

A Tectonic Theory of Moral Responsibility:

how a concern for patiency can make moral responsibility practices more fair

Lucas Jurkovic

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the

Doctorate in Philosophy degree in Philosophy

Philosophy Department

Faculty of Arts

University of Ottawa

© Lucas Jurkovic, Ottawa, Canada 2016

Table of Contents

Abstract	pg. iii
Acknowledgements	pg. v
Introduction	pg. vi
1- Moving into the Middle Ground between Compatibilism and Incompatibilism	pg. 1
2- What is Patency?	pg. 32
3- Empirical Work on Patency	pg. 60
4- The Homogenization Problem	pg. 95
5- The Tectonic Theory of Moral Responsibility	pg. 116
6- The Moral Significance of Neuroscience	pg. 151
Conclusion	pg. 194
References	pg. 200

Abstract

This thesis develops a tectonic theory of moral responsibility. The word ‘tectonic’ is used metaphorically, to bring to mind the interaction of tectonic plates in the earth’s crust. The theory of moral responsibility developed in the thesis posits that there are two aspects of persons that should be considered within moral responsibility judgments: agency and patiency. I suggest that the ways that these relate to and influence each other can be thought of as similar to the interactions between tectonic plates.

I use the term ‘agency’ in its usual sense to refer to the control component of moral responsibility judgments. A person must be an agent, i.e. have a normal degree of self-control in order to be held morally responsible for her actions. This is a vague take on agency, to be sure. However, the tectonic theory is meant to apply to any of the common conceptions of agency in the moral responsibility literature. The vagueness in my treatment of agency is thus intended to allow the tectonic theory to be pluralist in regards to various conceptions of agency on offer. Whether one subscribes to a reasons-responsiveness view of agency, or an identificationist view, the tectonic theory’s prescribed concern for patiency can function as a fairness-enhancing counterweight to the agentic considerations of either view.

The novelty of the tectonic theory derives from its use of the concept of patiency, which has been neglected in philosophical discussions of moral responsibility. Patiency refers to the features of a person and her life that she cannot control. The tectonic theory takes both agentic and patientic features of persons into account when making moral responsibility judgments. I argue that doing

so can enhance the fairness of moral responsibility judgments, while for the most part avoiding the pitfalls of more conventional approaches to moral responsibility such as compatibilism and incompatibilism. To develop the tectonic theory, I draw from work in philosophy, psychology, and cognitive neuroscience.

Acknowledgements

Many people were good to me while I developed this thesis. First and foremost, my partner Laura was relentlessly supportive, kind, and interested. Our wonderful life together has been a perfect setting in which to think. My mother, Martha Watt, and step-father, Michael Gareau, were also very supportive and I'd often dwell on their belief in me when in need of bolstering my own convictions about the project.

I should also thank my sisters, Leah and Rebecca, both for their support and all of the philosophical practice they provided me in dialogue over the years. My father, Josef, and brother, Jacob, were also sources of support, and many welcome diversions. Tansy, also a graduate student in philosophy, ensured that I was never without a real friend on this particular journey. I am also grateful to Dr. Andrew Sneddon, my supervisor, for all the encouragement and many helpful pointers. Finally, I thank Dr. Bruce Waller, whose work was the primary inspiration for this thesis. His generous participation as an examiner of the thesis meant so much to me.

Introduction

This thesis was provoked by my becoming unsettled reading Ishtiyaque Haji's book *Incompatibilism's Allure*.¹ Haji begins the book by recounting the story of Robert Harris, first introduced to the moral responsibility literature by Gary Watson in his paper, 'Responsibility and the limits of evil: variations on a Strawsonian theme.'² In brief, Harris suffered a severely abusive childhood, and grew up to be a vicious murderer. After recounting the story, Haji writes,

It is natural to wonder to what extent Harris was responsible for becoming the sort of person that he was. Was the pathway of his life fully set in stone by parental and societal neglect, social or genetic conditioning, sheer bad luck, and the like, or did he have any hand in molding it? If the former, and this inclines us to mitigate or remove blame, then we may be well on the way toward incompatibilism, loosely, the view that blameworthiness – or more generally, moral responsibility – is incompatible with being “determined” in certain ways.³

I was unsettled because I am inclined to judge that Harris should not be held morally responsible for his crimes, but not because I think that anything in his life determined that he would commit them. To claim that Harris was bound to become a criminal because of his upbringing strikes me as speculative. I do not see how one could be sufficiently certain about the eventuality of his crimes to invoke it as a basis for the mitigation of his moral responsibility. I accept that Harris' childhood made it quite likely that he would develop criminal tendencies, but justifying mitigation on these grounds alone is also unsatisfying to me.

¹ (Haji, 2009)

² (Watson, 1987)

³ (Haji, 2009, pg. 13)

As Haji sets it up,⁴ if Harris can be said to have had ‘any hand in molding’ his adult character, then there are no clear grounds for mitigating or removing his moral responsibility. To my judgment, Harris’ moral responsibility should be removed even if it is the case that he contributed to his own development, as indeed he claims he did.⁵ Thus, I set out looking for work on moral responsibility that explores ways to justify the tempering or elimination of moral responsibility without invoking determinism. I didn’t find much aside from the work of Bruce Waller. Waller is an eliminativist about moral responsibility, meaning that he endorses the comprehensive abolition of moral responsibility practices. However, Waller’s argument for this position is unusual in that he does not claim that moral responsibility is illegitimate because people’s behaviour is determined, but rather because moral responsibility practices do not look closely enough into the details of particular cases to be deemed fair.

I was convinced by Waller that moral responsibility practices are unfair, but I do not think that this is sufficient support for his eliminativist position. As Waller himself concedes when pressed by Daniel Dennett’s criticisms, as problematic as moral responsibility practices are, there are not really any alternatives.⁶ I therefore took up the challenge of finding a way to think about moral responsibility that can make related practices less unfair but that neither appeal to determinism nor call for eliminativism. This led to reflections about what sorts of facts about persons get left out of account in moral responsibility judgments and why. I was also confronted with the

⁴ It is worth noting that Haji’s remarks on Robert Harris are intended to introduce the standard incompatibilist view of such a case. My dissatisfaction with the view is not intended as a criticism of Haji’s book, which carefully takes stock of various arguments for incompatibilism. On the contrary, I reference Haji here because I take him to successfully articulate the substance of incompatibilist views.

⁵ Watson cites a newspaper article which quotes Harris’ sister as saying that Robert understood himself to have had an opportunity to be otherwise: “He told me he had his chance, he took the road to hell and there’s nothing more to say.” (Watson, 1987. Fn. 22)

⁶ Waller writes, “Dennett saves his toughest challenge for last. So you don’t think this “fair enough” system is acceptable? Fine; show us a better one. ... I should be honest: I don’t have such a system to offer.” (Waller, 2012)

challenge of explaining how such facts might be integrated into moral responsibility judgments without destabilizing them. These are the concerns that led me to the concept of patiency.

Patiency is not a common term in the moral responsibility literature or any philosophical literature for that matter. It has been used occasionally in the philosophy of action, wherein ‘patients’ refers to the recipients of action.⁷ This usage traces back to Aristotle, who theorized that actions involve agents who perform the action and patients that receive the action.⁸ Richard Taylor wrote about patiency in a paper called ‘Agent and Patient: is there a difference’⁹ that argues against the prospects of clarifying the nature of actions by sorting the world’s inventory into two classes: agents that act, and patients that are acted on. Taylor concludes that there is no stable way to distinguish between the things that act and those that are acted on. If Taylor is right, and I think he is, then there is no principled barrier to developing the concept of patiency as it applies to agents. In other words, agents are also patients. Even if a person performs agentic actions, she is still subject to numerous factors that her actions could not have altered. Such factors comprise a person’s patiency, and facts about them are the sort of thing that moral responsibility judgments often leave out of account.

Looking through both the philosophical and psychological literature on moral responsibility, I discovered a few authors who address patiency. There is empirical work from Kurt Gray, Daniel Wegner and colleagues, and Phillip Robbins and Anthony Jack and colleagues that seems to

⁷ See (Dretske, 1991), (R. Taylor, 1982), and (Karlsson, 2002)

⁸ See (Karlsson, 2002) for an evaluation of Aristotle’s conception of patiency in comparison to Dretske’s.

⁹ (R. Taylor, 1982)

indicate that whether a person is seen as an agent or a patient makes a difference to moral thinking. Roughly, when a person is seen as an agent, this inclines observers to hold her morally responsible for her behaviour, and when a person is seen as a patient, this inclines observers to grant her a moral status that entitles her to protection from harm. Amongst philosophers, there seems to have been only one person who worked directly on patiency, Soran Reader. Sadly, Reader died in 2012, but her paper ‘The other side of agency’¹⁰ compellingly argues that personhood is partially constituted by patiency and that philosophers’ typically agency-centric conceptions of personhood fail to do justice to the full suite of phenomena involved in being a person. With these ingredients, I put forward the tectonic theory of moral responsibility.

The tectonic theory sensitizes moral responsibility judgments to uncontrolled factors that are relevant to persons’ actions, but without the usual trappings of incompatibilist skepticism about what persons do control. I call the theory *tectonic* because it describes two layers of characteristics of a person – agentic and patientic characteristics. The geological metaphor is intended to conjure images akin to the interaction of tectonic plates in the Earth’s mantle: a dyadic system of agentic and patientic characteristics sliding along and influencing one another in subtle and sometimes unpredictable ways. Like two tectonic plates, agency and patiency provide enduring context to one another as well as the occasional episode of acute influence.

The basic idea is to swap the appeals to determinism that are usually present in critiques of moral responsibility for a concern for persons’ patiency. There is an important difference between

¹⁰ (Reader, 2007b)

thinking about a person as subject to determinism and as a patient. The thesis of determinism articulates a view of the world wherein persons are controlled by factors like genetics and acculturation. This is in conflict with the basic idea of morally responsible agency, because moral responsibility is supposed to be legitimized by persons' control over their own behaviour. It is difficult to square the idea that persons have some control over themselves with the idea that persons are controlled by genetic and environmental factors. Considering persons' patiency, in contrast, involves attending to factors that persons are unable to control, which coexists more easily with recognition of the fact that people typically do have the sort of self-control that legitimizes moral responsibility. Instead of having to choose between the view that persons are in control of themselves, and the view that persons are controlled by their genes and environment, the tectonic theory's use of patiency allows for considering both what people control and what they don't.

