73) notes that criticisms to the traditional conditional analysis of abilities, led many compatibilists to reject it and find shelter in Frankfurtian compatibilism which does not require the ability to do otherwise as a condition for moral responsibility.

[123] Hobart's formulation makes explicit reference to the notion of 'power'. He writes: "A person has a power if it is a fact that when he sets himself in the appropriate manner to produce a certain event that event will actually follow." (1934: 8) His account is a two-stage conditional analysis. As he writes: "There is a series, wish—will— act. The act follows according to the will (that is a law,—I do not mean an underived law) and the will follows according to the wish (that is another law)." (ibid: 8-9).

(CAA)[124]: An agent S has the ability to X, if and only if S would X, had S chosen/intended/decided/tried to X.

Although CAA is an analysis of the notion of 'ability', conditional analysts treat it also as providing us with sufficient materials to give an analysis of the notion of the 'ability to do otherwise'. Thus, they give the following analysis:

(CAADO): An agent S who performed action A, could have done B instead, if and only if S would have done B, had S chosen to perform B.

By giving this analysis, the classical conditional analyst claims that he has shown all there is to having free will. For any action A that we performed in the past, we were free to perform another action, if and only if we *would* perform action B, had we chosen or intended to perform action B. If an agent who performed A would still have done action A, despite an intention to B prior to A-ing, then she would lack free will with respect to that action; if an action would have occurred, despite the agent's intention, then the agent lacked free will with respect to it. If what we do is not a result of our intentions, it can hardly be called an action. For an agent who performed action A to have free will would be for that agent to have done B as a result of an intention to B, if it were true that she intended to B as opposed to A.

The conditional analyst of the ability to do otherwise holds that we can ascribe to an agent the ability to do otherwise, if a conditional statement is true of the agent. Furthermore, they hold that determinism is not incompatible with free will because it

[124] Frequently, the CAA is supposed to account for free will in terms of counterfactual power; the power to do otherwise if different conditions were actual. Although the difference is frequently obfuscated, there is a substantial difference between counterfactual power and two-way power, which is the *ceteris paribus* power to do otherwise; i.e. the power to do otherwise, given the same conditions.

is not necessarily incompatible with the truth of such conditional statements about the agent's abilities. In a deterministic universe it may be the case that some conditional statements of the relevant sort are false for specific agents in specific circumstances at a given time. But this does not mean that all of them must be false. Some of them can be true and therefore, determinism is compatible with free will because it does not necessarily present a threat to an agent's ability to do otherwise.

A.2.1 ***Traditional criticisms***

Criticism 1: Having the ability to do X does not require the consequent of the conditional statement. John Austin (1979) claims that possessing an ability does not require its manifestation in favourable conditions. After all, effort does not guarantee success. An agent can have the ability to do X, even if the conditional statement is false of the agent. Austin writes:

> It is not that I should have holed it if I had tried: I did try, and missed. It is not that I should have holed it if conditions had been different: that might of course be so, but I am talking about conditions as they precisely were, and asserting that I could have holed it. There is the rub (1979: 218 n.1)

Criticism 2: The truth of the conditional statement is not sufficient for attributions of ability. One can fail to possess an ability despite the fact that the conditional statement is true of her. An agent might not be able to do something, even if she would do it, if she tried to. The problem is that it might not be possible for her to try it. Keith Lehrer (1968) writes:

> Suppose that I am offered a bowl of candy and in the bowl are small round red sugar balls. I do not choose to take one of the red sugar balls because I have a pathological aversion to such candy. (Perhaps, they

> remind me of drops of blood and…) It is logically consistent to suppose that if I had chosen to take the red sugar ball, I would have taken one, but not so choosing, I am utterly unable to touch one. (1968: 32)

The truth of the conditional statement is not sufficient for ascriptions of ability because the agent might satisfy the conditional as a whole, while being unable to satisfy its antecedent. A conditional statement of the form 'if p then q' can be true for an agent, even if the agent is unable to fulfil the antecedent of the conditional; i.e. if p is not possible for that agent.

Criticism 3: The conditional analysis of abilities analyses the exercise of an ability to do something in terms of what the agent *would* proceed to do if it were the case that the agent chose/intended/decided/tried to do something (as in (1) above). The problem of the infinite regress comes forward when the ability to choose/decide/intend/try is held as analysable *in the same way* as the ability to perform an action (Schlosser 2017). Take for example, sentence (1):

(1) An agent is able to do X, if and only if the agent *would* X, if the agent chose/decided/intended/tried to do X.

If we are to analyze the ability to choose in the same manner, we get:

(CAAC) An agent is able to X (choose something), if and only if the agent would do X, if the agent chose to X.

CAAC (conditional analysis of the ability to choose) states: an agent is able to choose something –call it 'choosing to X'- if and only if the agent would choose X, if the agent chose to choose X. And so on, ad infinitum.

A.2.2 ***Reformed conditional analyses of abilities***

The criticisms described above lead many classical compatibilists to abandon the conditional analysis of ability and to embrace the views of P.F. Strawson (2003[1962]) and Frankfurt (1969) who downplay the importance of the ability to do otherwise, in their accounts of free will and moral responsibility. Following David Lewis's (1997) reformed conditional analysis of dispositions, new dispositionalists, like Michael Smith (2003), Kadri Vihvelin (2004, 2013) and Michael Fara (2008) give a reformed conditional analysis of ability in terms of a reformed conditional analysis of dispositions. In what follows, I will focus mainly on Vihvelin's account of free will, according to which, possessing free will consists in having a bundle of dispositions (Vihvelin 2004, 2013).

The project of reforming the conditional analysis of ability rests on providing a satisfactory conditional analysis of dispositions. Yet, like the simple conditional analysis of ability, the simple conditional analysis of dispositions is considered to require reformation as well. Lewis (1997) argues that the simple analysis of dispositions has been decisively refuted.[125] Due to cases of finks –circumstances that remove a disposition and make its manifestation impossible- ordinary dispositions can fail to manifest even if the stimulus is present.[126] Lewis describes the simple conditional analysis of dispositions (SCAD):

[125] I will claim that to the extent that this is meant as a compatibilist project, this statement is false. It conflates two standpoints we enter into when we engage in ascriptions of powers and abilities. To the extent that he is a compatibilist, Lewis talks either of general abilities, or does not understand that he is assuming the wrong standpoint to talk about the powers of objects in deterministic universes. This point will be argued in this chapter's A.2.3.

[126] The same has been claimed for masks. Michael Fara (2008) claims that the conditional analysis of abilities is wrong because "it rules out the possibility of masked abilities." (2008: 850)

> "Something x is disposed at time t to give response r to stimulus s, iff, if x were to undergo stimulus s at time t, x would give response r." (1997: 143)

He argues that such an analysis cannot account for cases in which finks are present. He claims that: "Such a disposition, which would straight away vanish if put to the test, is called *finkish*." (ibid: 144) That means that there are cases where, due to finks, an object x might undergo stimulus s, without response r manifesting.

New dispositionalists aim to give an account of abilities in terms of dispositions. Reforming the simple conditional analysis of dispositions by introducing clauses of fink-safe and mask-safe circumstances, is supposed to lead to a reformed conditional analysis of abilities, which perhaps will not be vulnerable to the traditional criticisms mentioned above. Fara writes: "An agent has the ability to *A* in circumstances *C* if and only if she has the disposition to *A* when, in circumstances *C*, she tries to *A*." (2008: 848) As Fara's analysis suggests, the attribution of the disposition to A is crucial for the attribution of the ability to A. Since to have abilities is understood in terms of having dispositions, conditions which would serve as masks or finks of such dispositions, would have the same effect upon abilities. Fara writes that "necessarily, an agent's ability to do something (in certain circumstances) is masked if and only if her disposition to do it when she tries to do it (in those circumstances) is in turn masked." (ibid: 848)

Fara holds that the reformed conditional analysis of abilities in terms of dispositions can overcome the problems of the simple conditional analysis of abilities. He writes:

> Since dispositions can be masked, the dispositional analysis correctly predicts that abilities can be masked as well. (…) The failure of the conditional analysis to account for cases of masked abilities is exactly analogous to the failure of conditional analyses of dispositions exactly

> analogous to the failure of conditional analyses of dispositions to account for cases of masked dispositions. (ibid: 850)

The most sophisticated attempt to reform the conditional analysis of ability is arguably that of Kadri Vihvelin (2004, 2013) who holds that to have the ability to do otherwise, is to possess a bundle of dispositions. Vihvelin thinks that the key to understanding ability, both in its 'wide' and 'narrow'[127] sense, is a sophisticated analysis of dispositions with clauses that specify the absence of finks in a series of cases in which the stimulus-conditions for the manifestation of a disposition are present. Vihvelin follows Lewis in holding that finks remove dispositions due to their interference with the bearer's intrinsic properties, which serve as a causal substratum for the manifestation of the disposition. Consequently, she provides an analysis which is sensitive to the possibility of such interference. According to Vihvelin (2013):

> S has the narrow ability at time t to do R in response to the stimulus of S's *trying to do R* iff, for some intrinsic property B that S has at t, and for some time t' after t, if S were in a test-case at t *and* S tried to do R *and* S retained B until time t', then in a suitable proportion of these cases, S's trying to do R and S's having of B would be an S-complete cause of S's doing R. (2013: 187)

The intrinsic property B that Vihvelin holds as central to her reformed conditional analysis of dispositions represents "the causal basis of the ability in question" (Clarke,

[127] Vihvelin uses the distinction between 'narrow' and 'wide' abilities to capture the distinction between abilities which are attributed to an agent on the basis of a capacity alone and those whose ascription to an agent require not just the physical capacity to use one's such skills at a particular time, but also the agent's access to opportunities, or circumstances which would not make the exercise of such ability impossible. As Clarke (2015) writes: "These are abilities to do things on certain occasions." (2015: 893) I consider the distinction between 'narrow' and 'wide' abilities to be co-extensive with that between 'general' and 'specific' abilities.

2015: 896). The reformed analysis marks a progress from the simple one due to the clause which specifies that it is successful at the absence of conditions that would prevent the relevant disposition from manifesting, if put to the test. By introducing the time interval between t and t', Vihvelin ensures that the property is in fact relevant to the manifestation of the disposition and did not just happen to be there, as a matter of pure coincidence. Furthermore, she ensures that the property has not vanished. Her emphasis on the idea of a test-case and her appeal to the time interval between t and t', suggest that she accepts Austin's point that an ability can be possessed without having to always manifest as a result of specific stimulus-conditions.

Vihvelin holds that the reformed analysis of dispositions will help in overcoming the traditional problems of the simple conditional analysis of abilities to act. Whereas the simple conditional analysis of ability cannot account for masked abilities (Fara 2008), the reformed analysis can, by the introduction of clauses which specify that success is not required for the attribution of abilities, since the agent is not required to manifest the disposition *in all cases* where they are put to the test.[128]

Vihvelin (2004, 2013) holds that the infinite regress objection is misplaced. She holds that the root of this misconception lies in the mistaken assumption that the ability to choose otherwise can ground the ability to do otherwise. She writes:

> Yes, when we say at someone, "if she had chosen to do X, she would have done X," we usually assume that she could have chosen to do X. But that is not because we think that having the ability to do X *requires* having the ability to choose to do X, but, rather, because we think that people typically have the *ability to choose* whether or not to do what they do, in *addition to having the ability to do* what they do. That is, we assume that persons have the bundle of

[128] For the view that the reformed analysis is successful in the case of finks, but not in the case of masks, see Clarke (2015).