The tectonic theory prescribes a two-step process for determining moral responsibility. First, whether a candidate for moral responsibility is indeed an agent should be ascertained. After an individual is deemed an agent, she should be looked at as a patient to give factors beyond her control an opportunity to mitigate or even vitiate her moral responsibility for a given action. This thesis argues that if this formula were to be operationalized, it is plausible that moral responsibility judgments will become less unfair. The remainder of the introduction will provide summaries of the different steps taken to make this case.

The first chapter of the thesis ‘Moving into the Middle Ground between Compatibilism and Incompatibilism,’ presents the pitfalls of the usual approaches to moral responsibility. Focusing on the work of Saul Smilansky, I argue that both compatibilism and incompatibilism are problematic in that each is prone to leaving significant details about particular cases out of account. I introduce the tectonic theory of moral responsibility and explain how it avoids these problems. The main goal of the chapter is to address and leave behind the traditional debate between compatibilism and incompatibilism.

The second chapter of the thesis, ‘What is Patency?’, focuses on Reader’s paper mentioned above. I present Reader’s arguments for including patency in conceptions of personhood, and show how being alert to persons’ patency can enhance the fairness of moral responsibility judgments. I also provide a definition of ‘patency’ based on Daniel Dennett’s analysis of the concept of control from *Elbow Room: on the varieties of free will worth wanting*.¹¹ As with agency, my appeal to Dennett is meant to provide a minimal definition of control so that the contrast between agency and patency can be made clear without requiring an adventure into existing and heavily freighted discussions about how best to articulate the control element of moral responsibility. The tectonic theory takes a pluralist approach to agency, applying to both reasons-responsive and identificationist views. Thus, Dennett’s relatively abstract definition of control serves well to enunciate the difference between agency and patency while leaving room for deploying the tectonic theory in conjunction with any extant partisan views about agency. Finally, at the end of the chapter I delineate different types of patency.

¹¹ (D. Dennett, 1984a)

The third chapter of the thesis, ‘Empirical work on Patience,’ presents the findings of Gray, Wegner and colleagues, and Robbins, Jack, and colleagues. Drawing on the work from Gray and Wegner’s team, I show that there is reason to expect the patientic considerations prescribed by the tectonic theory to make moral responsibility judgments less punitive. From the work of Robbins and Jack’s group, I develop an explanation for why moral responsibility judgments can so easily neglect patientic considerations in favour of a purely agentic rubric. Robbins and Jack suggest that neurological features of human brains are such that attending to a person’s agency and patience requires a shift in perspective or cognitive mode. This is important for explaining why the tectonic theory has a two-step structure.

The fourth chapter, ‘The Homogenization Problem,’ focuses on Waller’s critique of moral responsibility. I develop the concept of the homogenization problem to articulate what it is about moral responsibility judgments that makes them unfair. Drawing from Waller’s argument that such judgments fail to look closely enough at the details of persons and their actions, I propose that because moral responsibility judgments are inaugurated by consideration of whether a person is sufficiently agentic to be held responsible, unless something like the tectonic theory’s prescribed concern for patience is in place, such judgments unfairly homogenize across the class of agents. If they are to become less unfair, moral responsibility judgments should not rest at having classified a person as an agent, but instead consider a person’s patience so as to recover some of the nuance left out of account when establishing a person’s agency.

The fifth chapter of the thesis is called ‘A tectonic theory of moral responsibility.’ I present what I call the ‘basic argument for the tectonic theory.’ Based on a small number of what I take to be uncontroversial claims about persons’ actions, I conclude that moral responsibility judgments ought to take both agency and patiency into account. Additionally, I import another notion from geology to fill out the tectonic features of the theory. I suggest that particular tectonic moral responsibility judgments can articulate the dominance of either agentic or patientic factors through the concept of subduction. Subduction is a geological term for the process of one tectonic plate sinking below another and melting back into the earth’s mantle. I present a number of examples (including a detailed treatment of the Robert Harris case) that involve either agentic factors being judged to subduct patientic ones, or vice versa. I also suggest that compatibilist approaches to moral responsibility can usefully be characterized as conceiving of agency as generally subductive of patiency, and that incompatibilist approaches tend to go the other way, taking patientic factors as generally subductive of agentic ones.

The sixth and final chapter of the thesis, ‘The Moral Significance of Neuroscience,’ puts the tectonic theory to work adjudicating a debate about the moral and legal significance of neuroscience. I present a well-known paper by Joshua Greene and Jonathan Cohen called, ‘For the law, neuroscience changes nothing and everything,’¹² as well as responses from a number of critics. I show that a tectonic account of neuroscience’s moral significance can integrate the concerns of both Greene and Cohen, and their critics, while avoiding the problems of Greene and Cohen’s argument. I suggest that neuroscience can be expected to reveal predominantly patientic facts about persons, which can make justice systems more humane as recommended by Greene

¹² (Greene & Cohen, 2004)

and Cohen, but without calling morally responsible agency into question, which prospect is intolerable to their critics.

Chapter 1: Moving into the Middle Ground between Compatibilism and Incompatibilism

“Metaphysics make all one’s thoughts equally corrosive.”

Samuel Coleridge (MS notebook no. 8, pg. 23)

In the moral responsibility literature, some philosophers have become dissatisfied with the traditional dichotomy of compatibilism and incompatibilism. Often these are positions about whether free will is compatible with determinism. Determinism is the view that events in the universe can only go one way, i.e. that whatever happens was bound to happen given the universe’s physical laws.¹³ ‘Free will’ generally refers to the ability to do otherwise. If a person has free will in this sense, her behaviour on particular occasions should be purely a result of her mental states at the time, and these states must be under her control. The thesis of determinism represents a threat to the idea that persons have free will in this sense because if true, determinism entails that any person’s behaviour, whether it conforms to her mental states at the time or not, is entirely a function of the laws of the universe. Thus, the truth of determinism causes trouble for the view that persons have the ability to do otherwise. Compatibilism is the view that persons can be said to have free will even if determinism is true, and incompatibilism is the view that determinism would rule out free will.

¹³ Peter van Inwagen is well known for his formulation of the issues involved in considering the merits of compatibilism and incompatibilism. In *An Essay on Free Will*, he defines determinism thusly: “Determinism is quite simply the thesis that the past determines a unique future.” (van Inwagen, 1983, pg. 1)

Although the problem of free will and determinism can be addressed on a purely conceptual or metaphysical level, most philosophers' that work on it are interested in moral responsibility.¹⁴ Moral responsibility is related to these issues because 'free will' is often taken as a term for whatever it is about persons that makes them appropriate targets of moral responsibility judgments. Thus, many philosophers have shifted focus away from the metaphysical side of moral responsibility and toward the more purely social issues pertaining to moral responsibility practices. This has resulted in compromises involving the two priorities of doing justice to persons' metaphysical nature and doing justice to social issues connected to moral responsibility.

Saul Smilansky might be the most self-conscious compromiser of the bunch. Roughly, Smilansky believes that incompatibilism does greater justice to the way the world really is, i.e. persons do not actually have free will, but he also believes that this insight ought to be resisted so that the valuable institution of moral responsibility is not imperilled. To greet this predicament, Smilansky developed the position known as *willusionism*. This is the view that the incompatibilist insight to the effect that persons do not really have free will ought to be suppressed so that people continue to self-conceive as agents with the capacity to do otherwise. This is the only way, according to Smilansky, that the socially invaluable institution of moral responsibility can endure.

¹⁴ Manuel Vargas describes two distinct currents in the literature on free will that coalesced along these lines. In 'The Revisionist Turn: a brief history of recent work on free will,' he writes, "Conventional wisdom is that there are two major clusters of interests on the free will problem reflected in two populations that approach matters with interests grounded in distinct sub-fields within philosophy. The two groups are the *metaphysicians* and the *ethicists*. The former are more apt to focus on free will and the latter are more apt to focus on moral responsibility. And, the former are more apt to be incompatibilists and the latter seem more evenly split, perhaps even favoring compatibilism." (Vargas, 2010)

However, Smilansky has recently come to think that compatibilist forms of moral responsibility (which he takes to describe the actual practices of contemporary societies, hence the requirement that libertarianism be sustained as an *illusion*) give rise to injustice to such an extent that it might be advisable to episodically shed the libertarian illusion. Doing so would, according to Smilansky, allow for the actual injustice of moral responsibility practices to be recognized and addressed. Significantly, Smilansky characterizes the perspective that has shed the libertarian illusion as hard determinism, the view that nobody is morally responsible for their behaviour because the universe is deterministic, which entails that nobody has ever really been able to avoid the actions they perform. The purpose of this chapter is to correct Smilansky's reliance on metaphysics and articulate a substitute perspective that can render the injustice of moral responsibility practices without invoking deterministic metaphysics.

These two aspects of Smilansky's work, his allegiance to libertarianism as the only possible legitimizer of moral responsibility and his endorsement of oscillating between two different views on moral responsibility so as to balance the priorities of staying alert to injustice and preserving a community of responsibility, represent a strange blend of traditionalism and progressivism on the issue of moral responsibility. Appraising his views, then, provides a good opportunity to convey what is problematic about the traditionally dichotomous rubric that pits compatibilism and incompatibilism against one another in a once and for all battle for the legitimacy of moral responsibility, as well as what is promising about attempts to mix features of the two views.

I will focus on the most recent work by Saul Smilansky¹⁵ as a way of introducing the reasons for the ambivalence that some philosophers feel towards compatibilism and incompatibilism.

Additionally, I will perform a critique of Smilansky's response to the tension between them.

Smilansky encourages philosophers "to work out, in detail, where compatibilism and where hard determinism are more personally and morally salient, and how to integrate them."¹⁶ By his own admission, though, the prospects for this integration are poor, and either of the two perspectives alone is problematic. I will conclude by introducing my tectonic theory of moral responsibility and explaining how its concern for patiency is a more promising way of addressing this tension.

Smilansky's Ambivalence

Much of Smilansky's account of the pros and cons of compatibilism and incompatibilism will be familiar. Like many other philosophers, Smilansky takes compatibilism's central virtue to be that it allows for the preservation of moral responsibility practices. Smilansky's conception of incompatibilism's strengths is also the usual: incompatibilism takes seriously the fact that persons do not have free will in the traditional sense of being able to do otherwise. Smilansky is also alert to incompatibilism's most commonly pointed out weakness: it threatens to drain life experiences and social communities of all the meaning that depends on persons conceiving of themselves and one another as morally responsible. Because these views are more or less standard in the literature, I will only invoke these issues as needed to clarify Smilansky's proposal for addressing them.