> abilities which constitute the ability to make choices on the basis of reasons as well as abilities to do various things with their bodies (at 'will', for no reason other than impulse or 'mere' desire). There is no regress because someone (an animal, a young child) may have abilities of the second kind without having any or many abilities of the first kind. (2004: 442-443)

According to Vihvelin's response, we don't need to have the ability to choose something, in order to have the ability to do it. It seems that such a response misses something important. For the question here is not whether we have an ability to do something, as opposed to having the ability to choose it. The question is whether choosing is something we actively do; whether it belongs to the domain of agency or not. Since in the context of conditional analyses, actions are counterfactually dependent on choices, banishing choices from the domain of agency renders the exercise of agency counterfactually dependent on something over which the agent has no control. If choices are conceived of as distinct from doings, then we think of them as something that merely happens to us. As Steward writes:

> Decisions (unlike, for example, shootings) seem *essentially* to be actions: one cannot decide to decide something without being the decider, the agent of the decision. (…) The occurrence of this decision in him, because brought about by a process over which he has no control, seems to be just that: an occurrence *in him* not a decision by him." (2012: 179)

If we follow Vihvelin in claiming that choice is not an action, the ability to choose otherwise would not refer to our agency, but to the ability for something to happen to us otherwise. Hence, our agency would be passive in that, even in the paradigmatic compatibilist cases of free action, such actions would entirely depend upon circumstances over which we have no control and which lead to the formation of a decision or intention in us. If circumstances C (F n1) obtained, over which we are

powerless, action A1 would follow. If circumstances C (F n2) obtained, over which we are powerless as well, action A2 would also follow. Since in that case our actions would depend upon what is not under our control, our actions themselves would not be under our control.

The infinite regress objection is not countered for another reason as well. It shows that Vihvelin's account is not obviously consistent with the traditional compatibilist claim that free actions are those which are performed without impediments, as a result of the agent's choices and intentions. Assuming a distinction between the ability to choose to eat as opposed to postpone eating for the time being, and an ability to eat, it is beyond reasonable doubt that the latter does not require the former. We don't have to ascribe the ability to choose to eat to infants. Yet the ability to do otherwise cannot be independent of choice, if it is meant as empowering the agent's control over her actions. The challenge lies in that compatibilists (e.g. Hobbes 1982[1651], Hobart 1934) have claimed that such a notion of ability to do otherwise is inherently unsatisfactory due to the lack of the strong connection that has to exist between an action and some other mental state, which has to be in place for the former to constitute an action. For along compatibilist lines, if the ability to do otherwise is conceived of as independent of an agent's choice or intention, it ends up diminishing the agent's control, as opposed to enhancing it. In Hobart's words: "it is born at the moment, of nothing, hence it expresses no quality; it bursts into being from no source."[129] (1934: 7) If the ability to do otherwise is linked with the ability to choose to do otherwise, the infinite regress objection is still left unanswered.[130] If the ability to do otherwise is independent of the ability to choose to do otherwise, then it ends up being a concept of freedom that many compatibilists have explicitly rejected as incomprehensible. For according to compatibilists, the paradigm of free action is action

[129] Hobart extends this to undetermined acts, even if they are linked with choice.

[130] Perhaps the compatibilist could claim that it is not a vicious regress and that the ability to choose to choose x, is the same as the ability to choose x. It is not exactly clear to me whether this strategy could work.

as a result of a choice or an intention. To claim that one has an extra ability to do something, irrespective of the choice or intention that brings it forth, is not to account for a kind of action that most compatibilists would call free.

In order to avoid the risk of misrepresenting Vihvelin's point, a further word is in order. Assuming I have the ability to do otherwise than I choose at a given moment, the question arises as to whether doing otherwise is possible for me *as a result of* choosing to do otherwise or irrespectively of any choice I make, if I make one. If I choose to do otherwise than I chose to do before, and do it, I merely changed my mind with respect to what I choose to do.[131] But if I choose to do one thing and I end up doing otherwise without choosing to do otherwise, then that is either the result of weakness of will or of pure random behaviour.[132] But my changing my mind and acting otherwise, is not independent of my ability to choose to act otherwise. Thus, this line of thinking does not seem to help Vihvelin's point. If I choose to do something and do something else, I am either not doing what I am doing as a result of my choice or my action is random and again, is not an outcome of my choice. It seems that both cases are against what the compatibilist would characterize as a paradigmatic case of action. In a nutshell, it seems that the way that Vihvelin counters the objection from infinite regress does not secure what a compatibilist would call the right kind of connection between one's choices or intentions and actions. To say that we have the ability to do something irrespectively of an ability to choose to do it, seems to either loosen the ties between choice and action or make action counterfactually depend on something over which the agent has no control.

The lack of a satisfactory answer to the infinite regress objection raises the worry that the compatibilist cannot maintain that once Vihvelin's account gets clear of

[131] Nicholas Rescher (2015) makes a helpful distinction between 'preliminary' decisions and 'actual' decisions. The former refer to choices that refer to the future, which can be revoked if the agent changes her or his mind. The latter is irrevocable because it refers to a choice which has already been acted upon.

[132] It also seems to raise problems for those compatibilists who embrace Davidson's (1963) claim about the difference between reasons for φ-ing and reasons why the agent φ-ed.

defects, the response will be met. The problem is not the specific details of Vihvelin's account, but rather the whole endeavour. To summarize, an account of the ability to do otherwise will either: a) link the ability to do otherwise with choice, or b) give an independent account of the ability to do otherwise. In the former case, the infinite regress objection seems to be left unanswered, for we still need an account of the ability to choose and its relation to action. In the latter case, although Vihvelin gives an account of the ability to act independently of the ability to choose to so act, the leap cannot be made from an account of the ability to act, to an account of the ability to act otherwise, without shifting the concept of free will to one compatibilists have vociferously detracted in the past; that of pure random behaviour. Furthermore, the subject changes because at the absence of a link between choice and action, the compatibilist defends freedom of action, as opposed to freedom of will. And if actions counterfactually depend on conditions which do not belong to the domain of agency and are not under the agent's control, the performance of action cannot be under the agent's control. It seems purely accidental if the agent does as the agent intends to do.

But let us assume, for the sake of the argument that the infinite regress objection can be answered by conditional analysts. Certain considerations might be brought forth which might secure both the old and the reformed project of reducing statements about ability to conditional statements. It seems that an independent objection to the conditional analysis of ability can be given, to which we now turn.

A.2.3 ***The situatedness argument***

The conditional analysis of the ability to do otherwise, both in its old and reformed version, seems to end up losing sight of the agent's situatedness [Ch. III/B.2]. Such analyses claim that an agent is able to do otherwise in a deterministic universe, provided a qualified conditional statement is true of that agent. According to the conditionalist compatibilist, an inhabitant of a deterministic universe could have done

otherwise than she did, if it is true that if she chose or intended to do otherwise, then she *would* have done otherwise. Reflection on our situatedness allows us to see that conditional analysts of the ability to do otherwise, end up viewing the agent and her abilities in abstraction from the circumstances in which she is situated at any given time. The reason for this is that the manner in which they account for free will implies that for an agent who in C (F n) at t, freely performed action A at t+1, the necessary condition for doing otherwise, say B, at t+1, is that she were situated in C (F ~n) at t. This implies that the necessary condition for a situated agent –an agent in C (F n) at t- to freely do otherwise than she actually did, is that she were situated in circumstances that are different from the ones she was actually situated in, at the time before action. It is because no situated agent can be in such a position that the necessary condition for doing otherwise is necessarily non-actual for a situated agent. Therefore, if the necessary condition for a situated agent to do otherwise is necessarily non-actual for that agent, no situated agent can ever do otherwise, at any time. Since agents who inhabit deterministic universes, are also situated, along the lines of the conditional analysis of the ability to do otherwise, no agent in a deterministic universe can ever do otherwise. Thus, conditionalist compatibilism fails. In a nutshell, conditionalist compatibilists can only give a positive answer to the question whether an agent who performed action A in a deterministic universe could have done otherwise than A, at the expense of not treating the agent as a situated being. Therefore, in giving a conditionalist compatibilist account of an agent's ability to do otherwise, they cannot account for a *situated agent's* ability to do otherwise.

Agents are essentially situated beings. To speak of an agent as performing actions as well as possessing the ability to act in a particular way is tantamount to speaking of an agent *in certain circumstances which assume one form rather than another, at a given time t.*[133] If the ability to do otherwise is the ability to do otherwise

[133] Frequently, compatibilists protest that the incompatibilist holds such circumstances for an agent to be necessary. Doubtlessly, metaphysical necessitarians claim exactly that. But an incompatibilist does not have

in specific circumstances at a specific moment, then it becomes evident that along the lines of the conditional analysis of the ability to do otherwise, the necessary condition for the agent in given circumstances which assume a specific form at a time t, to do otherwise is the non-existence of those very circumstances in which the agent is actually situated at that time and to which, the question whether the agent could have done otherwise, must refer. Since an agent acts always in certain circumstances C which have form (F n) at a time t, the necessary condition for doing otherwise than what one does in a deterministic universe, is that the circumstances C in which the agent is situated at t, have a form other than (F n); for an agent in C (F n) at t, the necessary condition for doing otherwise at t+1, is that she were situated in C (F ~n) at t. Since the necessary condition for doing otherwise is necessarily non-actual for a situated agent in a deterministic universe, doing otherwise is not possible for the agent in circumstances C having form (F n) at t. Since no one can be in such a position, along the lines of the conditional analysis of ability there can be no ability to do otherwise in certain circumstances C at a given moment, whatever form they assume. It is not enough to implicitly or explicitly hold that situatedness matters without paying thorough attention to the form the agent's situatedness assumes. And this holds both for choosing, forming intentions, as well as acting.

The challenge can be demonstrated in another way. If we are essentially situated agents, the question whether an agent had a specific ability to do otherwise is elliptical for the question whether an agent in given circumstances, having one form rather than another, at a given moment, had the ability to do otherwise in the next moment. Our initial question,

---

to be a necessitarian. The upshot of the situatedness objection is that determinism places necessity in the relation between the totality of circumstantial factors prior to the formation of the intention and the formation of that intention, as well as in the relation between the totality of states of affairs at the moment prior to action and the performance of the action, irrespectively of the independent modal status of the relata.

Q: Was the agent able to do otherwise?

assumes the following form:

Q*: Was the agent in C (F n) at t able to do otherwise at t+1?

The situatedness argument against conditionalist compatibilism amounts to the following charge: the compatibilist has to change the subject in giving a positive answer. For a positive answer to Q* based upon the conditionalization of the ability to do otherwise, will eventually take the following form: yes, because if the agent were in C (F ~n) at t, then the agent would do otherwise. And since the necessary and sufficient condition for doing otherwise is absent for an agent in C (F n) at t, the agent in C (F n) at t, cannot do otherwise at t+1.To illustrate this better, let us take an example.

Assume deterministic world W. An inhabitant of it, Matthew, engages upon deliberation and performs action A at t+1. Could Matthew at t, have performed action B at t+1 instead?[134] Now, if we ask question Q, it is easy to miss what can be seen more clearly if we ask question Q*. Given the fact that Matthew was a situated agent at time t before performing action A, he was situated in circumstances C which had form (F n). Once we see that Q is elliptical for Q*, the question whether Matthew at t, could have done B at t+1, is elliptical for the question whether Matthew in C (F n) at t, could have performed action B at t+1. If the answer is positive, then both A and B are possible for an agent in C (F n) at t, which is in conflict with the definition of determinism.[135] If

[134] A response that would not respect situatedness is the following: yes, because it is not metaphysically necessary that Matthew formed intention A at t. The situatedness argument does not claim that it has to. First, it is not a matter of metaphysical necessity of Matthew's intending to A at t; it is a matter of causal/physical necessity, given Matthew's situatedness at t-1. Necessity, in the causal/physical sense, is placed in the causal relation between causes and effects. Humean views, like those of Beebee and Mele (2002) and Bernard Berofsky (2012) will be treated in this chapter's section B.2.2.

[135] The conception of determinism I have in mind here is a non-Humean one. Humean versions of determinism will be discussed towards the end of this chapter. It will be argued that neo-Humean compatibilist accounts

the answer is negative, then Matthew could not, in deterministic world W in C (F n) at t have performed action B at t+1. The conditional analysis may only give a positive answer by changing Matthew's situating circumstances at t. They may answer: yes, Matthew could have performed B if Matthew were not situated in C (F n) at t. But this neglects that to ask Q is to actually ask for Q*, at pains of treating Matthew's situatedness as an accidental feature of him. The necessary and sufficient condition for doing otherwise (B in this case) is necessarily absent for Matthew in C (F n) at t. So, the conditionalist compatibilist changes the subject, by answering to a different question or by misunderstanding what is involved in it.

Neglect of situatedness leads us to neglect that Q is elliptical for Q*. It says: yes, the agent could have done otherwise because if the agent were situated differently, the agent would have done otherwise. This amounts to changing the subject. For, to ask whether an agent could have done otherwise is to ask whether a *situated* agent could have done otherwise. Answering by treating situatedness as inessential is to change the subject by giving an answer to a different question or by deliberately treating situatedness as something inessential. Let us pay attention to possible objections.

One objection concerns the role of intentions in an agent's actions. The compatibilist may claim that situatedness, in this sense, is not that relevant because the way that an intention arises is not what is relevant to free action. What is relevant is whether the action itself has been produced by an intention, however the latter is formed. An action is free if it is caused by an agent's intentions and it is true that if the agent intended to do otherwise, then the agent would do otherwise.

It is true indeed that in a deterministic universe the following scenario is possible: i) Matthew intends at t to perform A at t+1, ii) Matthew performs A at t+1,

---

secure lack of causal necessitation by undercutting the agent's power to effect a change on the environment by the exercise of free will. Talk of it being 'up to humans' to partly determine laws of nature, ultimately involves an epistemic nature of determination which applies in the same way to everything in the universe, rather than a causal one.

and iii) it is true that if Matthew intended at t to perform B at t+1, then Matthew would have performed B at t+1. It does not follow from i), ii) and iii) that Matthew at t could have done otherwise than A at t+1. The reason is twofold. First, it ends up being an account of a general ability. Second, it does not fulfil its basic desideratum; it is not an account of the ability to *do* otherwise; for if the ability to do otherwise is elliptical for the ability to do otherwise at t+1, as a result of an intention at t to do otherwise at t+1, then the compatibilist faces the following two questions: When is the agent able to do otherwise at t+1? At t or at t-1? If the agent's situatedness is to be respected, then the answer whether the agent can do otherwise has to refer to an agent in C (F …) at a given moment in time. In other words, due to his situatedness, the ability to perform B at t+1, has to be ascribed to Matthew at the moment prior to action or at the moment prior to the formation of the intention to so act.

Let us consider whether Matthew has such an ability in C (F n) at time t. By the assumption of determinism, at time t, Matthew has only one physically possible course of action at t+1; action A. This means that action B is physically impossible for Matthew at t; the necessary condition for doing B is necessarily absent for Matthew in C (F n) at t, since it requires what cannot be true of anyone in C (F n) at t; namely, that they are situated in C (F ~n) at t. Since the necessary condition for doing B at t+1, is necessarily absent for Matthew in C (F n) at t, Matthew at t cannot do otherwise than A at t+1. Thus, the agent at t does not have the ability to do B at t+1.

Let us consider whether Matthew has such an ability to do B –without losing the ability to do A- at time t-1. Could the agent in C (F n) at t-1 have formed a different intention at t? The answer is negative. Why? Because, given C (F n) at t-1 and the assumption of determinism, the way that universe turned out to be at t, which includes states of affairs that in this case involve Matthew's intentions, represents the totality of what was physically possible (and by implication, not physically impossible) to occur in that world, given the totality of states of affairs, i.e. the specific form that the totality of what exists in that universe assumed at t-1. For Matthew, who in deterministic world W intended at t to A at t+1, to have intended at t to B at t+1, he would have to be in C

(F ~n) at t-1. The necessary condition for Matthew at t to intend to do otherwise at t+1, is that prior to the formation of an intention to so act, he were situated in circumstances which assumed a different form than the one they actually assumed at the moment prior to the formation of his intention to act, *in that deterministic universe*. But since agents are situated beings, to claim that the agent has the ability to form another intention at t, has to make reference to when and where the agent has such an ability, at pains of talking about general abilities. Thus, what applies for ascriptions of the ability to do otherwise, applies as well to ascriptions of the ability to intend to do otherwise. The reason for that is that the bearer of both abilities is essentially a situated being; situated always in circumstances which have a specific form rather than another, at particular times. To answer positively to the question: 'could Matthew, who inhabits a deterministic universe, have intended to do otherwise?' on the basis of a conditional analysis of ability is to change the subject. If situatedness is to be respected, this question is elliptical for: 'could Matthew in C (F n) at t-1, who inhabits a deterministic universe, have intended at t to do otherwise at t+1?' A positive answer on the grounds of a conditional statement would be: yes, because if Matthew were situated in C (F ~n) at t-1, then he would form an intention at t, to B as opposed to A at t+1. And here is where the subject is changed. This answer neglects that when we ask whether Matthew could have intended to do otherwise, we are talking about a situated agent, not someone whose situatedness is an accidental feature.

Another objection that may be raised against the situatedness argument is that it leads to necessitarianism. If it is necessary that the agent is situated in circumstances having a particular form rather than another at a given time, then there is no other way the universe could have been instead. This objection rests on a fallacy of equivocation. It is one thing to claim that so long as we speak of an agent, it is necessary that this agent is situated in circumstances which have one form rather than another at each time. It is quite another thing to claim that the specific form that these circumstances assume at each time is itself necessary. It is the first claim that I make and the latter

that might lead to the belief that the situatedness argument implies the truth of metaphysical necessitarianism. To see that they are different, it is easy to see that the first claim is consistent with indeterminism. Each inhabitant of an indeterministic universe is situated in circumstances which have one form rather than another – indeterminism does not imply the existence of contradictions in states of affairs- at a given time t. It is also true that the same inhabitant will be situated in that same universe in circumstances that will also have a specific form, rather than another form at t+1. The whole point of indeterminism is that which specific form they will assume at t+1 is not causally or metaphysically necessitated by the form they assume at t. Since necessitarianism is inconsistent with causal indeterminism, unlike the position I make, the position I make is not implied by necessitarianism; it is consistent with it, although it does not rest on it. This means that it is true, independently of whether the universe we inquire about is a necessitarian one or not. Therefore, claiming that so long as we speak of an agent, it is necessary that the agent of whom we speak will be situated in circumstances having a specific form rather than another at a given time, is not to embrace metaphysical necessitarianism.

The final objection I will mention concerns what can be referred to as neo-Humean versions of determinism. Such an objection claims that it is a modal fallacy to claim that it is necessary for Matthew to do A at t+1. Matthew's world as well as its laws, could have been otherwise. Furthermore, each individual property or collection of properties could have been otherwise. Thus, it is not the case that Matthew could not have done B at t+1. We can reply to this objection by focusing on the notion of necessity that it employs. If necessity is conceived of as a metaphysical category, where X is necessary if X is true in all possible worlds, then it was not necessary for Matthew to do A at t+1. But that is not the notion of necessity that is involved in the situatedness argument. Necessity is meant in a causal/physical sense of a non-Humean sort and is located in the particular causal relation between the sufficient causal conditions of Matthew's A-ing at t+1 and Matthew's action at t+1, independently of the modal status of either of them. Furthermore, such a relation of necessity between cause and effect is

again not a relation of metaphysical necessity; it is not implied by it that every pair of deterministic universes which are identical at a time have to be identical at the next moment.

The upshot of the situatedness argument is that the project of giving a conditional analysis of the ability to do otherwise fails because it fails to account for *a situated agent's* ability to do otherwise. Since agents are situated, agents *in a deterministic universe* are situated as well. Therefore, conditionalist compatibilism, the project of employing the conditional analysis of the ability to do otherwise in order to account for compatibilism between free will and determinism, fails as well. However sophisticated an analysis of such an ability is given, no matter how much we inquire about what clauses to include in it, the very idea of basing the ability to do otherwise on the truth of a conditional statement ends up losing sight of human situatedness. Thus, it seems that if the reformed conditional analysis of the ability to do otherwise is the best candidate for both FW (c2) and FW (c4) respectively, and such an analysis fails because it is conditional –not because it is simple or not- then it seems that FW (c2) and FW (c4) are not the best characterization of the concept of free will. They cannot serve the compatibilist's ultimate purpose which revolves around its essence: the position that free will is compatible with determinism.

This leaves us with FW (c1) and FW (c3) as the final rival conceptions. As mentioned before, FW (c3) is essentially linked with neo-Humean conceptions of powers, such as Humean compatibilism. This position will be briefly discussed again in section B.2.2. If this conception is sensible, defenders of free will have cause to celebrate. I will proceed as though it does not make sense due to suspicions that it ultimately eliminates agency or changes the discussion by changing the concept of agency so much, that it loses its ordinary meaning. For this reason, I will settle with FW (c1) as the best characterization of free will.

Settling with a notion of free will, allows us to answer what would make it impossible. The answer is given by the negation of the essential elements of FW (c1),

which are: a) fundamentality of power, b) active conception of 'doing', and c) alternative possibilities. Since these three elements form the essence of free will, free will would be made impossible by anything that involves commitment to at least one of the following positions: a) the denial of the fundamentality of power,[136] b) anti-realism or reductionism about active 'doing', and c) the denial of alternative possibilities possessed by a situated agent.

The situatedness argument rebuts the conditional analysis of the abilities to do otherwise, implicit in Wedgwood's (2013) compatibilist treatment of PAP and OIC.[137] No agent can be under the scope of an obligation or a normative reason if it is not possible for that agent to violate it. Since talking about an agent is essentially talking about a situated being, no agent in particular circumstances, having one form rather than another at a given time t, has both the ability to satisfy the demand generated by normative considerations and violate it, in a deterministic universe. Since this is the condition for being under the scope of normative considerations, it follows that no agent in a deterministic universe can be under the scope of normative obligations and reasons. Certain doubts still remain with respect to the conception of determinism that is involved in the situatedness argument. To counter them, I will devote section B to defining determinism and arguing that the two dominant categories of conceptions of determinism are incompatible with free will.

[136] It would be circular to reject FW (c3) on the basis that it denies the essence of FW (c1). For that would presuppose the latter. No such argument is or will be given. The suspicions that will be mentioned do focus on the implications that the reducibility of power seems to have for agency, but not on the grounds that the reducibility of powers is seen as making free will impossible, in light of the truth of FW (c1).

[137] It does not seem to be the case that Wedgwood would be interested in embracing a position like Humean compatibilism. But even if he wouldn't, others might be. Again, this is a route they might take, however plausible it might be.

Section B: ***Is determinism committed to what would make free will impossible?***

In the first section, we concluded that the best characterization of free will holds it to be the fundamental, specific power to actively do otherwise. Assuming that agents had and have free will in certain cases, with respect to those cases, they could have done otherwise than they did, as well as have alternative possibilities open to them with respect to what they will do now and in the future. In section A, it was maintained that a position is incompatible with free will if it is committed to the denial of at least one of its essential elements.

Having answered the first question, it is time to give an answer to the second one; whether determinism is actually committed to the negation of any essential constituent of free will. As argued for in the previous section, the necessary conditions for free will are: a) the fundamentality of powers, b) the active conception of 'doing' and c) alternative possibilities. A potential incompatibility of either element with a version of determinism renders free will incompatible with it. If there can be a version of determinism which is compatible with conditions a), b) and c), then the answer to this section's question will be negative with respect to that form of determinism. I will argue that there are two general conceptions of determinism which are ultimately incompatible with free will. What I will call robustly causal determinism is inconsistent with alternative possibilities and arguably, the active conception doing. What I will call neo-Humean determinism is inconsistent with the fundamentality of specific powers and ultimately, active doing. The first renders it impossible for us to do otherwise at each particular moment, whereas the latter secures lack of causal necessitation of our actions by denying that anything, including the agent, can effect a change on the environment in the relevant sense; action becomes a cosmic accident.

Like the notion of 'free will', the notion of 'determinism' is a complex one. If we put aside theological and logical determinism, the key notions that are involved in how the notion is traditionally understood concern laws of nature, causation and

possibility. According to Van Inwagen's frequently quoted definition, determinism is "the thesis that there is at any instant exactly one physically possible future." (1983: 3) I will focus on versions of determinism which involve the notion of 'physical possibility' in their formulation. Conceptions of determinism vary but there seems to be a conceptual pendulum swing between conceiving of determination as governance by causes and governance by laws.[138] At the one end, determinism can be conceived of as governance by causes, which are non-nomic.[139] At the other end, determinism can be conceived of as governance by laws, which are non-causal in the sense that implies the productive relation between causes and effects. This is the position explored by Beebee and Mele (2002) and argued for by Berofsky (2012). In between, we find some intermediary positions where the agent is governed by both and in which governance by laws may be reducible or at least as basic as governance by causes.[140] This seems to me to be the position of those who embrace deterministic forms of conceptions of natural laws as second-order relations between universals, like David Armstrong (1983). In the rest of the section, I will describe recurrent issues in defining determinism [B.1], discuss some influential statements of the deterministic sentiment [B.1.1] and give two definitions of determinism; robustly causal determinism and neo-Humean determinism [B.1.2]. The chapter will conclude that both are incompatible with free will [B.2.1, B.2.2].