¹⁵ (Smilansky, 2012)

¹⁶ (Smilansky, 2012, pg. 238)

Smilansky's account of what is problematic about compatibilism, on the other hand, is less usual. He describes compatibilism as both giving rise to and systematically obscuring injustice. In order to capture the injustice involved in what he calls "compatibilist forms of life," Smilansky thinks it is necessary to revert to a hard determinist perspective that characterizes all behaviours of all persons as inevitable and thus exempt from moral responsibility. But this reversion is short lived because Smilansky needs to return to a compatibilist view in order to lay the blame for the injustice on the way that compatibilism fosters complacency about it. This appears to be the only way that Smilansky thinks the injustice of moral responsibility practices can be compelling.

My tectonic theory of moral responsibility departs from these last two points. I do not think that hard determinism (or any particular metaphysical position) is required to render the injustice involved in moral responsibility practices. Nor do I think that compatibilism necessarily breeds complacency towards it. I do agree with Smilansky that complacency toward the injustices of moral responsibility practices is common and severe enough that theories of moral responsibility should address it. In contrast to Smilansky though, I think that the remedy for this needs to be instituted within moral judgments themselves, rather than within the metaphysical stances that purportedly act as premises for them. At the end of this chapter, I will introduce the tectonic theory to explain how this can be done. For now, I turn to a more detailed presentation Smilansky's account.

Traditionally, compatibilism and incompatibilism are doctrines about the theoretical compatibility of free will and determinism. The problems that Smilansky discerns in these views,

however, are more concerned with social and normative issues pertaining to moral responsibility practices. In fact, he reverses the conventional direction of influence between the free will/determinism issue and practical concerns about moral responsibility. Rather than taking verdicts on the compatibility issue to be instructive in regards to the practical issues concerned with moral responsibility, Smilansky thinks that inferences about free will and determinism should be driven by the practical issues.¹⁷

I think that Smilansky's approach is the right one insofar as it gives the practical side of moral responsibility issues priority over the more traditional metaphysical issues. While I will dispute Smilansky's contention that libertarian beliefs are required for sustaining the moral responsibility system, the core of his appraisals of compatibilism and incompatibilism can be decoupled from this commitment. To complete the decoupling, Smilansky's characterization of compatibilism as involving injustice needs to be disconnected from his claims about compatibilist justification for moral responsibility being shallow. Additionally, it must be shown that an incompatibilist perspective is not required for making the relevant injustice compelling. I turn to these tasks in the next two sections.

¹⁷ There are other philosophers with this sort of approach. Daniel Dennett sloganeered on its behalf: "Don't try to use metaphysics to ground ethics . . . put it the other way around: use ethics to ground what we should mean by our 'metaphysical' criterion." (Dennett, 2003, pg. 297) Peter van Inwagen is another example. He thinks that moral responsibility is sufficiently valuable that interest in preserving it should govern the stance one takes to the conceptual issues: "To deny the free-will thesis is to deny the existence of moral responsibility, which would be absurd. . . . It is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think, should we become compatibilists." (van Inwagen, 1983, pg. 223)

Smilansky on Compatibilism's Shallowness

Smilansky characterizes compatibilism as involving serious trade-offs. The chief virtue of compatibilism is that it enables the appreciation of agency, which makes all sorts of valuable social states of affairs possible.¹⁸ Smilansky's criticism of compatibilism is supported by an argument that he calls "the trap." Roughly, Smilansky takes compatibilism to make it harder to see "the depths of the injustice and tragedy involved [in the compatibilist life of responsibility]."¹⁹ Both of these ideas deserve elaboration.

In the following passage, Smilansky addresses the extent to which moral responsibility practices can be justified by the non-libertarian forms of agency endorsed by compatibilists:

Any justification for the manifold social practices which discriminate and then harm (or benefit) some people, on the basis of their behavior, is also shallow. Their shallowness does not make them unimportant: compatibilist forms of life are the "best game in town", and the Appreciation of Agency, as I explicated it, is crucial for human dignity and a decent personal and social life.²⁰

This passage makes it clear that Smilansky distinguishes between conceptual and pragmatic forms of justification for moral responsibility practices. Conceptually, he insists that any justification that does not affirm libertarian doctrines is shallow, which I reject for broadly Strawsonian reasons that will be presented in the next section. Here, my interest is to attend to the pragmatic side of things. Smilansky claims that if compatibilism can sustain the appreciation of agency it thereby sustains human dignity and makes decent personal and social lives available

¹⁸ "In maintaining the life of responsibility and the Appreciation of Agency, we also allow some dignity, empower individuals, and permit the establishment of a decent social order, as far as possible." (Smilansky, 2012, pg. 237)

¹⁹ (Smilansky, 2012, pg. 237)

²⁰ (Smilansky, 2012, pg. 235-236)

to persons. His remark about compatibilism being the best game in town suggests that these pragmatic goods are not securable by other means.

I confess that I do not see how being the only route to human dignity and decent personal and social lives could be considered shallow justification for moral responsibility practices. While I see no reason to endorse the exclusivity clause involved here, until an alternative way of securing human dignity et al. is offered, I consider moral responsibility practices quite formidably justified insofar as they provide these. Importantly, this is far from a novel position; many philosophers take moral responsibility practices to be susceptible to this sort of pragmatic justification. In fact, one of the more profuse currents of the moral responsibility literature, the discussion about the reactive attitudes begun by P.F. Strawson,²¹ is founded upon rejection of the view that moral responsibility requires metaphysical justification of any kind.

Even outside the large Strawsonian literature, there are philosophers who explicitly value what moral responsibility practices offer to societies far more than they value the extent to which such practices match up with people's intuitions about the nature of human freedom and the universe in which it resides. Daniel Dennett has no allegiance to either libertarian or Strawsonian principles but is content to rest his case in favour of moral responsibility on what it offers to society: "The only way I can see of securing civilization (which I see as about as "intrinsically"

²¹ (P. F. Strawson, 1962)

good as anything I can think of) is with a system of moral responsibility that includes punishment for those found responsible for their misdeeds.”²²

Thus, although Smilansky takes compatibilism to be regrettably out of phase with libertarian doctrine, he is still alert to the considerable social value of taking individuals to harbor moral responsibility-conferring agency. Setting the libertarianism issue aside, Smilansky shares in the common view that compatibilist forms of moral responsibility can hardly fail to be worth cultivating, given how many aspects of our lives are owed to its provenance. Smilansky calls attention to many such aspects. He takes compatibilist conceptions of agency to allow for a “community of responsibility,” that creates “the context within which respect for persons flourishes.”²³ He also notes that compatibilism comports fairly well with both our “inherent phenomenology” of ourselves as agents,²⁴ and with our sense of personal development.²⁵

In sum then, Smilansky is content to concede that “there is a basic substrata of compatibilist abilities and capacities, through which almost all adults almost all of the time can control their actions, and behave responsibly with respect to them.”²⁶ Still, he insists that there is good reason

²² (Dennett, 2012)

²³ (Smilansky, 2012, pg. 226)

²⁴ “Being a person, an agent, whatever is about to happen that I am planning to do is, from the perspective of choosing, up to me. I am not passive: I do have some measure of control over what I do. At the moment of acting, I have enough control; I can stop behaving nastily.” (Ibid)

²⁵ “As one grows up, one takes charge of the application of one’s capacities through one’s actions, makes them one’s own and identifies with them, and presents him or herself to the world as accountable for his or her choices and behavior. Such gradually maturing agency is inherently concerned with responsibility: one sees oneself and acknowledges to others that one’s choices and actions are a proper basis for being reacted to and treated in accordance with them.” (Smilansky, 2012, pg. 225)

²⁶ (Smilansky, 2012, pg. 226)

“to resist the compatibilist attempt to claim that when these conditions are met, there is nothing else that we need to worry about.”²⁷ I turn, thus, to what else there is to worry about.

Smilansky on Compatibilism's Injustice

Smilansky's criticism of compatibilism takes the form of an argument that he calls 'the trap.'

The Trap is a thought experiment about a man who commits a crime and goes to jail, which ruins his life. Smilansky asks the reader to consider that through the use of some advanced technology it is known well in advance that this person will commit a crime. Smilansky adds that this diagnosis of future wrongdoing is not set in stone, but rather might be avoided if steps are taken on the part of society to keep the individual in question from accumulating the relevant dispositions and/or encountering the relevant inducements to criminal activity. Smilansky notes, of course, that we do not have the technology required to perform this sort of diagnosis in particular cases, but he reminds the reader that we have sufficient sociological data to know that many people will be led into criminality by their circumstances. Such individuals, he describes as 'trapped.'

Smilansky's major move in the trap argument is to induce contrition in his readers regarding trapped individuals. He does this by emphasizing that there are measures society might have taken to prevent the individual from becoming a criminal. He writes, "we established the social institutions and, despite knowing that he would fall were we not to change things, we nevertheless decided not to. The man is trapped, and we are responsible for the parameters that

²⁷ (Smilansky, 2012, pg. 226)

trapped him.”²⁸ Smilansky does not intend the trap story to be an exotic occurrence. He does not see an important difference between the statistical sociological information already at hand that speaks to the eventuality of criminality on the part of people in particular social stations and the imagined prediction that is facilitated by advanced technology. Further, he explicitly states that the trap story is intended as a commentary on the way things go in the real world.²⁹

Because there are measures that could be taken to prevent trapped individuals from becoming criminals, Smilansky claims that we ought to feel complicit in the crimes of trapped people whenever they occur. Importantly, he specifies that we should feel complicit in both the harm that befalls the victims of such crimes, as well as in having been complacent about the conditions that eventuated the criminal behavior. He writes, “What my story is intended to emphasize is the sense in which the person is trapped by our policies: to repeat, we know beforehand what he is destined to do unless we save him but we decide—for our reasons—not to do so. We let him harm his victim, and (compatibilistically justly) fall to his doom.”³⁰

Smilansky’s appeal to what could have been done to preempt the criminal behavior of trapped individuals is curious to me. In the first place, I do not see why it is so important that measures are not taken to preempt criminality. Additionally, I think Smilansky fails to acknowledge the complex difficulties that would be confronted by any persons or institutions intending to take such preemptive measures. This is not to say that I wouldn’t applaud efforts to lessen the

²⁸ (Smilansky, 2012, pg. 219)

²⁹ “This is not a case of unjust provocation or inducement targeting any given individual, but rather an exploration of how an advanced contemporary liberal-democracy, a well ordered society, routinely operates.” (ibid)

³⁰ (Smilansky, 2012, pg. 219)

likelihood of criminality amongst “trapped” individuals, and the concomitant reduction of complacency regarding the fate of such individuals. These would be wonderful developments, but I do not think that the problematic elements of compatibilism are conveyed by noting their absence.