[138] Helen Beebee (2000) argues that on the Humean sense, laws do not govern. I think that it is perhaps better to claim that they do govern, but not in the sense that non-Humean laws of nature are said to govern.

[139] Advocates of such forms of determinism may deny the nomological character of causation. Although many deny that determinism can be non-nomological, there seems to me to be no contradiction in holding that they are compatible.

[140] To maintain that there can be another position between the two ends of our conceptual pendulum's swing, in which the agent is governed by both and in which facts about governance by causes reduce to facts about governance by laws, is redundant because it amounts to restating one end of the pendulum's swing; neo-Humean determinism.

B.1 ***Issues in defining determinism***

An adequate formulation of determinism has to avoid the problem of triviality. There is a very trivial manner in which we can use the notion of determinism, and its very triviality consists in its lack of specific content. The phrase: *Que sera, sera*, is often used to characterize the deterministic sentiment, which is falsely interpreted as necessarily having a fatalistic character (Earman 1986). Trivial talk involves phrases such as:

(1) What will be, will be.

(2) The future will be brought about by the past.

(3) Necessarily, there is only one form that the future may assume at time t.

Such phrases are totally inadequate to characterize determinism as a doctrine. Sentence (1) can be interpreted in various ways. It can be seen as a tautology. What it doesn't show us is the manner in which what will be, will come about. It is equally true in a deterministic as well as an indeterministic universe. Even worse, there are no reasons to deny that it will be true in a universe where everything is uncaused. Sentence (2) is also vague. It seems like the terms 'future' and 'past' are used as general terms that are meant to cover specific states of affairs that hold at a given moment, conceived of as causes and effects. If the future being 'brought about' by the past, means that certain states of affairs will be causally produced by antecedent states of affairs, we are told nothing positive about the *manner* in which the effect will be brought forth by its cause. All we are told is that the future is not uncaused and it is prudent to remember that 'not being uncaused' stands for something different for theorists who embrace different conceptions of causation, let alone different conception of laws of nature. Factors that

obtain at one moment could either causally necessitate the 'future' or probabilify it; i.e. make it more likely; and only in that sense, such sentences are not empty of content.

The ambiguity of sentence (3) is rather interesting. It can be read as a robust necessitarian statement or as a purely logical one; the future, conceived of as a totality of states of affairs that obtain at a given moment, can only take one form at a time because if it took more than one form at a time, a contradiction would be true. If we hold to the law of non-contradiction, we are committed to the belief that contradictions cannot be true. The problem with maintaining that statement (3) is an adequate conception of determinism is that it can be true even if causal indeterminism is true as well. In an indeterministic universe, there would still be no contradictions; i.e. it cannot be the case that a state of an indeterministic universe can take a contradictory form. But the fact that it is settled that the future will assume *one* form at each instant – because the future cannot have a contradictory form- doesn't by itself imply that out of all the possible forms it can take, it is also settled which *specific form* it will assume at each instant. The notion of 'possibility' which is claimed by determinists to characterize certain states of affairs and not recognized with respect to others, is not the 'logical' one. As John Earman (1986) writes: "Laplacian determinism entails one kind of non-trivial inevitability: given the way things are now, the future can't be other than it will be, where the 'can't' is the 'can't' of physical impossibility." (1986: 18) Unless the notion of possibility we are referring to isn't specified, our use of determinism is vague and is open to interpretations which would be compatible with indeterminism. A conception of determinism as compatible with causal indeterminism is hardly a satisfying account for the notion of determinism, which is traditionally interpreted as posing a threat to free will. Besides appealing to the notion that effects are not uncaused, an informative definition of determinism has to refer to the *manner* in which effects are brought about by their causes. Once again, it is the nature of the *relation* between cause and effect which is important in characterizing a causal relation or a universe as deterministic.

B.1.1 *Traditional formulations*

Determinism is associated with a certain worldview which essentially concerns the manner in which living and non-living entities are situated in the causal nexus. Expressions of it abound in both philosophical and non-philosophical literature. It has been interpreted as posing a threat to free will since it is meant to capture a particular way in which human beings are situated in the world, which deprives them of alternative possibilities of action at any given moment. According to such a worldview, determinism poses a threat to free will because the way in which the performance of human actions is accommodated within its confines leads to a conception of human beings as essentially passive playthings of the forces of their surrounding environment; a picture in which the agency of human beings, despite its vast array of phenomenal accompaniments, is not essentially different to that of entities which we hold as paradigmatically unfree. In a pithy description of the view, William James writes:

> What does determinism profess? It professes that those parts of the universe already laid down absolutely appoint and decree what the other parts shall be. The future has no ambiguous possibilities hidden in its womb: the part we call the present is compatible with only one totality. Any other future complement than the one fixed from eternity is impossible. The whole is in each and every part, and welds it with the rest into an absolute unity, an iron block,[141] in which there can be no equivocation or shadow of turning. (James, 1956[1897]: 150)

The most famous exposition of determinism can be found in Pierre-Simon Laplace (2015[1820]). I will show how robust and Humean understandings of laws of

[141] It might be prudent to add that this 'iron block' doesn't itself have to be metaphysically necessary. Spinozists might view it that way, but being a necessitarian is not the only way of being a determinist.

nature can lead to different interpretations of Laplace's[142] formulation and to robustly causal and neo-Humean determinism, respectively. Laplace writes:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble outline of this intelligence. (Laplace 2015[1820]: 12)

Laplace's formulation of the deterministic vision is taken to characterize *physical* determinism. Although criticized,[143] it frequently serves as the template upon which different conceptions of determinism are developed. In the next section, I will show how, once applied to Laplace's remark, different conceptions of laws of nature give rise to metaphysically substantive doctrine of robustly causal determinism (RCD) on the one hand, and neo-Humean determinism (NHD) on the other.

[142] Although Laplace's definition seems to me to be entirely inconsistent with the regularity theory of laws, I will proceed in this chapter as if this is not the case. The reason for this claim is that the intelligence Laplace imagines computes the laws of nature from knowledge of the totality of states of affairs of *one* instant, and hence, does not compute them from the entirety of states of affairs of that universe. Were it to compute the laws of nature from the totality of states of affairs of that universe, laws would be entirely redundant in making the past, present and future appear before its eyes. Nevertheless, the hermeneutical issue of interpreting what Laplace meant does not determine whether another meaning can be given to the idea of determinism.

[143] Earman (1986) criticizes Laplace's formulation as ending up "equating determinism with predictability." (1986: 7) There is no reason to deny that in principle, if one holds that causation is nomological in character. But it seems that determinism does not require causation to be nomic.

B.1.2 *Two versions of Determinism*

A robust view of laws is committed to the existence of irreducible powers whose operation is causal in nature. Facts about such robust causal relations do not boil down to facts about constant conjunction of events. Thus, there is no need to change the terms involved in Laplace's formulation. The state of affairs that holds in the universe at a given instant and which is an effect of its previous state of affairs, is produced by it. Furthermore, the manner in which it is produced is that of causal/physical necessitation. In a characteristically non-Humean spirit, advocates of such a conception of determinism need to posit a productive relation between states of affairs that are causes and states of affairs that are their effects. Therefore, robustly causal determinists conceive of determinism as the position according to which:

RCD (Robustly causal determinism): everything that happens is causally necessitated to happen by antecedent factors.

Appealing to causal necessitation is not the same as appealing to metaphysical necessitation. If something is causally necessitated by its causes, it doesn't mean that it has to obtain in all possible worlds. Yet, RCD appeals to a strong modal force in that it states that given a set of conditions, the effect must follow. Given the specific form of state of affairs (SFSA henceforth) obtaining at a time t, there is only one specific form of state of affairs that is physically possible at time t+1. This is not merely logical, since what is not part of that specific form of state of affairs cannot happen at that time. Deterministic law governedness, construed along the lines of robustly causal determinism implies that deterministic laws capture the manner in which causes bring about their effects. To the extent that a statement of the totality of states of affairs at a given instant and a comprehensive statement of laws of nature *entail* the totality of states of affairs that obtain or will obtain at the next instant, such 'entailment' is

metaphysically rich. It represents the causal necessitation that characterizes causal relations that are instantiated in a deterministic universe.

Neo-Humean determinism (NHD henceforth) rests on the denial of any productive relation between causes and their effects. Thus, the notion of causal necessitation employed in RCD cannot be used by the neo-Humean determinist. In contrast to robustly causal determinists, advocates of NHD argue for a reductively causal form of determinism, where facts about causal relations reduce to facts about non-accidental regularities of events.

The essential difference between RCD and NHD is the former's emphasis on causal necessitation and the latter's emphasis on 'entailment'. Whereas the advocate of RCD claims that a state of affairs is brought about by the operation of forces upon the previously existing states of affairs in the manner of causal necessitation, advocates of NHD refer to the concept of entailment as a logical concept, meant to refer to relations among propositions or statements[144] rather than productive relations of effects by their causes. This formulation is frequently called the 'entailment thesis'[145]:

NHD (Neo-Humean Determinism): A comprehensive statement of the totality of states of affairs that obtain in a universe at an instant, in conjunction with a comprehensive statement of that universe's laws of nature, entail the comprehensive statement of the totality of states of affairs that obtain at the next instant, or at any other instant in that universe's duration.

[144] Whether entailment is conceived of as holding between propositions or statements seems to be an outcome of a thinker's semantical commitments. Propositions, as opposed to statements, are supposed to be abstract entities. Thinkers who are do not accept the existence of abstract entities, may formulate NHD in terms of statements.

[145] Such 'entailment' is also characteristic of nomological versions of RCD, yet in RCD it reflects a necessitating productive relation, whereas in NHD, it emphatically does not.

Deterministic law-governedness, construed along the lines of neo-Humean determinism is incompatible with a metaphysically robust notion of entailment (which is supposed to reflect causal necessitation). The relation of entailment in NHD is purely epistemic; it doesn't have to reflect any robustly causal productive relation. To inhabit a universe governed by deterministic laws does not imply that human actions are causally necessitated. Bernard Berofsky (1987) writes:

> In order for an action A to be determined, it must be the case that a sentence whose truth entails the existence of A follows logically from a law or conjunction of laws L together with a conjunction of sentences S which state initial conditions." (1987: 84)

Due to their explicit rejection of robustly causal relations, advocates of NHD hold that the relation 'B follows from A' is not metaphysically robust. This allows them to counter the idea that determinism is inextricably tied with causal/physical necessitation. The notion of 'entailment' that is employed by the neo-Humean determinist is an epistemic relation that does not reflect causal necessitation.

B.2 ***The question of compatibility***

Having defined RCD (robustly causal determinism) and NHD (neo-Humean determinism), we can now proceed to answer the main question of section B. Since we defined two notions of determinism, we will proceed separately to inquire whether each of them is committed to denying a necessary condition of free will; i.e. free will's fundamental constituents. If not, then free will and that form of determinism which will not be committed to denying its essential constituents will be compatible after all. But if it is shown that both notions of determinism are committed to any of: a) the reducibility of powers, c) denial or reductionism of 'active' doing, and c) denial of

alternative possibilities, then determinism will be shown after all, to be incompatible with free will. Let us take each concept separately.

B.2.1 ***Free will and RCD***

Robustly causal determinism is the thesis according to which, anything that happens is causally necessitated to happen. In the words of Richard Taylor: "Determinism is the thesis that whatever occurs occurs under conditions given which nothing else could have occurred." (1958: 224) The actions that take place, are causally necessitated to be performed. Causal necessitation precludes alternative possibilities, and as such, renders it impossible for the agent in the circumstances she is situated, to do otherwise. Given antecedent conditions in which the agent is situated, her action is necessitated and therefore, it is not up to her whether she will exercise the relevant capacities or not. Therefore, RCD commits us to a passive conception of abilities where their exercise is the causally necessary effect of the way the agent is acted upon. Whether their exercise will lead into the agent altering objects in the agent's environment or to changes internal to the agent, the agent does not have two-way powers.