Fortunately, Smilansky’s diagnosis of compatibilism’s costs can survive without appeal to the motif of observers’ complicity. The fundamental problem with compatibilism is that it eventuates injustice. Smilansky writes, “the purpose of this thought experiment is to indicate the limitations of compatibilism, to show why even under ideal compatibilist conditions, when we do right, according to compatibilism, injustice prevails.”³¹ The injustice in question might be made more compelling by noting that compatibilism doesn’t necessarily foster motivation to lessen the degree of injustice in society, but the injustice Smilansky discusses can be articulated without this embroidery.

Smilansky describes the challenge facing compatibilists as “directly ethical.” He writes, “the very people whom the compatibilist describes as free, morally responsible, and deserving of blame and punishment for committing a given criminal act are seen to be, in fact, trapped. They are literally the victims of circumstances.”³² Smilansky’s insistence on adding that such individuals are “trapped by us” betrays a commitment to the idea that injustice ought to be construed as issuing from agents. As mentioned, to the extent that this motif can compel interest in lessening social injustice, I think it is worth emphasizing. Nevertheless, it is possible to look at

³¹ (Smilansky, 2012, pg. 219)

³² (Smilansky, 2012, pg. 220)

what Smilansky describes as injustice and find it worth addressing even in the absence of feelings of complicity.

The core feature of the injustice that Smilansky takes compatibilism to involve appears to be that it fails to take seriously behavioural factors that lie outside of individuals' control. He writes, "even if we grant the compatibilist everything that he or she wants, people will still spend twenty years in prison, when (as non-believers in libertarian free will) we must say that in fact there is no other way in which their lives could have unfolded, and which was within their *control*."³³ It is important to clarify that Smilansky's concept of control derives from libertarian doctrine. It is the sort of control that would neutralize the trap argument. That is, if libertarian free will were actually possessed by persons, this would mean that persons cannot be trapped.³⁴ So Smilansky takes recognition of the absence of libertarian control to be necessary to discerning traps. Perhaps it is, but it is not sufficient.

As naturalism has risen over recent decades, few contemporary philosophers believe that libertarian doctrine can describe reality.³⁵ Most philosophers are compatibilists and they vary

³³ (Smilansky, 2012, pg. 236, italics added)

³⁴ "With libertarian free will, while circumstances are important, they do not determine the outcome. Thus if we as a society have otherwise been reasonable in setting up the moral game, it is the criminal—rather than we—who is morally responsible for his ending up in prison, because of the way he chose to play the game." (Smilansky, 2012, pg. 221)

³⁵ Robert Kane is an interesting hold-out. Kane's theory is fascinating (Kane, 1998). He argues that the human brain is of an order of complexity sufficient for it to be the case that indeterministic quantum phenomena sometimes determine an individual's decision. Kane is persuasive that this could be the case at least at times of dilemma. He argues that persons sometimes need to make choices between two mutually exclusive options that can both be willed by an individual, with quantum phenomena intervening only to tilt the balance toward one of the two choices. Kane insists that this would be genuine could-have-done-otherwise free will. Critics insist that it makes no sense for individuals to be held responsible for the outcomes of quantum events. I do not have clear intuitions as to whether a

greatly in the extent to which their theories recognize moral responsibility practices to involve injustice. Bruce Waller and Daniel Dennett provide an instructive pair. Waller thinks that compatibilist conceptions of free will are tenable but that the unfairness of moral responsibility practices is nevertheless sufficient to call for their abolition. Dennett, on the other hand, endorses a compatibilist form of free will and insists that moral responsibility practices are unfair to a minor and tolerable degree. This shows that giving up libertarianism cannot be equated with recognition of there being problematic injustice within moral responsibility practices.

It is plausible then that Smilansky's reasoning is subject to further commitments that explain why he thinks adopting an incompatibilist perspective is the only way to glimpse what is problematic about moral responsibility practices. I suggest that this commitment be described as *the forensic stance*. By a 'forensic stance' I mean a perspective that aims to maximize its chances to answer the question of whether an action was bound to occur. Smilansky appears to be committed to the idea that the eventuality of an action must be forensically non-controversial in order for moral redress of the action to be unjust. Note that this is not a view about whether human action is deterministic, but rather a normative claim about the role that determinism should play within the process of applying philosophical oversight to moral responsibility practices. For Smilansky, confronting the injustice of moral responsibility practices must be done from a perspective from which the actions of persons are seen as inevitable.

quantum-mechanically determined choice between two endorsed options constitutes contra-causal agency and so I will not pursue the matter any further. See (D. Dennett, 2003) for a charitable critique of Kane's program.

But is it necessary to invoke determinism in order to render the injustice at work in moral responsibility practices? The whole of this thesis will constitute a sustained effort to show that it is not. In this chapter, my goal is just to present Smilansky's reasons for encouraging episodes of incompatibilist thinking to act as a counterweight to the problems that afflict the predominantly compatibilist status quo, and to articulate a plausible alternative. For Smilansky, it appears to be crucial to criticizing compatibilism that trapped individuals are understood to be doomed,³⁶ but Smilansky's trap argument itself provides support for a non-metaphysical characterization of the injustice that prevails in the moral responsibility system.

As mentioned, Smilansky starts out describing the trap with a thought experiment involving advanced technology that makes strong predictions about people's behavior in the future.³⁷ The theoretical viability of this sort of technology appears to require the truth of determinism,³⁸ but Smilansky soon shifts away from the idea of technologically facilitated predictions to consideration of sociological statistics, which do not depend on the truth of determinism. He writes of the trap argument,

This story is unreal only in one important respect: we do not in fact possess the high level of knowledge described here, and hence we are incapable, as a rule, of such prediction on

³⁶ "The hard determinist case which I have just made, through the story of The Trap, captures an important sense of injustice. And this seems to me now very hard to deny. The absence of predictability means that the criminal to be did not know that he was doomed through our socially designed set up, but this ignorance does not change the reality, which is that, in fact, he was doomed." (Smilansky, 2012, pg. 223)

³⁷ "Let us assume, now with far less empirical basis but not in a way that should matter to us for present purposes, that it was known years in advance that he would commit the sort of crime that he committed, and end up in prison. The available technology is such that this could be predicted ... as long as nothing further would be done." (Smilansky, 2012, pg. 218)

³⁸ At the beginning of the section presenting the trap argument, Smilansky asks his readers to assume that the story takes place in a deterministic world: "Assume a world without libertarian free will. In fact, just for the sake of simplifying the story (it does not change things morally), assume that it is a deterministic world, rather than one which combines determinism with random indeterminism, which does not enhance human control. We wish to explore compatibilism, and compatibilists have no particular interest in indeterminism, so should agree to this." (Smilansky, 2012, pg. 217)

the individual level. We do, however, have reason to believe that, statistically, many people are going to fall in this way, unless we change our economic, penal, or policing policies, and we do not change them. Morally, the fact that our knowledge is only statistical does not seem to matter very much.³⁹

If Smilansky is right that there isn't a morally significant difference between the advanced technology-enabled predictions and the inferences from sociological statistics, then why not just invoke the sociology? Plausibly, Smilansky includes the bit about advanced technological predictions because of his allegiance to the forensic stance that he takes to be a virtue of incompatibilism. It can be asked, though, whether it is really required that the crimes of socially disadvantaged people (trapped folks) be even theoretically subject to flawless forecasting in order to make claims about injustice in moral responsibility practices. Some might agree with Smilansky that predictive certainty is required to support the injustice claim because without it, any particular trapped individuals can be construed as having been able to evade their traps. So long as someone enjoys an opportunity, however remote, to resist the influences of their environment and upbringing, she can be ascribed an ability to do otherwise, which can be taken to ground her moral responsibility should her resistive efforts fall short of preventing criminal behaviour.

This is the sort of stance Smilansky thinks compatibilism encourages. Immediately after the passage quoted in footnote 14, which explains how the trap argument is intended to describe contemporary developed societies, Smilansky writes, "It is crucial to see that this practice is something in which compatibilists find no fault."⁴⁰ This claim is not accurate. There are many compatibilists who find plenty of fault in the ways that moral and legal responsibility are meted

³⁹ (Smilansky, 2012, pg. 219)

⁴⁰ (Smilansky, 2012, pg. 219)

out to individuals whose lives made morally condemnable behaviour more likely.⁴¹ The most sophisticated among them is Bruce Waller. It is somewhat misleading to describe Waller as a compatibilist because he is an eliminativist about moral responsibility, but he is very much a compatibilist about free will, which he takes to survive skepticism about libertarianism.⁴² His critique of moral responsibility does not rely on the sort of forensic certainty championed by hard determinism.

Waller's view is that injustice is inevitable within the moral responsibility system, not because people's behavior can theoretically be foretold, but because moral responsibility judgments are systematically insensitive to factors that lie beyond individuals' control. Waller's position does not insist that uncontrollable factors such as individuals' formative environments eliminate their capacity to avoid particular actions. In contrast, Waller finds a way to consider that which individuals do not control without thereby divesting them of all agentic capacity. A full treatment of Waller's view will be presented in chapter three. For now, I appeal to Waller just to show that determinism does not need to be invoked in order to render the injustice that Smilansky (and others) take to afflict the moral responsibility practices of our societies.

It is feasible to be agnostic about whether uncontrollable factors fully determine individuals' behavior and at the same time suggest that moral responsibility judgments should take such factors into account. Smilansky's claim that compatibilists find no fault in moral responsibility

⁴¹ Dennett writes, "of course I also agree with [Waller] and many others that our American system of justice and punishment is obscenely mis-instituted for the most part, but I want to reform it, not abandon it." (D. Dennett, 2012)

⁴² (Waller, 2011, ch. 3)

judgments being insensitive to the uncontrollable factors that contribute to individuals' behavior betrays his residual allegiance to the traditionally dichotomous rubric of compatibilism and incompatibilism. According to such a rubric, free will and moral responsibility are either legitimate despite the contributions of uncontrolled factors, or they are illegitimate because of them. It is true that compatibilists (with the interesting exception of Waller) do not find the insensitivity of moral judgments to uncontrollable factors to be a fault sufficient to call for the elimination of moral responsibility, but this does not entail that compatibilists must refrain from seeing this as a fault. In chapter three, where I present a non-eliminativist form of Waller's critique of moral responsibility, a detailed account of how one can be simultaneously for moral responsibility and lucid about the fundamental injustice it involves will be presented. For now I only wish to articulate this alternative to Smilansky's view as a way of showing that injustice in moral responsibility practices does not require an incompatibilist perspective.