RCD commits us to a view of our actions that conceives of them as the exercise of one-way powers. In robustly causal deterministic universes the agent has no control over whether her powers are going to be exercised or not at any given moment. Since RCD involves the view that the causal relation involves causal necessitation of the effect by its cause, the formation of the agent's intention as well as the agent's choices and actions, are necessitated by factors over which the agent has no say. That renders the agent's powers fundamentally one-way powers; powers which necessarily manifest given certain conditions. Therefore, robustly causal determinism is committed to two out of the three conditions which would make free will impossible. RCD is inextricably linked with the denial of alternative possibilities and the denial of the active conception

of ‘doing’, since it treats all powers as being one-way powers. A fortiori, robustly causal determinism is incompatible with free will. Let us formalize the argument:

Argument against the compatibility of free will and robustly causal determinism.

1) If RCD is true, then only one SFSA is physically possible at an instant, given the SFSA that held at the previous instant.

2) Mutually exclusive courses of action, A1 and A2, cannot both be part of the same SFSA; that would involve the possibility of a contradictory SFSA.

3) If A1 was causally necessitated to happen given the SFSA at the previous instant, A2 was not physically possible given that very SFSA, and vice versa.

4) A situated agent’s ability to do otherwise requires *at least* two courses of action to be physically possible at an instant, given the SFSA holding at the previous instant.

5) From (2) we get that the ability to do otherwise requires at least two SFSA to be physically possible at an instant, given the SFSA holding at the previous instant.

6) What requires at least two SFSA to be physically possible at an instant cannot be compatible with what allows for only one.

7) From (1), we get that RCD allows for only one SFSA to be physically possible at an instant, given the SFSA holding at the previous instant.

8) Therefore, from (4), (6) and (7), RCD cannot be compatible with the ability to do otherwise.

Premise (1) follows from the definition of robustly causal determinism. RCD allows for only one SFSA to be possible at an instant, given the SFSA holding at the previous instant, rendering thus, what is not involved at that SFSA to be physically impossible. Stathis Psillos (2002) writes: "According to determinism, every event that occurs has a fully determinate and sufficient set of antecedent causes. Given these set of causes, its probability of happening is unity." (2002: 251) The probability of what is not involved in the SFSA at an instant, given the SFSA at the previous instant, is 0.

Premise (2) follows from the law of non-contradiction.[146] Two alternative courses of action cannot exist in the same sense, at the same time and in the same universe.[147] One cannot have raised a hand and not raised it, simultaneously and in the same sense. An SFSA at an instant which involves the raising of a person's hand is an SFSA that does not involve that person not raising the same hand, at that instant and in the same sense of 'raising' a hand.

Premise (3) states that for the robustly causal determinist, what is physically possible to take place at an instant, given specific antecedent conditions, is what given

[146] Perhaps those who accept Mill's thesis, according to which the fact that we cannot conceive of a contradiction does not entail its impossibility, will reject premise 2. If the criticism of radical empiricism in Ch. II/C is correct though, they cannot do so without denying that we can have knowledge, which would pose strain on the intended knowledge claim: free will is compatible with determinism.

[147] A logical empiricist could make the following formulation: two things cannot be involved in the same description, because they are mutually exclusive. This formulation though, seems to make the law of non-contradiction as referring to *descriptions* of states of affairs, as opposed to states of affairs themselves. This seems to me to be essentially the main defect of logical empiricism; in making the law of non-contradiction referring to descriptions, the law itself does not reflect something about the world; it reflects our adoption of a convention of description. So in rejecting something, the logical empiricist loses also sight of truth, because she says: I do not reject something because it is false; I reject it because this is the convention I use in rejecting *and no question arises as to whether this convention is correct* or related to truth in any way; for if there was such a way, we would fall into the rationalist position where such laws are discovered and signified by, rather than constructed by our conventions. This amounts to another way in which logical empiricism falls into the same trap with radical empiricism; they lose sight of truth.

these conditions *will* take place at that instant. What will not happen, was not physically possible given the previously holding SFSA. If antecedent conditions necessitate A1, A2 was not physically possible to take place given these exact antecedent conditions. If antecedent conditions necessitate A2, A1 was not physically possible given these conditions. Since the probability of that which is causally necessitated, given certain conditions equals to unity, if A1 is causally necessitated, the probability of A2 is zero, and vice versa.

Premise (4) states that for one to have the ability to do otherwise, she would have to be able to engage in either of at least two different courses of action, whose outcomes, from premise (2) cannot be part of the same SFSA. The agent would have to have two-way powers whose manifestation would not be certain given the antecedent conditions, which advocates of RCD hold as causally necessitating action.

Premise (5) is a consequence of premises (2) and (4). It states that in order for it to be physically possible that she can act either way, it would have to be the case that at the moment prior to action, each SFSA –either of the SFSA that involve mutually exclusive actions- was physically possible at the next instant. If I live in a deterministic universe and I deliberate whether I will pull my rifle's trigger or not, then only one SFSA that the universe assumes at the time of my action, is physically possible given the SFSA holding at the instant prior to my action. That SFSA will either involve me pulling the trigger or not. It cannot involve both. Assuming determinism, one SFSA is physically possible. If that SFSA is trigger-pulling-involving, then it would be physically impossible for me at the moment prior to action, not to pull the trigger at the next moment because all other courses of action or non-action would have to be involved in SFSA that are not physically possible given the SFSA holding at the instant prior to my action. If that SFSA does not involve me pulling that trigger, since it is the only possible SFSA given the SFSA holding at the previous instant, it would make it physically impossible for me to have pulled the trigger.

Premise (6) is self-evident. It states that what requires two possibilities to be open at a given instant, cannot be compatible with what allows for only one. Premise (7) invokes the definition of robustly causal determinism as an indication that it offers only one genuine physical possibility, whereas the ability to do otherwise, as we saw above, requires more than one. Sentence (8) is the conclusion which by combination of (4), (6) and (7) states that RCD cannot be compatible with free will, construed as the ability to do otherwise.

As we have seen, RCD does not seem to be compatible with free will. It negates at least two of its constituent elements. It is explicitly committed to the denial of alternative possibilities of action and also to the denial of active 'doing'. If RCD is true, then our actions consist in exercises of one-way powers because we have no control over whether they are going to be exercised or not. We may act according to our will, but if we do so, it is ultimately not up to us. As Thomas Reid writes:

> For to say that what depends upon the will is in a man's power, but the will is not in his power, is to say that the end is in his power, but the means necessary to that end are not in his power, which is a contradiction. (1969[1788]: 266, Book IV, ch.1)

A further question can be raised as to whether RCD is committed to the denial of the fundamentality of powers. On the one hand, if it does, it is dangerously close to a Humean view. This seems very counter-intuitive and if it ends up to be the case, it will be entirely ironic. One reason in favour of such a view consists in pointing out that the robustly causal determinist does seem to analyse abilities and powers after all. To say that 'X has the power to R', does seem to be analysed in terms of 'X will R', after all. Maybe the notion of 'power' cannot be squared with determinism. Whether this is the case or not, there is no room to expand on it here.

B.2.2 ***Free will and NHD***

Neo-Humean determinism is the position according to which, the SFSA that holds at a given instant in a universe, in conjunction with the comprehensive statement of that universe's laws of nature, entail the proposition or statement that describes the SFSA that holds at the next instant, or at any other instant, in that universe. Given that conjunction, the proposition that describes the SFSA holding at any other instant, follows logically. In NHD, as opposed to RCD, the relation of 'entailment' is metaphysically innocuous and does not reflect any existent relation of causal necessitation. It is purely epistemic. Advocates of NHD can consistently hold that although the relation of logical necessity holds, there was nothing that causally necessitated the agent to act in the way the agent ended up acting. Their commitment to a Humean theory of laws of nature allows them to claim that facts about causation rest on nomic facts –facts about a universe's laws of nature- which in turn, supervene on non-nomic facts (Loewer, 2012) and that the conjunction has the form it has, because the proposition that is entailed by it, is about something that has already taken place and the fact that it has, has already been taken into account in the determination of that universe's laws of nature.[148] Thus, there was nothing that causally necessitated the agent to act as the agent did as opposed to another way (Beebee and Mele, 2002).[149]

Its denial of causal necessitation implies that NHD allows for alternative possibilities. Yet the very essence of NHD revolves around its commitment to the

[148] This seems to me to show that Laplace's (2015[1820]) formulation is not consistent with neo-Humean determinism; for, according to Laplace, the intellect can comprehend the laws of the universe and consequently, how the universe was, and will be, by taking into account the SFSA, including powers, of a single moment, whereas in the neo-Humean case, laws of nature are comprehended after everything (the SFSA the universe assumed in its entire life) has been taken into account.

[149] In this respect, they hold Humean compatibilism to face the problem of luck, just like metaphysical libertarianism does.

reducibility of powers; for if powers were fundamental, there would be no way of holding NHD as being distinct from RCD. Facts about causal relations would not rest on facts about laws, themselves based on non-nomic facts. Therefore, NHD denies a crucial aspect of what we frequently associate with free will; the fundamentality of powers. Although it allows for alternative possibilities as the universe unfolds, it does so in a way that seems to diminish the agent's causal power. Lack of causal necessitation is purchased at the price of denying, in a characteristically Humean way, the agent's powers as abilities whose exercise will bring forth an outcome. Of course, the neo-Humean determinist will protest that nothing is missing because the sense of causation that is employed does not require a robust productive relation between cause and effect as a truth-maker for statements about causal relations. Nevertheless, it seems that something is lost from agency if there is no such productive relation in causation. The ability to do otherwise comes dangerously close to the ability for something to happen otherwise.

This picture of the universe depicts it as a long list of paintings. We may organize the paintings in such a way that an intelligible story can be told. We have a painting of someone with her hand raised. A painting of the same person with her hand a bit further down, and then another with her hand even lower. In each painting there are two small apples that she might pick. And then there is a painting with her touching one branch. Each painting could be different. In this sense, she could have picked the other apple. But there is no productive relation between cause and effect. It seems that if we deny that the agent is causally necessitated *because we deny the productive relation between causes and effects*, to say that a person could have done otherwise is just like saying that out of all the pieces of a collection of paintings, one painting could have been different. In the neo-Humean sense of agency, the activity of human beings reduces to being depicted in one way rather than another. And it is precisely for this reason that something seems to be lost from agency. At the absence of other reasons in favour of belief in the compatibility of NHD with free will, we have strong reasons to believe that NHD is incompatible with free will; not because it leaves us with no alternative

possibilities, but because it is difficult to see how the alternative possibilities it allows us are alternative possibilities of action.

Conclusion

This chapter motivates incompatibilism about free will and determinism. It does so, in order to counter attempts to develop a compatibilist account of OIC and PAP, whose proponents aim to turn the main argument of the thesis [Ch. III/A] into an argument for compatibilism instead of libertarianism about free will.

In section A, it is maintained that the notion of free will has been usually understood as referring to 'the ability to do otherwise'. Due to the various distinctions of its constituent parts, this complex notion admits as many interpretations as their possible combinations. The table of four characterizations of free will that is developed, is comprised of the following: FW (c1) is the fundamental/irreducible specific power to actively do otherwise. FW (c2) is the fundamental/irreducible specific power to passively do otherwise. FW (c3) is the non-fundamental/reducible specific power to actively do otherwise. FW (c4) is the non-fundamental/reducible specific power to passively do otherwise. The section continues [A.2.3] with a criticism of conditional analyses of the ability to do otherwise, which are the best example of FW (c2) and FW (c4). An independent argument is given against both views, which attacks directly their core assumption, which amounts to the claim that in order for an agent to be able to do otherwise, a conditional statement of a specific kind has to be true of that agent. It is argued that this cannot account for a situated agent's ability to do otherwise. And to the extent that such analyses are employed in the project of grounding the compatibility between free will and determinism, it is argued that because agents in deterministic universes are also situated agents, such a project is inherently flawed. The way that conditionalist analysis account for the ability to do otherwise leads to the conclusion that the necessary condition for doing otherwise, is necessarily non-actual for a situated

agent at each moment. If the necessary condition for doing otherwise is absent for an agent at each moment, then the agent cannot do otherwise at any moment. The rejection of FW (c2) and FW (c4) leaves us with FW (c1) and FW (c3) as the final rival conceptions of free will. It is held that the former is the best characterization of free will and that the latter refers to the power that agents are claimed to have by advocates of the Humean compatibilist position, which is distinct from Hume's compatibilism. If Humean compatibilism is plausible, then libertarianism about free will can be vindicated along its lines. The chapter proceeds as if this is not the case and discussion of Humean compatibilism is postponed until section B.2.2. Section A concludes that any position that is committed to the denial of fundamental powers, of active doing and of alternative possibilities, is incompatible with free will.