It is important to note though, that a rejection of libertarianism is plausibly helpful and possibly necessary for rendering the injustice of moral responsibility practices. As noted above, Smilansky thinks that if people did have libertarian free will, this would neutralize the trap argument. It would be so because the truth of libertarianism would entail that people cannot be trapped. Libertarianism can be described as a view that sets limits on the extent to which people's actions can be influenced by factors that they do not control. Because it posits that persons always have a capacity to do otherwise, however significant factors external to a person may be, the person herself remains an arbiter of their influence on her conduct. Therefore, if the epistemic ambitions of the forensic stance are abandoned, it is important that libertarian beliefs not be around to deflate the importance of external influences upon a person's conduct.

Libertarianism simply does not contain within itself the means to capture the injustice Smilansky intends to convey with the trap argument. This failure, though, is instructive. The reason that a libertarian perspective is unable to do this is that it posits a faculty of self-control within persons that cannot be outpaced by influences from uncontrolled factors, such as the conditions Smilansky describes as making up traps. It is a virtue of compatibilism that it doesn't foreclose on the possibility of non-agentic factors making significant contributions to actions. To be sure, compatibilists tend to place the emphasis on agentic factors being robust enough for moral responsibility to be legitimate, but the view typically recognizes that circumstances beyond a person's control shape both an individual's agentic efforts as well as the context in which they occur.⁴³

Thus, although it seems important that libertarianism is rejected in order to appreciate the injustice in moral responsibility practices, this does not mean that incompatibilism needs to be embraced. The importance of rejecting libertarianism is just that it frees up space to include in moral responsibility judgments factors that are not subject to an agent's control. Smilansky does not provide any reason to accept that one needs to construe persons as controlled by such factors in order to explain that injustice can result from failing to consider them. His explanation of what is problematic about holding trapped people morally responsible for their behavior, together with his remarks about how libertarianism (if true) would diffuse the trap argument, suggest that the relevant injustice is constituted by moral responsibility judgments failing to do justice to factors

⁴³ Fischer explicitly states this: "It is better to think of compatibilism as conceding from the beginning that we are thoroughly subject to factors entirely outside our control. Nevertheless, according to the compatibilist, we can still exhibit a meaningful and robust sort of control." quoted in (Smilansky 2012, pg. 220)

that cannot be altered by a person's agentic efforts. The only metaphysics required is a resistance of libertarianism, which Smilansky himself concedes to be an illusion.

And rejecting libertarianism ought to be easy. It does not follow from the fact that recognizing the non-existence of libertarian free will is required for confronting injustices within moral responsibility practices, that such practices can only be satisfyingly legitimized by libertarian beliefs. This inference, which appears to animate Smilansky's willusionism, is logically flawed. It can be characterized either as perpetrating the fallacy of denying the antecedent, or of sliding illicitly between conditional and biconditional forms of the claim about libertarian free will granting legitimacy to moral responsibility. There is no non-question begging way to get from the claim that if libertarianism is true, then moral responsibility is legitimate to the claim that the only way for moral responsibility to be legitimate is if libertarianism is true.

Formally, the first idea can be expressed as a conditional: $LFW \rightarrow MR$. In order for the inference about the necessity of libertarianism to work, however, the connection between libertarianism and moral responsibility would have to take the form of a biconditional: $LFW \leftrightarrow MR$. There are two types of reasons to resist the biconditional formulation of the relation between libertarian free will and moral responsibility: conceptual arguments such as those from Frankfurt and Dennett, and empirical work from psychologists and philosophers. I will briefly present these reasons, and then present my criticism of, and alternative to, Smilansky's suggestion of oscillating between compatibilist and incompatibilist perspectives to cope with the deficiencies of each.

Arguments against LFW ↔ MR

The word ‘alternativism’ is often used to describe the view that the libertarian conception of free will is required for moral responsibility. ‘Alternativism’ is so used because an individual who possesses free will is supposed to have the ability to do otherwise, i.e. to choose from multiple alternative courses of action. In *Elbow Room: the varieties of free will worth wanting*,⁴⁴ Daniel Dennett makes a convincing case that intuitive support for alternativism typically derives from reactions to cases of coercion, the import of which for broader discussions about moral responsibility in general is highly questionable.⁴⁵ For example, when an individual is forced to behave in a certain way by another person, say via a gun to the head, most people’s intuition will be that the coerced individual should not be held responsible for behaviours perpetrated while the gun is at her head. These sorts of observations do lend some credibility to the idea that the availability of alternatives has some connection to moral responsibility, but it falls short of establishing the close connection conveyed by the biconditional formulation above.

The nature of the connection between alternative possibilities and moral responsibility was effectively scrutinized and found wanting by both Harry Frankfurt and Daniel Dennett. In ‘Alternate Possibilities and Moral Responsibility,’⁴⁶ Frankfurt criticizes what he calls ‘the principle of alternate possibilities’ by providing a counterexample to the doctrine. As stated, the doctrine holds that individuals should only be held responsible for actions about which it can be

⁴⁴ (D. Dennett, 1984a)

⁴⁵ (Dennett, 1984a, ch. 1)

⁴⁶ (Frankfurt, 1969)

said that they could have done otherwise. The example involves a man named Jones who has decided on some action. Meanwhile another man named Black is watching Jones and planning to intervene coercively if Jones begins to deviate from the course of action. Black's interest is to ensure that Jones performs the action already decided upon and the reader is to suppose Black to be perfectly effective at both discerning any deviancy on Jones' part regarding the performance of the action and at manipulating Jones so that he gets back on course. As it happens, Jones does not drift from his decision and performs the action without Black having to intervene. Thus, Jones' action is one about which it cannot be said that he could have done otherwise, but because Black was inactive, leaving Jones' decision to unfold as it normally would, Jones is intuitively morally responsible for his action.

While Frankfurt's argumentation has been challenged and endorsed many times over,⁴⁷ the least that can be said of his argument is that it has shifted the burden of proof onto those who maintain that an ability to do otherwise is crucial for moral responsibility. To my judgment, this burden has not been met. However, my own judgment is unimportant because Dennett's case against alternativism is more difficult to quibble with on account of the way it appeals to facts about the way the world is, as opposed to Frankfurt's reliance on fairly exotic examples. In 'I could not have done otherwise: So What?'⁴⁸ Dennett advances what I call the *uniqueness lemma*.

The uniqueness lemma is a very general claim about situations. As Dennett construes it, alternativism is committed to the claim that individuals should only be held responsible if they

⁴⁷ (Copp, 1997) is sceptical, for example, while (Fischer, 2012) largely accepts and builds on Frankfurt's work.

⁴⁸ (D. Dennett, 1984b)

could have done otherwise in the exact same circumstances. Dennett takes the prospects for such a principle to be very dim for the reason that the question of whether someone could have done otherwise in a particular situation is fundamentally empirically inaccessible. The reason for this is that reality is not a tape that can be rewound to a certain point in time and investigated. After an action, the situation in which the action occurred is lost forever in two ways. First, Dennett contends that the perpetrator of an action can never be returned to the exact same state in which he performed the action.⁴⁹ Second, the most viable conception of the world's physics involves quantum indeterminacy, which Dennett takes to entail that any particular situation is freighted with unique quantum phenomena which as a matter of physical principle cannot be expected to recur.⁵⁰

Thus, Dennett can be taken to add to Frankfurt's conceptual salvo, a more naturalistic means of scepticism regarding alternativism. The circumstances of an action, those under which an ability to do otherwise is thought to be important, are always unique, including the states of the individual that acts. The uniqueness lemma doesn't refute alternativism, but drains it of any practical significance. Dennett writes, "ordinary people, when interested in assigning responsibility, do not in fact pursue inquiries into whether their fellows could have done

⁴⁹ "People ... are also designed to be so sensitive to the passing show that they never can be in the same microstate twice. One doesn't even have to descend to the atomic level to establish this. People learn, and remember, and get bored, and shift their attention, and change their interests so incessantly, that it is as good as infinitely unlikely that any person is ever in the same (gross) psychological or cognitive state on two occasions. And this would be true even if we engineered the surrounding environment to be "utterly the same" on different occasions-if only because the second time around the agent would no doubt think something that went un-thought the first time, like "Oh my, this all seems so utterly familiar; now what did I do last time?" (D. Dennett, 1984b, pg. 559)

⁵⁰ "If our responsibility really did hinge, as this major philosophical tradition insists, on the question of whether we ever could do other- wise than we in fact do in exactly those circumstances, we would be faced with a most peculiar problem of ignorance: it would be un- likely in the extreme, given what now seems to be the case in physics, that anyone would ever know whether anyone has ever been responsible. For today's orthodoxy is that indeterminism reigns at the subatomic level of quantum mechanics." (Ibid, 557)

otherwise. Now we see a reason why they would be unwise to try: the sheer impossibility of conducting any meaningful investigation into the question.”⁵¹

In addition to Dennett’s and Frankfurt’s more philosophical challenges to alternativism, a number of empirical studies call the doctrine into question. Experimental philosophers have performed many investigations into the connection between determinism and moral responsibility.⁵² Although the studies are varied, a striking pattern emerges from their results. Known as the abstract/concrete paradox, subjects of these studies have consistently shown a willingness to accept that determinism vitiates moral responsibility in scenarios typified by abstraction while continuing to hold individuals responsible for even determined behaviour when the actions are described concretely.

Although the results of experimental philosophers’ work on would-be defeaters of moral responsibility remain controversial, particularly with respect to the explanation of the abstract/concrete phenomenon, it is clear that the traditional conception of alternativism fails to track at least some of the factors to which people’s moral judgments respond. Broadly speaking, I take these results to support a Strawsonian view of moral responsibility, according to which metaphysical considerations about ultimate authorship are idle in comparison to consideration of actors’ actual lives when it comes to moral responsibility judgments. At the least, traditional

⁵¹ (Dennett, 1984b, pg. 558)

⁵² See (Nichols & Knobe, 2007), (Sinnott-Armstrong, 2008), (Roskies & Nichols, 2008), (Nahmias and Murray, 2011), (Mandelbaum & Ripley, 2012), and (Knobe, 2014)

incompatibilists who take determinism to be a threat to moral responsibility need to find a way to explain these results, and so far none is compelling.

Smilansky's Oscillatory Coping Strategy vs. The Tectonic Theory of Moral Responsibility

In this section, I will recapitulate and clarify elements of Smilansky's proposal of oscillating between compatibilist and incompatibilist perspectives as a way of mitigating the shortcomings of each. The problems with this proposal are not difficult to diagnose, and Smilansky is aware of them. Once the demerits of the proposal are explained, I will present my tectonic theory of moral responsibility and explain how it can avoid the pitfalls of Smilansky's proposal.