In section B, it is argued that determinism is not compatible with free will. Various complications in defining determinism are discussed and popular expressions that are thought of as describing determinism are found to be inadequate. It is not enough to say that a future state of affairs will be caused by a past state of affairs and the laws of nature. What an adequate conception of determinism has to account for is the *manner* in which the effect is brought about by its cause. Two versions of determinism are singled out: robustly causal determinism and neo-Humean determinism. The former claims that effects are causally necessitated by their causes and that such relations of causal necessitation are reflected by statements about X entailing Y. The second, due to its reversal of the hierarchy of the facts that are involved in the order of understanding of the first one, is committed to no such causal necessitation. It is frequently expressed in terms of 'entailment' of the totality of states of affairs that obtain in a universe at one instant from the statement that captures the totality of states of affairs that obtain in the universe at another instant, in conjunction with the statement that captures that universe's laws of nature. In section B.2.1, an argument is given against the compatibility of free will and robustly causal determinism and in section B.2.2, it is maintained that neo-Humean determinism secures lack of causal necessitation of an agent's action at a high price. It secures us

alternative possibilities, although in a way that is hard to see how they are alternative possibilities of action.

Conclusion of the thesis

This thesis defends libertarian free will. It builds on the insight that beings endowed with rational powers are free, in virtue of their rational faculties. Frequently, arguments for this view are presented in a Kantian spirit. This makes it more difficult to understand its main claim as a genuine ontological claim about the world, as opposed to how the mind must see think of the world. In other words, they purport to be synthetic and a priori, but their Kantian mantle makes such an aspiration inherently unstable. To address this issue, advocates of such arguments have to break with the Kantian conception of synthetic a priori and consciously embrace a traditional –pre-Kantian- conception of it.

The argument of the thesis starts with the claim that rationality has a normative nature. In a sense, rationality is connected with normative reasons. And it proceeds to show that in order to be under the scope of normative reasons, a person has to have libertarian free will.

In Ch. I, the notions of reason, reasons and rationality are explored [Ch. I/A] and the position that a kind of rationality –substantive/objective rationality- is normative is defended [Ch. I/B]. It is maintained that reason, conceived of as *pure reason*, is a faculty that allows us to grasp necessary facts and relations. Reasons are presented as justifiers/favourers, whose justifying/favouring function can be non-reductively understood in terms of what they support. If A is related to B, then the relation of support between A and B, renders A as constituting or generating a reason R to believe in B or perform physical action B. It is mentioned that normative reasons, due to their justificatory nature, raise the question of their scope. In matters of thinking, just like in matters of physical action, it is not always the case that we are under the scope of normative reasons. It is argued that a frequent mistake in thinking about when an agent is under the scope of normative reasons, occurs when the conditions for being under the scope of normative reasons are treated as identical to the conditions that have to obtain in order for an obligation to be successfully discharged. This leaves no room for

error, in the sense of an unjustified act, whether mental or physical. If a moral obligation is not discharged, then the agent cannot be said to be unjustified in acting as she did because in order to be justified or unjustified in what she did, she had to be under the scope of that moral obligation. If an epistemic reason is not followed, then the agent cannot be irrational in light of that reason because the agent was not under its scope to begin with. It is argued that to resolve this issue, this identity of conditions has to be broken. To do so in a way that is sensitive to the agent's knowledge and abilities requires us to recognize alternative possibilities of action as a requirement for being under the scope of normative reasons and obligations. Furthermore, the same rationale can be applied to broadly causal conceptions of reasoning that end up equating 'engaging in reasoning' with 'engaging in reasoning *correctly'*. This also leaves no room for incorrect –i.e. unjustified- reasoning. In section B, two objections to the normativity of rationality are discussed. The first amounts to the claim that rationality is not essentially connected with reasons. The second is based on an error theory about normativity in general, which pronounces normative to be 'queer' because it is not possible to reduce it to a naturalistic worldview. It is argued that the implications of the first objection are vastly exaggerated. There does seem to be a distinction between substantive/objective rationality and structural/subjective rationality. In contrast to the latter, the former is essentially connected with normative reasons. To deny the normativity of the latter is not to deny the normativity of the former, at pains of embracing a radical epistemic skepticism about justification and knowledge, in general. The second objection stems from naturalism and is treated in a twofold way. First, it is argued that it leads to a radical epistemic skepticism. Second, it is argued that it is unmotivated and that the pressure to naturalize everything is unjustified. Embracing a non-naturalistic worldview, which is argued for in Ch. II, makes the error theory about normativity unmotivated.

The task of chapter II is twofold. It aims to develop a rationalistic theory of a priori justification in order to reframe the main argument of the thesis in a non-Kantian

framework and to rebut physicalism about the mind and its faculties. The rationale of the chapter can be capsulized in the claim that a debate about a priori justification is an ontological debate as much as it is an epistemic one and that without some a priori component in our thinking, we are threatened with a radical epistemic skepticism. A debate about the ultimate scope of the mind's faculties is, simultaneously, a debate about their nature.

Conceptual matters that concern the notion of a priori justification are discussed in Ch. II/A. A priori justification is defined as justification that is independent of sense-experience, where independence is conceived of in a temporally restricted sense. Sense-experience is required as an enabling condition of a priori justification, but not as what ultimately serves as the source of such justification. In this respect, the notion of a priori justification that is employed is closer to Kant than to the one that features in the Platonist myth, according to which, the soul has perfect knowledge of the Forms before it is attached to a body, in virtue of which sense-experience arises. Ch. II/B involves the presentation of a deficiency argument against empiricism, as well as the presentation of certain dominant strategies with respect to the debate of a priori justification, which are: rationalism, Kantianism, logical empiricism and radical empiricism. For issues of space, responses to Kantianism and logical empiricism have been omitted and radical empiricism is challenged in Ch. II/C. In Ch. II/C it is argued that radical empiricism with respect to a priori justification leads us eventually to a radical form of epistemic skepticism. Essentially, it is characterized by its commitment to Mill's thesis which claims that inconceivability does not entail impossibility. This leads to a thoroughly unsatisfactory treatment of epistemic principles. It is argued that epistemic principles that constitute a theory have to fulfil at least two functions. The first function instructs us when a belief revision has to take place. In those circumstances, specified by the principle that fulfils the first function, the principle that fulfils the second function instruct us what specific revision to make. Radical empiricism involves an unstable commitment to the principle of non-contradiction as the main epistemic principle which determines when a belief revision has to take place.

It is argued that radical empiricism ends up losing sight of truth in that due to its commitment to Mill's thesis –the position according to which the fact that contradictions are inconceivable does not imply that they are false- it ends up with an uncertainty as to whether the cases that we think require a belief revision, are cases where we have encountered a truth or a falsity, as well as whether employing further principles in such cases, gets us closer to truth or not [Ch. II/C.1.1]. Furthermore, it is argued that Mill's thesis commits us to a problematic diagnosis of all normative practices, since we can never know if someone satisfies a standard *as opposed to violating* it, or violates it *as opposed to satisfying* it, or has done both, whatever that may mean [Ch. II/C.1.2]. In Ch. II/C.1.3, certain objections are raised against the rationalistic framework that was developed in the chapter's previous sections. They are answered by pointing out that their advocates tacitly presuppose an empiricist ontology of the mind's faculties. Once we bear in mind that the debate about the ultimate scope of the mind's faculties is simultaneously a debate about their nature, rejecting rational insight on the basis of presupposing an empiricist conception of their nature is circular and misidentifies the nature of the debate.

Ch. III involves the development of the main argument of the thesis and the discussion of six objections to it. It aims to link substantive/objective rationality with libertarian free will. It starts with the premise that substantive/objective rationality is essentially connected with normative reasons. To be under the scope of normative reasons, one needs to have alternative possibilities. Such alternative possibilities are ascribed to situated agents at a time t, in circumstances C (F n); circumstances C having form n. Thus, for an agent to be under the scope of normative reasons, the agent in C (F n) at t, needs to be able to either meet the demand constituted or generated by a reason or violate it, non-accidentally. This seems to require libertarian free will. To the extent that we possess the powers of substantive/objective rationality, we possess libertarian free will. In conjunction with the presuppositions of the thesis, according to which, knowledge requires internalist justification and radical skepticism is false, it

follows that to the extent that we can have knowledge, we possess libertarian free will. Objections are then discussed and answered [Ch. III/B.1-B.6]. These objections revolve around reducing normative facts to evaluative facts [B.1], giving a compatibilist account of OIC and PAP [B.2], claiming that libertarianism involves randomness that is inimical to normative responsibility [B.3], wholly or partly rejecting the importance of alternative possibilities for normative responsibility [B.4], eliminating epistemic normativity in light of doxastic involuntarism [B.5] and maintaining the possibility of a-rationalism [B.6].

Ch. IV involves a rejection of compatibilism about free will and determinism, in order to counter attempts to understand the main argument of the thesis as an argument for compatibilism as opposed to one for libertarianism about free will. Such an attempt is discussed in Ch. III/B.2. In section A, free will is defined as the ontologically fundamental/irreducible specific power to actively do otherwise. The most sophisticated conceptions of free will that do not embrace this definition arise from the compatibilist attempt to give a conditional analysis of the ability to do otherwise. Such attempts are discussed [Ch. IV/A.2-A.2.2] and rejected [Ch. IV/A.2.3]. In section B, two distinct conceptions of determinism are identified and discussed. Robustly causal appeals to causal necessitation, whereas neo-Humean determinism tacitly rejects it. It is claimed that free will is compatible with neither.

A final word has to be said about further research. In a sense, all topics involved in this research require further attention. To be more specific, I will focus on the presuppositions of this project, some positions that are not adequately defended, as well as some questions that arise from this research.

This project is heavily contextualized by its presuppositions and their usual effect amounts to alienating those who do not accept them. Such presuppositions are: i) the falsity of epistemic skepticism, ii) an internalist conception of epistemic justification as a requirement for knowledge, iii) the OIC principle, iv) a verification transcendent conception of truth and v) a conception of reasons as irreducibly normative. Future research on all five topics is salient for enriching it. Skepticism has

always accompanied philosophy. Externalist conceptions of justification are quite popular and the presupposition of their falsity in this thesis may alienate those who embrace them. The OIC principle seems to be on a steadier footing, due to the fact that most ethicists seem to accept it. Nevertheless, the fact that a position is widely embraced does not mean that it is true. The topic of truth is vast and is not so much explored in the thesis. In a sense, accepting a conception of truth that does not transcend verification would make the lives of those who broadly embrace transcendental arguments much easier. Finally, the topic of reasons is vast and really interesting. Many positions exist, such as instrumentalism and non-cognitivism that are not discussed in the thesis. Further research on all these topics would help in the development of the rationalism I defend here and its contextualization in the wider philosophical landscape.

Apart from presuppositions, other positions have been briefly mentioned and not so much discussed. In chapter I, I mentioned physicalist theories of content. Further research on this topic would help in the development of a more sophisticated philosophy of mind for the rationalist framework developed in this thesis. In chapter II, Kantianism and logical empiricism have not been discussed further. A full defense of a rationalist conception of a priori justification has to take a stand on the challenges these schools of thought present us with. Furthermore, dialetheism in logic hasn't been discussed in this thesis. Apart from the fact that dialetheism seems to be a response to semantic paradoxes, the literature that is developed around it is quite elusive and unpredictable in its developments. Advocates of such positions will not find this thesis appealing to them. It is in the nature of this thesis to open more questions than it answers.