Smilansky's ambivalence is clear. He writes, "Philosophically, we need to abandon the Assumption of Monism on the compatibility question and try to work out, in detail, where compatibilism and where hard determinism are more personally and morally salient, and how to integrate them."⁵³ He does not explain how to do this, however. Smilansky only goes as far as claiming that philosophers need to "combine and balance the compatibilist and hard determinist positions."⁵⁴ His proposal for achieving this is vague and unsatisfying, but it also represents a departure from his willusionism. Smilansky writes that "Perhaps we need to think also how to take tentative steps outside of the Bubble [that keeps the illusion of libertarian belief active], and against the pragmatic conservatism that, in general, I am advocating in our current situation."⁵⁵

⁵³ (Smilansky, 2012, pg. 238)

⁵⁴ (ibid)

⁵⁵ (ibid)

There are two main problems with this proposal, and they have already been broached. The first is the idea that a compatibilist perspective according to which persons exert meaningful agentic control over themselves and can be held morally responsible for large portions of their behavior must be abandoned in order for the injustice of moral responsibility practices to be appreciated. This was rejected along with the forensic stance that motivated it. It may be intuitive for Smilansky that if one can doubt that an action was inevitable, it is thereby enshrined as deserving of moral responsibility, but there is no necessary connection between the two. As his invocation of statistical sociological data indicates, it is entirely possible to not know that an action was inevitable and still insist that non-agentic aspects of the actor's life should mitigate her moral responsibility. And as Dennett argued, whether an action was inevitable within a particular situation is not an answerable question.

I did grant that it is probably important that libertarianism be rejected for accessing the injustice endemic to moral responsibility systems, but I also showed that there is ample reason to do this anyway, both from the conceptual work of Frankfurt and Dennett, and the empirical work that gave rise to the abstract/concrete paradox. These considerations, however, are only important for demotivating Smilansky's willusionism.

With respect to the proposed oscillation between compatibilism and incompatibilism, there is a significantly greater pragmatic issue that afflicts the incompatibilist side of things. Smilansky's conception of incompatibilism is the view that all human action is destined to occur because of

determinism. The pragmatic problem with this is that it makes trouble for the community of responsibility, thus threatening to drain human lives of social meaning and dignity.

To simply take all persons as wed to particular futures by deterministic forces is intolerable for Smilansky. He describes the view as problematically homogenizing across the class of agents: “Certainly everyone ought to be treated as innocent for no one can, for hard determinists, become guilty, whatever he or she does. Justifying differentiating reactions and policies would seem to be inherently problematic for, we recall, no one deserves to be treated worse than another.”⁵⁶ Thus, according to Smilansky, both compatibilist and incompatibilist perspectives involve deficits of detail regarding persons and their actions. A compatibilist, claims Smilansky, must be complacent regarding uncontrolled factors that contribute to persons’ actions, and an incompatibilist cannot do justice to that which a person does control.

I disputed the idea that a compatibilist must be complacent about uncontrolled factors, but this required invoking the unusual example of Waller’s views. There is something partly right about the charge. Compatibilist views do not typically emphasize the contributions uncontrolled factors make to individuals actions. However, this asymmetry can be explained by the over-arching motif of the moral responsibility literature. Philosophers tend to strive for maximally general claims, and within the moral responsibility literature, authors typically aim to address the legitimacy of moral responsibility judgments all at once, as though all such judgments are of a

⁵⁶ (Smilansky, 2012, pg. 227)

piece. I think that this is pernicious with respect to the building of theories of moral responsibility.

The relative insensitivity of compatibilism to uncontrolled factors is a result of the way that incompatibilists present the presence of such factors to comprehensively vitiate moral responsibility. Given that compatibilists are mostly in the business of rebutting incompatibilist views, it is understandable that they de-emphasize that which persons do not control. The for-all-the-marbles state of the literature results in compatibilists understandably prioritizing emphasis on what people can control as a way of denying that an absence of libertarian free will is sufficient to abandon moral responsibility. In effect, the compatibilist emphasis on agency serves to indicate that moral responsibility is not as unjust as the incompatibilists would have it. Even though we do not have libertarian free will, claim compatibilists, persons have a significant degree of agency, insensitivity to which is more pragmatically problematic than is compatibilism's relative insensitivity to that which persons do not control.

If it were necessary to choose between compatibilism and incompatibilism as described above, pragmatic considerations would appear to favour compatibilism. But this represents a false choice between two views as to how much significance should be assigned to uncontrolled factors. Uncontrolled factors need not be judged to be either overridingly significant (as incompatibilists have it) or insignificant in comparison to what a person can control (as compatibilists retort). The prospect of a philosophically decisive account of the legitimacy of moral responsibility seems to have resulted in a sub-disciplinary norm according to which

ascribing significance to that which persons are unable to control is equivalent to denying the importance of that which they can control.

But the rubric just described only makes sense if we are concerned to (de)legitimize moral responsibility judgments as a single unified phenomenon, and this is not my concern. In fact, I do not think the motif of legitimacy is apt at all in this context. My view is that moral responsibility judgments are inevitable and valuable within human societies, and that they are intrinsically unfair. This is akin to a compatibilist view insofar as I claim that moral responsibility practices cannot be undermined by considerations of determinism, but I depart from compatibilism in insisting that the robustness of moral responsibility practices is not a function of their fairness. If compatibilism were just a claim about whether moral responsibility is here to stay, then I would be pleased to identify as a compatibilist. However, if compatibilism is the claim that moral responsibility is justified because it is fair, I reject it. Moral responsibility is justified despite being unfair.

There is no contradiction in construing moral responsibility to be both valuable and intrinsically flawed. There is however a valuable tension. If the moral responsibility system can be expected to endure and it is known that it will inevitably give rise to unfairness, the task of the theorizer becomes one of minimizing the unfairness involved without undermining the system as a whole. These intertwined priorities are the motivation for my tectonic theory of moral responsibility.

My tectonic theory of moral responsibility is designed to sensitize moral responsibility judgments to uncontrolled behavioural factors, which I call a person's patiency, without thereby desensitizing them to what an individual does control, thus preserving an appreciation of a persons' agency. I describe my theory of moral responsibility as *tectonic* because it describes two layers of characteristics of a person – agentic and patientic characteristics. The geological metaphor is intended to conjure images akin to the interaction of tectonic plates in the Earth's mantle: a dyadic system of agentic and patientic characteristics sliding along and influencing one another in subtle and sometimes unpredictable ways. Like two tectonic plates, agency and patiency provide enduring context to one another as well as the occasional episode of acute influence.

The tectonic theory prescribes a two-step process for determining moral responsibility. First, whether a candidate for moral responsibility is indeed an agent should be ascertained. After an individual is deemed an agent, she should be looked at as a patient to give factors beyond her control an opportunity to mitigate or even vitiate her moral responsibility for a given action. I will argue that if this formula were to be operationalized, it is plausible that moral responsibility judgments could become less unfair.

My purpose, then, is to apply the methods philosophers have used to develop extant forms of compatibilism based on psychological conceptions of agency to the uncontrolled factors that have almost exclusively been valued through the lens of determinism-invoking incompatibilism. Compatibilists have focused on different ways that an individual's mental life can be connected to action so as to justify moral responsibility. In contrast, I will focus on aspects of individual's

lives to which a person's mental life cannot connect in the same ways, thereby locating aspects of persons and their behaviour for which persons should not be held morally responsible, and which may make significant contribution to their actions. Such phenomena comprise a person's *patiency*.

An emphasis on the deterministic nature of persons was a suitable counterweight to claims about the existence of traditional forms of free will, but since compatibilists are now more concerned with the psychological details of agency, something else about persons needs to be emphasized to effectively criticize moral responsibility. The foil for contemporary compatibilist accounts of agency should be patiency. Developing the concept of patiency into a form in which it can figure into moral responsibility judgments and explaining why it can contribute to such judgments are the burdens of the next two chapters.

Chapter Two: What is Patience?

“A balanced philosophical account of the person might start by acknowledging that in much if not most of life we do not choose, and that, indeed, the application of choice to some aspects of life makes no sense.”

(Soran Reader, *The Other Side of Agency*)⁵⁷

The concept of patience traces back to Aristotle’s account of action. Aristotle reasoned that every action involves an agent, who perpetrates the action, as well as a patient, who receives or undergoes the action. Since, philosophers have yet to find much use for the concept of patience. There are a few authors working on the philosophy of action who have used the concept, such as Fred Dretske, Richard Taylor, and Mikael Karlsson.⁵⁸ None of these works, however, addresses the significance of patience for morality.⁵⁹ Dretske and Karlsson both work toward concepts of action that can encompass not just intentional human action, but the doings of plants, animals and even machines. Taylor’s paper ‘Agent & Patient: Is there a Distinction?’ presents a deflationary analysis of action theory on the grounds that the criteria according to which agents are distinguished from patients are entirely pragmatic, in the sense that they cannot be taken to track metaphysical truths of any kind.

Because the tectonic theory of moral responsibility aims to integrate both the agentic and patient aspects of a person, this result from Taylor is welcome for my purposes. My proposed remedy for the unfairness of moral responsibility judgments is pitched at the level of what it is to be a person. I think that a small adjustment here can have a big impact on the larger system at

⁵⁷ (Reader, 2007b)

⁵⁸ (Karlsson, 2002; Taylor, 1982; Dretske, 1991)

⁵⁹ Although it does not work with the concept of patience, Andrew Sneddon’s book *Actions and Responsibility* (Sneddon, 2006) does connect the conceptual concerns of philosophy of action to considerations of moral responsibility.

work. In chapter four, I will argue that moral responsibility judgments are inherently prone to leaving important facts about persons out of account. More specifically, I argue that the moral responsibility judgment's habitual concern for a person's agency tends to direct focus away from patientic facts about a person, facts which might mitigate or even in some cases eliminate moral responsibility judgments once recovered.

In order to make this case, I should begin with an exploration of the concept of patiency, which is the purpose of this chapter. I will begin by developing a definition of patiency, drawing mainly from the work of the late Soran Reader, who is the only philosopher to address patiency as a subject in itself. Reader argues that any conception of the person that does not account for patiency fails to capture a number of important aspects of human lives. Because Reader's concern is to rehabilitate the concept of the person, linking her work to the tectonic theory will require clarifying the role that control, and more importantly, a lack thereof, plays in her discussion. To do this, I will present Daniel Dennett's analysis of the concept of control from *Elbow Room: The Varieties of Free Will Worth Wanting*.⁶⁰ At the end of the chapter, I will present a tentative taxonomy of different types of patiency.

Defining Patiency

The late Soran Reader's paper *The Other Side of Agency*⁶¹ is the most important philosophical work on the surprisingly neglected subject of patiency. Reader does not elaborate on the classical

⁶⁰ (D. Dennett, 1984a)

⁶¹ (Reader, 2007b)

definition of patiency. She writes, “I use the word 'patient' as the correlate of 'agent.' A being is a 'patient' in this sense when it is acted on.”⁶² I have a minor worry with this formulation, concerning the phrase ‘acted on.’ From such a description, it might be thought that an individual’s patiency only ranges over the results of other people’s actions, excluding the influence of impersonal factors such as genetic endowment or cultural environment. For many, it might not be natural to think of an individual’s genes or culture as *acting on her*. As will soon become clear, Reader does not constrain the concept of patiency in this way, and I therefore modify the above formulation as follows: an individual’s patiency refers to the myriad circumstances and characteristics incurred over the course of a life as a result of factors that are not susceptible to significant modification through the exertion of her agency.⁶³

Rather than a correlate of agency, then, I propose that patiency be thought of as the complement of agency. If a person’s agency is comprised of their control over facets of their lives and selves, her patiency is comprised of everything else that she cannot control. Of course, this definition is broader than is necessary for the tectonic theory, whose concern for patiency is keyed to uncontrollable factors relevant to the performance of a particular action. I do not think that this is problematic, however, because the concept of agency is also broader than is required for moral responsibility judgments. People have control over many more aspects of themselves and their lives than can be usefully considered in connection to particular cases of moral responsibility. Many people, for example, are multi-lingual and thus have a capacity to control the language in

⁶² (Reader, 2007b, pg. 581)

⁶³ It is important to note that an individual’s patiency may be partially comprised of circumstances and characteristics that are not wholly due to external factors. I have in mind external factors that have been actively consented to. Still, once an individual accepts a role that she might have refused (say, to supervise a dissertation), pertinent circumstances accrue to her patiency insofar as they cannot plausibly be refused once the relevant role has been accepted and can thereafter be uncontroversially classified as flowing from non-agentic factors.

which they express themselves, but it is difficult to see how this feature could be relevant for moral responsibility judgments.⁶⁴

This brings up the matter of how to determine if a particular patientic factor is relevant to a moral responsibility judgment. I would like to be minimally prescriptive on this issue. The tectonic theory prescribes that moral responsibility judgments consider a person as a patient, but deliberately leaves room for disagreement about which patientic aspects of a person or her life are relevant to the moral responsibility judgment at hand, as well as the extent to which particular patientic aspects of a person should mitigate her moral responsibility. In other words, the tectonic theory aims to provide a conceptual apparatus that explains what it is that people disagree about when they disagree about moral responsibility, rather than aiming to eliminate such disagreement.

The basic ideas do need to be defended, however. The tectonic theory prescribes that if some aspect of a person's life was not subject to her control and played a role in bringing about the action for which her moral responsibility is being determined, then it should be considered as a possibly mitigating factor for moral responsibility. There are two ideas here that need unpacking: what it is to have control over something, and what it is for something to contribute to bringing about an action.

⁶⁴ Of course, it would be possible to construct a case in which speaking a second language was relevant to a person's moral responsibility for an action, but my point is just that many features of persons' agency are frequently irrelevant to moral responsibility claims.

Daniel Dennett's analysis of control can serve to clarify the concept of patiency. Dennett appeals to cybernetics and automata theory, wherein the "root idea of control" has acquired some technical clarity. Dennett writes, "A controls B if and only if the relation between A and B is such that A can drive B into whichever of B's normal range of states A wants B to be in."⁶⁵ This definition has implications regarding what it is to be a controller as well as what is required for something to be a controllee. A controllee must have a nature that makes it possible for it to be in different states. A controller, on the other hand, must be capable of harboring desires regarding which of the possible states its controllee ought to be in and also be capable of satisfying those desires.

Thus, being a controller has three major components. There is an epistemic element: to be in control of something, one must know the different states it can be in, and how to bring them about. There is also a psychological element: to be in control of something, one must have desires regarding which of its possible states one would like it be in. Finally, there is a causal requirement: to be in control of something, one must have and seize an opportunity to cause it to be in the desired state. While Dennett's concern in developing this definition was to show that free will skeptics who claim that persons are controlled by the laws of nature are working with an underanalyzed concept, it serves equally well in supporting a definition of patiency. A person is a patient with respect to any phenomena that either cannot transpire in different ways (cannot

⁶⁵ (D. Dennett, 1984a, pg. 52); note that Dennett's formulation here describes the necessary conditions for being in control of something. Simply possessing knowledge about a things possible states as well as how to put it into these states does not yet constitute control. The control must still be exercised. This point is made by James Stacey Taylor in his book *Practical Autonomy and Bioethics* (J. S. Taylor, 2009). I fill this gap with the 'causal component' of control described below.

be a controllee) or any phenomena that cannot be adjusted according to the person's desires (is not subject to the relevant controller).

The notion of a phenomenon contributing to bringing about an action is more unruly. A natural way to approach this issue is through counter-factual analysis: if an action would not have occurred if it weren't for the fact that phenomenon X transpired, then X contributed to bringing the action about. This strategy only goes so far, however, because it fails to distinguish between enduring background conditions that contribute to the occurrence of all human actions, and patientic aspects of particular persons' lives that play a more direct role in bringing about their actions. I do not think that is a problem for the tectonic theory, though.

The principled interest in focusing on distinctive features of a person and her action is already a feature of moral responsibility judgments even if such judgments do not involve considerations of patientcy. In other words, while it is important that the patientic side of tectonic moral responsibility judgments is steered toward distinctive features of the judgee, I believe that this will mostly take care of itself because moral responsibility judgments begin as particularized assessments of persons' actions. A moral responsibility judgment is intrinsically interested in what it is about the judged person that led to the performance of a particular action. The mere fact that a person is the subject of a moral responsibility judgment thus already sets her apart from others. The inclusion of patientic considerations as a second phase of such judgments can be expected to inherit the larger judgment's concern for what is peculiar to the subject of the judgment. I take this to neutralize the worry about a counterfactual account of a patientic factor

being relevant to an action. There is no reason to expect that a tectonic moral judgment would suffer from an undue focus on background conditions.

Thus, counter-factual claims, like “if she had not grown up in a severely racist culture, she would not have become a racist,” can be used to enunciate the patientic considerations within tectonic moral responsibility judgments. I would like to emphasize that such claims be construed as a natural way to describe subjective judgments, which should be understood to be defeasible. Put otherwise, counter-factual claims asserting the relevance of patientic factors to a moral responsibility judgment at hand are fine for rendering particular views about how the factor affects the person’s degree of moral responsibility for an action, but one should be careful about insisting on the truth of such claims.

In the previous chapter, I argued that certainty regarding the inevitability of a particular action is not required in order for moral responsibility judgments to be sensitive to pertinent patientic factors. Later, in chapter four, I will explain why tectonic moral responsibility judgments should be self-consciously uncertain with respect to whether a person’s action was made inevitable by patientic factors. For now, with a definition of patientic and a sketch of how it can figure in moral responsibility judgments in hand, I turn back to Reader’s work to both clinch the case that a person’s patientic is always there to be considered, as well as illuminate some of the reasons why it is worth considering.

For the most part, Reader works toward showing that an agentic conception of the person entails a number of patientic phenomena. That the patientic phenomena typically obscured by agency-centric thinking about persons can be morally significant is a further point that Reader asserts, but does not systematically explain. This is not intended as a criticism, however. The patientic phenomena she delineates are for the most part obviously morally significant or obviously not morally significant, and her paper is no less compelling for failing to analyze why this is so.⁶⁶ Since my focus is on moral responsibility, however, it is important for me to establish that the patientic phenomena that Reader shows to be entailed by agentic aspects of a person involve a lack of control.

Reader's case begins with claims about an 'agentic bias.' This is a term for how the concept of a person, whether used by non-specialists, or in philosophers' efforts to delineate a person's essential characteristics tends to focus on the powers possessed by individuals to modify themselves and their circumstances. This obscures and devalues all that individuals do not have the power to change. For Reader the agentic bias accounts for the way that patientic features of an individual are considered negligible if they are considered at all despite that way that they (in Reader's words) "shape the lives of persons so profoundly."⁶⁷

⁶⁶ Additionally, Reader's lone monograph *Needs and Moral Necessity* (Reader, 2007a), which predates the paper under discussion, lays out a conception of ethics as a response to the needs of things that addresses the question of why patientic phenomena are morally significant. Roughly, the claim is that patientic phenomena involve persons' needs, and since all of ethics is conceived by Reader as a response to needs, patientic phenomena are thereby endowed with moral significance. While I find Reader's work on needs to be both compelling and probative, my argument has no allegiance to it.

⁶⁷ (Reader, 2007b, pg. 579)

Reader's first task is to show that philosophers do indeed betray an agentic bias. She writes that this bias is seldom made explicit and instead "presumed, mentioned if at all as a gesture in passing, a presumption, a shared starting point for any reflection on persons."⁶⁸ She offers Charles Taylor's work in Sources of the Self as an exemplification of this tendency.⁶⁹ Reader's second goal is to illuminate the shortcomings of a purely agentic conception of the person. To do this, she explores four features that she takes to be constitutive of such a conception: *action*, *capability*, *freedom* and *independence*. First, Reader presents reasons for skepticism regarding the extent to which these properties inform a stable and satisfying conception of the person, and then she provides patientic counterparts that are entailed by these agentic features.

Reader on Action

Reader's criticism of taking action to be a distinctive feature of personhood is drawn from Richard Taylor.⁷⁰ Late in his career as a philosopher of action, Taylor faced the challenge of legitimizing the intuitive claim that persons are the only things that genuinely act. Taylor considered multiple ways of meeting this challenge and concluded that none were viable. Rather than actions providing a means to demarcate persons from the rest of the world, Taylor concludes that any such demarcation must be a decidedly pragmatic affair. Reader approvingly glosses Taylor's conclusion as follows: "human actions and persons as agents are distinguished not by intrinsic features, but by the ordinary practices we use to decide what is and isn't an action

⁶⁸ (Reader, 2007, pg 579)

⁶⁹ "Charles Taylor describes his project at the beginning of *Sources of the Self* as 'tracing.. .our modern notion of what it is to be a human agent, a person, or a self.' Treating 'agent' as equivalent to 'person' or 'self' here is not meant to introduce any new or difficult idea. It is simply meant to gesture at a common understanding on which Taylor will rely." (Ibid)

⁷⁰ (R. Taylor, 1982)

or an agent when we encounter them."⁷¹ It follows from this that the idea of demarcating persons from non-persons merely on the basis of action is suspect.