Finally, several questions arise from this research that are interesting to explore. The list of rationalists that do not embrace libertarianism about free will is too long to mention. Developing a sophisticated libertarian rationalism seems to be an endeavour that seems worth the effort. The challenges that it has to face revolve around the

viability of a rationalism that does not embrace the principle of sufficient reason in the physical/causal domain and the moral/motivational domain. Since these two come together in the case of action, it seems that what is required to shed more light in the viability of this position revolves around the topic of weakness of will, or akrasia.

Perhaps the list of what remains to be done is an exhausting one and a frustrating one to read as well. One might think that this project answers nothing and creates more problems than it aspires to give a solution to. All that this thesis aspires to do is to make the claim that if we accept certain presuppositions, it may be the case that far from being enemies of free will, rationalism and human situatedness might prove to be its greatest allies.

Bibliography

Alston, W. 1988. "The Deontological Conception of Epistemic Justification" In *Philosophical Perspectives*, vol.2: 257-299

Alvarez, Maria. 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press.

–2013. "Agency and Two-Way Powers." In *Proceedings of the Aristotelian Society*, vol. 113, part I: 101-121.

Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Austin, John. 1979. "If's and Can's", In his *Philosophical Papers*, 3rd edition. Oxford: Oxford Clarendon Press, pp: 205-232.

Ayer, A.J. 1946. *Language, Truth and Logic*. New York: Dover

–1954. "Freedom and Necessity", In his *Philosophical Essays*, London: Macmillan, pp: 271-284.

Bealer, G. 1996. "*A Priori Knowledge and the Scope of Philosophy*", In *Philosophical Studies*, vol. 81, (2-3): 121-142.

–1999. "A Theory of the A Priori", In *Philosophical Perspectives*, vol.13: 29-55.

Bedke, M. 2010. "Might All Normativity be Queer?" In *Australasian Journal of Philosophy*, vol. 88, (1): 41-58.

Beebee, Helen. 2000. "The Non-Governing conception of Laws", In *Philosophical and Phenomenological Research*, vol.61, (3): 571-594.

Beebee, Helen and Mele, Alfred. 2002. "Humean Compatibilism", In *Mind*, vol.111, (442): 201-223.

Benacerraf, P. 1973. "Mathematical Truth", In *Journal of Philosophy*, vol. 70, (19): 661-679.

Berofsky, Bernard. 1987. *Freedom from Necessity*. London: Routledge.

–2012. *Nature's Challenge to Free Will*. Oxford: Oxford University Press.

Berker, S. 2013a. “Epistemic Teleology and the Separateness of Propositions”, In *The Philosophical Review*, vol. 122, (3): 337-393.

–2013b. “The Rejection of Epistemic Consequentialism”, In *Philosophical Issues*, vol. 23, (1): 363-387.

Blanshard, B. 1939. *The Nature of Thought*, vol.1. London: George Allen & Unwin Ltd.

–2013[1962]. *Reason and Analysis*. London: Routledge.

Boghossian, P. 2000. “Knowledge of Logic”, In Paul Boghossian & Christopher Peacocke (eds.), *New Essays on the A Priori*, Oxford: Oxford University Press, pp: 229-254.

–Boghossian, P. & Peacocke, C. (Eds.). 2000. *New Essays on the A Priori*, Oxford: Oxford University Press

BonJour, L. 1985. *The Structure of Empirical Justification*. Cambridge, MA: Harvard University Press.

–1998. *In Defense of Pure Reason*. New York: Cambridge University Press.

–2002. *Epistemology: Classic Problems and Contemporary Responses*. New York: Rowman and Littlefield.

Bratman, M. 1981. “Intention and Means-End Reasoning”, In *The Philosophical Review*, vol. 90, (2): 252-265.

–1987. *Intention, Plans and Practical Reason*. Stanford: CSLI.

Broome, J. 2005. “Does Rationality Give Us Reasons?” In *Philosophical Issues*, vol.15, (1): 321-337.

–2007. “Does Rationality Consist in Responding Correctly to Reasons?” In *Journal of Moral Philosophy*, vol. 4, (3):349-374

–2013. *Rationality through Reasoning*. Oxford: Blackwell.

Burge, T. 1993. “Content Preservation”, In *Philosophical Review*, vol. 102, (4): 457-488.

Carnap, R. 1937. *The Logical Syntax of Language*. London: Routledge and Kegan Paul.

Casullo, A. 2003. *A Priori Justification*. Oxford: Oxford University Press.

–2015. "Four challenges to the a priori-a posteriori distinction", In *Synthese*, vol. 192, (9): 2701-2724.

Clarke, Randolph. 2005. "Agent Causation and the Problem of Luck", In *Pacific Philosophical Quarterly*, vol. 86, (3): 408-421

–2009. "Dispositions, Abilities to Act, and Free Will: The New Dispositionalism" In Mind, Vol.118, (470): 323-351

–2015. "Abilities to Act." In *Philosophy Compass*, vol. 10, (12): 893-904

Copp, D. 1997. "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility", In *Nous*, vol. 31, (4): 441-456.

–2008. "'Ought' implies 'can' and the derivation of the Principle of Alternative Possibilities" In *Analysis*, vol.68, (1): 67-75.

Cowie, C. 2016. "Good News for Moral Error Theorists: A Master Argument Against Companions in Guilt Strategies", In *Australasian Journal of Philosophy*, vol.94, (1): 115-130.

Cuneo, T. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

Das, R. 2017. "Bad News for Moral Error Theorists: There Is No Master Argument Against Companions in Guilt Strategies", In *Australasian Journal of Philosophy*, vol. 95, (1): 58-69.

Davidson, D. 1963. "Actions, Reasons and Causes", In *Journal of Philosophy*, vol. 60, (23): 685-700.

–2004. "Paradoxes of Irrationality", In his *Problems of Rationality*, Oxford: Oxford University Press, pp: 169-188

Della Rocca, Michael. 2008. *Spinoza*. New York: Routledge.

Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press.

Earman, John. 1986. *A Primer on Determinism*. Boston: D Reidel Publishing Company.

Egonsson, D. et al. (Eds.). 2001. *Exploring Practical Philosophy: From Action to Values*, Burlington, Vt.: Ashgate Press

Enoch, D. 2006. "Agency, Schmagency: Why Normativity Won't Come From What is Constitutive of Agency", In *Philosophical Review*, vol. 115, (2): 169-198.

Fara, Michael. 2008. "Masked Abilities and Compatibilism" In *Mind*, Vol. 117, (468): 843-865.

Fischer, J.M. & Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.

Fodor, J. 1990. A Theory of Content and Other Essays. Cambridge, Mass.: MIT Press.

Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility", In *Journal of Philosophy*, vol.66, (23): 829-839.

–1971. "Freedom of the Will and the Concept of a Person", In *Journal of Philosophy*, vol. 68, (1): 5-20.

–1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Garrett, D. 2015. *Hume*. New York: Routledge

Ginet, C. 2001. "Deciding to Believe", In Matthias, Steup (Ed.), *Knowledge, Truth and Duty*, Oxford: Oxford University Press, pp: 63-76.

Goldman, A. 1979. "What is Justified Belief?" In George Pappas (ed.), *Justification and Knowledge: New Studies in Epistemology*, London: D. Reidel Publishing, pp: 1-25.

Graham, P.A. 2010. "In Defense of Objectivism about Moral Obligation", In *Ethics*, vol. 121, (1): 88-115.

–2011. "'Ought' and Ability", In *Philosophical Review*, vol. 120, (3): 337- 82.

Grzankowski, A. 2014. "Can and the Consequence Argument", In *Ratio*, vol. 27, (2):173-189.

Hacker, P.M.S. 1996. *Wittgenstein's Place in Twentieth-Century Philosophy*. Oxford: Blackwell.

Haji, I. 2012. *Reason's debt to freedom: Normative appraisals, Reasons and Free Will*: Oxford: Oxford University Press.

–2018. "Obligation, Responsibility and History", In *The Journal of Ethics*, vol. 22, (1): 1-23.

Hampton, J. 1998. *The Authority of Reason*. Cambridge: Cambridge University Press.

Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.

–2001. "General Foundations versus Rational Insight", In *Philosophy and Phenomenological Research*, vol. 63, (3): 657-663.

Hasker, W. 1973. "The Transcendental Refutation of Determinism", In *Southern Journal of Philosophy*, vol. 11, (3): 175-183.

Hobart, R.E. 1934. "Free Will as involving Determination and Inconceivable without it", In *Mind*, vol. 43, no.169: 1-27.

Hobbes, Thomas. 1982[1651]. *Leviathan*. London: Penguin Classics.

Hobhouse, L. T. 1911. *Liberalism*. [EBook] Available at: <https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/hobhouse/liberalism.pdf> [Accessed 10 December 2019].

Hook, S. (Ed.). 1958. *Determinism and Freedom in the Age of Modern Science*. London: Collier Books.

Huenemann, Charlie. 2008. *Understanding Rationalism*. New York: Routledge.

Hume, David. 1985[1739]. *A Treatise of Human Nature*. Edited by Ernest Mossner. London: Penguin Books.

–2007[1748]. *An Enquiry into Human Understanding*. Edited by Peter Millikan. Oxford: Oxford World's Classics.

Jackson, F. 1991. "Decision- Theoretic Consequentialism and the Nearest and Dearest Objection," In *Ethics*, vol.101, (3):461-482.

James, William. 1956[1897]. "The Dilemma of Determinism", In his *The Will to Believe and other essays in popular philosophy*, New York: Dover, pp: 145-183.

Jordan, J.N. 1969. "Determinism's Dilemma", In *The Review of Metaphysics*, Vol.23, (1): 48-66.

Juhl, C. and Loomis, E. 2010. *Analyticity*. New York: Routledge.

Kane, R. (Ed.). 2011. *The Oxford Handbook of Free Will*, 2nd edition. Oxford: Oxford University Press

Kant, I. 1929[1787]. *Critique of Pure Reason*. Translated by Norman Kemp Smith. London: Macmillan.

Katz, J.J. 1998. *Realistic Rationalism*. Cambridge, MA: MIT Press.

Kelly, T. 2003. "Epistemic Rationality as Instrumental Rationality: A Critique", In *Philosophy and Phenomenological Research*, vol. 66, (3): 612-640.

–2007. "Evidence and Normativity: A reply to Leite", In *Philosophy and Phenomenological Research*, vol. 75, (2):465–474.

Kiesewetter, B. 2011. "Ought and the Perspective of the Agent." In *Journal of Ethics and Social Science*, Vol.5, (3): 1-24.

–2017. *The Normativity of Rationality*. Oxford: Oxford University Press.

Kim, J. 1988. "What is Naturalized Epistemology?" In *Philosophical Perspectives*, vol. 2: 381-405.

Kitcher, P. 1980. "A Priori Knowledge", In *The Philosophical Review*, vol. 89, (1): 3-23.

Kolodny, N. 2005. "Why be Rational?" In *Mind*, vol.114, (455): 509-563.

–2007. “State or process requirements”, In *Mind*, vol. 116, (462): 371-385.

Koons, R.C. & Pickavance, T.H. 2015. *Metaphysics: The Fundamentals*. Oxford: Wiley Blackwell.

Korsgaard. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

–2009a. *Self-Constitution: Agency, Identity and Integrity*. Oxford: Oxford University Press.

–2009b. “The Activity of Reason”, In *Proceedings and Addresses of the American Philosophical Association*, vol.83, (2): 23-43.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Laplace, Pierre-Simon. 2015[1820]. *A Philosophical Essay on Probabilities*. Translated by F.W. Truscott & F.L. Emory. New York: Chronicon Books.

Lavin, D. 2004. “Practical Reason and the Possibility of Error”, In *Ethics*, vol.114, (3): 424-457.

Lehrer, Keith. 1968. “Cans without Ifs”, In *Analysis*, vol.29: 29-32.

Leite, A. 2007. “Epistemic Instrumentalism and Reasons for Belief: A Reply to Tom Kelly’s “Epistemic Rationality as Instrumental Rationality: A Critique”.”, In *Philosophy and Phenomenological Research*, vol.75 (2): 456-464.

Lewis, David. 1981. “Are we free to break the laws?” In Gary Watson (ed.), *Free Will*, 2nd edition. Oxford: Oxford University Press, pp: 122-129.

–1986. *A Plurality of Worlds*. Oxford: Blackwell Publishing.

–1997. “Finkish Dispositions”, In *Pacific Philosophical Quarterly*, vol. 47: 143-158.

Lockie, R. 2018. *Free Will and Epistemology*. London: Bloomsbury

Loewer, Barry. 2012. “Two accounts of laws and time”, In *Philosophical Studies*, vol. 160 (1): 115-137.

Lord, E. 2017. “What You’re Rationally Required to Do and What You Ought to Do (Are the Same Thing!)”, In *Mind*, vol.126, (504): 1109-1154.

Lowe, E.J. 2008. *Personal Agency*. Oxford: Oxford University Press.

Lucas, J. 1970. *The Freedom of the Will*. Oxford: Oxford Clarendon Press.

Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin

Mayr, Erasmus. 2011. *Understanding Human Agency*. Oxford: Oxford University Press.

McKenna, M. and Pereboom, D. 2016. *Free Will: A Contemporary Introduction*. New York: Routledge.

Mele, A. 2003. “Agent’s Abilities”, In *Nous*, vol. 37, (3): 447-470.

–2006. *Free Will and Luck*. Oxford: Oxford University Press.

Mill, J.S. 1884[1843]. *A System of Logic*. London: Longmans, Green, and Co.

–2015. *On Liberty, Utilitarianism and other Essays*. Oxford: Oxford University Press.

Millikan, R.G. 1984. *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Moore, G.E. 1969[1912]. *Ethics*. London: Thornton Butterworth Limited.

Nelkin, D. 2008. “Responsibility and Rational Abilities: Defending an Asymmetrical View”, In *Pacific Philosophical Quarterly*, vol. 89, (4): 497-515.

–2011. *Making Sense of Freedom and Responsibility*. Oxford: Oxford University Press.

O’Connor, T. 2000. *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press.

–2011. “Agent Causal Theories of Freedom”, In Robert Kane (ed.), *The Oxford Handbook of Free Will*, 2nd edition, Oxford: Oxford University Press, pp: 309-328.

Olson, J. 2011. “Error Theory and Reasons for Belief”, In Andrew Reisner & Asbjørn Steglich-Petersen (eds.), *Reasons for Belief*, Cambridge: Cambridge University Press, pp: 75-93.

–2014. *Moral Error Theory*. Oxford: Oxford University Press.

Owens, D. 2000. *Reason without Freedom: The Problem of Epistemic Normativity*. London: Routledge.

Pap, A. 1958. *Semantics and Necessary Truth*. New Haven: Yale University Press.

Pappas, G. (Ed.). 1979. *Justification and Knowledge: New Studies in Epistemology*, London: D. Reidel Publishing

Parfit, D. 2001. “Rationality and Reasons”, In Dan Egonsson et al. (Eds.), *Exploring Practical Philosophy: From Action to Values*, Burlington, Vt.: Ashgate Press, pp: 17-40.

–2011. *On What Matters*, vol.1. Oxford: Oxford University Press.

Peacocke, C. 2004. *The Realm of Reason*. Oxford: Oxford University Press.

Pereboom, D. 2014a. *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.

–2014b. “The disappearing agent objection to event-causal libertarianism”, In *Philosophical Studies*, vol.169, (1): 59-69.

Priest, G. 2006. *Doubt Truth to be a Liar*. Oxford: Oxford University Press.

Psillos, Stathis. 2002. *Causation and Explanation*. London: Routledge.

Putnam, H. 1983. “Two Dogmas Revisited”, In *Realism and Reason: Philosophical Papers*, vol.3. Cambridge: Cambridge University Press.

Quine, W. V. O. 1953. “Two Dogmas of Empiricism”, In his *From a Logical Point of View*, Cambridge, MA: Harvard University Press, pp: 20-46.

–1969. “Epistemology Naturalized”, In his *Ontological Relativity and Other Essays*, New York: Columbia University Press, pp: 69-90.

–1986. *Philosophy of Logic*, 2nd edition. Cambridge, MA: Harvard University Press.

Quinton, A. 1964. "The A Priori and the Analytic", In *Proceedings of the Aristotelian Society*, vol. 64, (1): 31-54.

Raz, J. 1999. *Engaging Reason*. Oxford: Oxford University Press.

–2011. *From Normativity to Responsibility*. Oxford: Oxford University Press.

Reid, Thomas. 1969[1788]. *Essays on the Active Powers of the Human Mind*. Edited by Baruch Brody. Cambridge, Mass: MIT Press.

Reisner, A. & Steglich-Petersen, A. (Eds.). 2011. *Reasons for Belief*, Cambridge: Cambridge University Press

Rescher, N. 1988. *Rationality: A Philosophical Inquiry into the Nature and the Rationale of Reason*. New York: Oxford Clarendon Press

–2015. *Free Will: A Philosophical Reappraisal*, 2nd edition. New York: Transaction Publishers.

Robertson, S. (Ed.). 2009. *Spheres of Reason: New Essays in the Philosophy of Normativity*. Oxford: Oxford University Press.

Russell, Bertrand. 1998[1912]. *The Problems of Philosophy*. Oxford: Oxford University Press.

Russell, Bruce. 2014. "A Priori Justification and Knowledge", *In Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/apriori/> [Accessed 5 February 2019].

Salmon, W. C. 1967. *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Scanlon, T.M. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.

–2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.

Schlosser, Mark. 2017. “Traditional Compatibilism Reformulated and Defended.” In *Journal of Philosophical Research*, vol.42: 277-300.

Schnall, Ira. 2001. “The Principle of Alternate Possibilities and ‘Ought’ implies ‘Can’”, In *Analysis*, vol. 61, (4): 335-340.

Schopenhauer, A. 1969[1819]. *The World as Will and Representation*, vol.1. Translated by E.F.J. Payne. New York: Dover Publications.

–1974[1813].*On the Fourfold Root of the Principle of Sufficient Reason*. Translated by E.F.J. Payne. Indianapolis: Hackett.

Segal, G. & Sober, E. 1991. “The Causal Efficacy of Content”, In *Philosophical Studies*, vol. 63, (1): 1-30.

Shafer-Landau, R. 2009. “A Defense of Categorical Reasons”, In *Proceedings of the Aristotelian Society*, vol. 109, part 2: 189-206.

Shah, N. 2011. “Can Reasons for Belief be debunked?” In Andrew Reisner & Asbjørn Steglich-Petersen (eds.), *Reasons for Belief*, Cambridge: Cambridge University Press, pp: 94-107.

Sinnott-Armstrong, W. 1984. “‘Ought’ conversationally implies ‘can’”, In *Philosophical Review*, vol. 93, (2): 249-261.

Skorupski, J. 2010. *The domain of reasons*. Oxford: Oxford University Press.

Slagle, J. 2016. *The Epistemological Skyhook: Determinism, Naturalism and Self-Defeat*. New York: Routledge.

Smith, Adam. 2009[1759]. *The Theory of Moral Sentiments*. Edited by Ryan Patrick Hanley. New York: Penguin Books.

Smith, Michael. 2003. “Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion”, In Sarah Stroud & Christine Tappolet (eds.), *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press, pp: 17-38.

Southwood, N. 2008. “Vindicating the Normativity of Rationality” In *Ethics*, vol. 119 (1): 9-30.

Spinoza, Benedict. 1996[1677]. *Ethics*. Translated and Edited by Edwin Curley. London: Penguin Books.

Steup, M. (Ed.). 2001. *Knowledge, Truth and Duty,* Oxford: Oxford University Press

–2008. "Doxastic Freedom", In *Synthese*, vol. 161, (3): 375-392.

Steward, Helen. 2012. *A Metaphysics for Freedom*. Oxford: Oxford University Press.

Strawson, P.F. 2003[1962]. "Freedom and Resentment." In Gary Watson (ed.), *Free Will*, 2nd edition. Oxford: Oxford University Press, pp: 72-93.

Streumer, B. 2007. "Reasons and Impossibility", In *Philosophical Studies*, vol.136, (3): 351-384.

–2017. *Unbelievable Errors: An Error Theory about All Normative Judgments*. Oxford: Oxford University Press.

Stroud, S. & Tappolet, C. (Ed.). 2003. *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press.

Tahko, T. 2009. "The Law of Non-Contradiction as a Metaphysical Principle", In *Australasian Journal of Logic*, vol. 7: 32-47.

Tanney, J. 1995. "Why Reasons may not be Causes", In *Mind and Language*, vol. 10, (1-2): 103-126.

Taylor, Richard. 1958. "Determinism and the Theory of Agency", In Sidney Hook (ed.), *Determinism and Freedom in the Age of Modern Science*. London: Collier Books, pp: 224-230.

Thomson. 1992. "On Some Ways in Which A Thing Can be Good", In *Social Philosophy and Policy*, vol.9, (2): 96-117.

–2008. *Normativity.* Chicago: Open Court.

Thurow, J. C. 2008. "The A Priori defended: a defense of the Generality Argument", In *Philosophical Studies*, vol. 146, (2): 273-289.

Van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.

–2000. “Free Will Remains a Mystery”, In *Philosophical Perspectives*, vol. 14: 1-20.

Vetter, Barbara and Jaster, Romy. 2017. “Dispositional accounts of abilities” In *Philosophy Compass*, (12): e12432. Available at: <https://doi.org/10.1111/phc3.12432> [Accessed 20 March 2018].

Vihvelin, Kadri. 2004. “Free Will Demystified: A Dispositionalist Account”, In *Philosophical Topics*, vol. 32: 427-450.

–2013. *Causes, Laws and Free Will: Why Determinism doesn’t matter*. Oxford: Oxford University Press.

Vranas, P. 2007. “I Ought, Therefore, I Can”, In *Philosophical Studies*, vol. 136, (2): 167-216.

–2018. ““Ought” implies “Can” but Does Not Imply “Must”: An Asymmetry between Becoming Infeasible and Becoming Overridden”, In *Philosophical Review*, vol. 127, (4): 487-514.

Watson. G. 1975. “Free Agency”, In *Journal of Philosophy*, vol. 72: 205-220.

–(Ed,). 2003. *Free Will*, 2nd edition. Oxford: Oxford University Press.

Wedgwood. 2006. “The Normative Force of Reasoning”, In *Nous*, vol.40, (4): 660-686.

–2007. *The Nature of Normativity*. Oxford: Oxford University Press

–2009. “The Good and the Right Revisited”, In *Philosophical Perspectives*, vol. 23: 499-519.

–2013. “Rational ‘Ought’ Implies ‘Can’”, In *Philosophical Issues*, vol. 23: 70-92.

–2017. The Value of Rationality. Oxford: Oxford University Press.

Whittle, A. 2010. “Dispositional Abilities”, In *Philosophers’ Imprint*, vol.10, (12): 1-23.

Wick, W. 1964. “Truth’s Debt to Freedom”, In *Mind*, vol. 73, (292): 527-537.

Widerker, D. 1991. “Frankfurt on ‘Ought implies Can’ and Alternative Possibilities”, In *Analysis*, vol. 51, (4): 222-224.

Wittgenstein, L. 1961[1922].*Tractatus Logico-Philosophicus*. Translated by D.F. Pears & B.F. McGuiness. London: Routledge and Kegan Paul.

Williams, B. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.

Wolf, S. 1980. “Asymmetrical Freedom”, In *The Journal of Philosophy*, vol. 77, (3): 151-166.

–1990. *Freedom within Reason*. Oxford: Oxford University Press.

Yaffe, G. 1999. “‘Ought’ implies ‘can’ and the principle of alternate possibilities”, In *Analysis*, vol. 59: 218-222.

–2005. “More on ‘ought’ implies ‘can’ and the principle of alternate possibilities”, In *Midwest Studies in Philosophy*, vol.29: 307-312.

Young, J. 2005. *Schopenhauer*. London: Routledge

Zimmerman, M. 2008. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.