Even if we were to somehow stabilize the basis on which the actions of agents are distinguished from the doings of non-agents, Reader would claim that the resultant account of personhood must still be impoverished. There are two reasons for this. First, even perpetrations to which ascriptions of 'action' are non-controversial will inevitably involve patientic phenomena. To sloganeer on Reader's behalf: to act, one must be acted upon. For example: "When I hit you, I suffer your resistance to the blow ... If I do not suffer that resistance ... I have not done that action."⁷² The second reason is that an agent's action will commonly involve patients who are the recipients of the action.

The moral significance of the first idea, that patientic experience is an inevitable alloy of agentic exertions, can be made clear by contemplating soldiers suffering from post-traumatic-stress disorder. I hope it is not controversial to claim that there are at least some behaviors, the performance of which by someone suffering from PTSD would be subject to less or no moral responsibility while the same behavior performed by a psychologically healthy person would be subject to a harsher judgment.⁷³ If this is right, it shows how even when a person has a normal degree of self-control and exerts her agency to both approve of and execute an action, the very

⁷¹ (Reader, 2007, pg. 584)

⁷² (Ibid, pg. 589)

⁷³ Imagine a sufferer of PTSD who, during a date, is jolted back to troubling memories by some content in a film, and dramatically terminates the date. Such a person is likely to be spared some of the harshness that would be meted out in a judgment of a psychologically healthy person who dramatically ends a date on account of simply disliking the film.

performance of that action can result in incurring states and traits that cannot be controlled, and which therefore have significance for moral responsibility judgments.

Reader's second point about patiency being partially comprised of being on the receiving end of others' actions might seem not to need any emphasis. That actions typically yield either victims or beneficiaries appears to be a point already assimilated into ordinary as well as philosophical discourse about personhood and moral responsibility. However, Reader argues that the role that considerations of victims/beneficiaries plays in thought about persons and morality is problematic, and plausibly a result of the agentic bias. In a self-described 'slimmed down' version of 'The Other Side of Agency,' called 'Agency, Patiency, and Personhood',⁷⁴ Reader provides a good summary of what it is about philosophers' treatment of victims that is problematic. She writes,

There is an aspect of the non-agential human being which has particular moral and political importance, and to which philosophers have generally been sensitive. Human action that aims to harm is violence, and persons suffering such harm are victims. If philosophical efforts in ethics and politics are agentially focussed on the violence done by the perpetrator, or on what bystanders should do to punish or prevent it, the victim will typically be objectified, and our concern will be what can be done to restore that being to full personhood or agency. This implicit denial of full personhood to victims compounds the harm they suffer: our communities shun them, pressing them to hide what they suffer and to 'recover' their humanity by 'getting past' their victimhood, to become 'survivor' agents.⁷⁵

To theorize while in the grips of the agentic bias, thus, predisposes accounts of victim-like patiency toward neglecting the actual substance of individuals' victimhood in favor of emphasizing the ways in which a person (or other people) might act in opposition to the forces responsible for her victimization. I will say more about this in chapter four, wherein I present a

⁷⁴ (Reader, 2010)

⁷⁵ (Reader, 2010, pg. 202)

modified form of Waller's critique of moral responsibility. For now it is sufficient to note that even when theories recognize patiency, typically in the form of victimhood, it is still relegated to a secondary status, providing the background conditions against which an individual's presumed agency functions, instead of being recognized as an important phenomenon in itself.⁷⁶

Reader on Capability

Reader is keenly aware that few theories of personhood enshrine action as the only necessary feature of personhood. More often, an individual's capacities or abilities are given a central role. This allows for personhood to countenance what a person is theoretically able to do, instead of just treating a person as the sum of what they in fact end up doing. Reader mentions John Locke and John Rawls as examples of philosophers that have made ability central to personhood, but targets the more contemporary views of Amartya Sen and Martha Nussbaum in order to show what is problematic about this sort of theorizing. She writes that, "thinkers like Sen and Nussbaum seem to want to conceal human vulnerability to helplessness and dependence behind brave talk of capability. I think the agential conception of persons is what explains this."⁷⁷

These remarks hint at the problems with taking capabilities to be constitutive of personhood. Capabilities are only significant and thus only need positing when they are unexerted. Exerted capabilities are actions and once an action has been perpetrated, there is no longer any reason to inquire after an individual's capability to perform that action. Thus the danger of conceiving of

⁷⁶ There is one way in which victimhood has been frequently connected to moral responsibility: manipulation cases. These will be discussed later in this chapter where I develop a tentative taxonomy of different types of patiency.

⁷⁷ (Reader, 2007b, pg. 600)

persons as possessed of capabilities is that it can easily lead to a failure to do justice to factors that contribute to a capability going unexercised.

There is a disturbing scene from the film 'The Believer'⁷⁸ that exemplifies this well. In the film, Ryan Gosling plays a neo-nazi who is required to undergo sensitivity training with holocaust survivors after he is convicted of vandalizing a synagogue. After listening to one of the survivors' story about a nazi soldier callously murdering his son in front of him, Gosling's character asks dismissively 'So what did you do?,' and proceeds to list a number of behaviours of which the holocaust survivor was presumably capable of at the time such as attempting to gouge out the nazi soldiers eyes. This is a clear incidence of abuse of the agentic conception of persons as defined by capabilities. If it is true at all that the man whose son was murdered by the nazi soldier was capable of acting in opposition to his plight, it is trivially true and easily dwarfed by considerations of all the uncontrollable factors that surrounded him. Once such factors are accounted for, it is clear that it is more sensible to consider the man whose son was murdered in his patiency, rather than stressing the minute quantum of agency he bore within the situation.

Reader presents a similar thought in a moving discussion of the fight or flight dichotomy. In a circumstance of conflict, it is often said that a person will either take a stand against that which conspires against her wishes (fight), or extricate herself from the situation (flight). Reader notes that this is woefully inadequate. Although most people are capable of fighting or fleeing most

⁷⁸ (Bean, 2001)

situations, there is no reason that emphasis on these options should diminish the importance of individuals who do neither. Reader writes,

The idea that either fight or flight must always be possible is one of the fantasies of the agential conception ... [Sometimes] Endurance is the only option. It is not action, it does not show positive capability, it is not chosen or independent. Nevertheless, it is a way to be a person in adversity. Far from being an easy or self-deluded option, endurance is difficult and courageous.⁷⁹

I think that these reflections are sufficient to show that a focus on individuals' capabilities can result in failing to do justice to patientic factors that either explain why capabilities were not exerted in particular situations, or which led to an absence of capability in a particular situation. Additionally, capacity-centric assessments of individuals' behavior can result in debate as to whether a capability was present and unexerted or absent. One can always ask of a person what she could do or might have done to improve her lot, and the query can always terminate in a plausible speculation as to the presence of an unexerted capacity. The fact that this line of thought is more or less always available does not, of course, mean that it is always useful or fair to follow it.

It is worth noting that emphasizing a person's capacity can have positive aspirational value in a wide variety of cases. If a student is struggling, it will likely be more helpful to emphasize ways in which the student can act so as to improve her work than it would be to investigate the relatively inalterable facets of the student and her situation. But Reader's work provides the means to stress the defeasibility of this principle. Regardless of the extent to which a person is capable of particular actions, focusing just on this capacity can lead to a problematic indifference

⁷⁹ (Reader, 2007b, pg. 597)

to uncontrolled factors that explain why a given capacity does not develop or goes unexercised in a particular situation.

Reader on Freedom

Reader argues that once “free” acts are placed in their terrestrial context, “to be free in one respect is to be constrained in another”⁸⁰ as when in a free conversation, participants remain bound by linguistic grammar. This is an example of a patientic phenomenon that is obviously not morally significant, though it still speaks to the way that being unconstrained (i.e. agentic) in any significant way involves being constrained in others. Philosophers’ conception of freedom is typically more complex, though, and Reader’s more probative remarks on freedom target the more sophisticated notions of freedom as constituted by rationality and independence.

Addressing the idea that freedom consists of rationality, Reader notes that “the process of deliberation is always and necessarily incomplete”⁸¹ because the apprehending of one reason must involve failing to apprehend others. Reader sees philosophers’ as prone to abstracting rationality away from the contexts in which it occurs. Her point is that rational deliberation is inevitably subject to constraints deriving from patientic features of persons and their circumstances. This idea, that a person’s capacity for rational deliberation is subject to patientic limitations, is a favourite theme of Waller’s and will be dealt with in detail in chapter four. In regards to the putative independence of a free person, Reader’s account of patienty is clearly a collection of uncontrolled behavioural factors. She writes “If we insist ... that there is a sense in

⁸⁰ (Reader, 2007b, pg. 597)

⁸¹ (Ibid)

which persons are independent, we will be forced to acknowledge that such 'independence' is only ever partial, and is woven, when it is, from multiple dependencies."⁸² Reader eloquently notates some of the dependencies on which human 'independence' rests (healthy body, light, air, freedom from illness, that the earth be turning, family and social support etc.). Whether one's body is healthy, or one's formative environment is providential are phenomena that proceed independently of a person's agency.

The notion that freedom consists in a capacity to act according to independent rational reflections is all but ubiquitous in the compatibilist literature. While not as pernicious as the homogenization that incompatibilism perpetrates across the class of persons, if a theory of moral responsibility simply posits that rationality is criterial for morally responsible action, this too results in problematic homogenization. It is a lesser homogenization, across the class of agents, but still represents an abstraction away from the myriad factors that can limit the degree and effectiveness of rational reflection of which a person is capable in a particular situation. If the problem with incompatibilism is that it exfoliates persons' rational capacities and efforts out of the moral responsibility calculus, then the problem with compatibilism is that it scrubs away other sorts of significant detail about inalterable features of the context of cases of rational action. This is a lesser evil to be sure, but one which merits attention all the same.

The claim that moral responsibility judgments inevitably and problematically homogenize across the class of agents will be argued for in detail in the next chapter. Here, my purpose has just been to present Reader's work to illuminate the impoverishment of purely agentic conceptions of the person. Reader presents many reasons that people can and should be thought of as patients

⁸² (Reader 2007b, pg. 591)

throughout their lives regardless of fluctuations in their degree of agency. Her main contribution to the development of my tectonic theory of moral responsibility is the idea that patiency comprises a coherent and important aspect of a person, and that patientic features are necessitated whenever an individual displays agentic features. Though she does not explicitly claim that agency-focused accounting of individuals' characteristics and circumstances are prone to homogenizing across the class of persons, it is clear that she thinks a concern for patiency allows for analysis of differences between people at higher resolution.

The basic idea behind Reader's work, that a concern for patiency can serve to recover significant details about persons and actions that get obscured when we are only concerned with persons' agency, can be explained in quite ordinary terms. Although there are compatibilists who acknowledge that persons' agency comes in degrees,⁸³ for the most part philosophers who support moral responsibility treat agency as a threshold phenomenon.⁸⁴ This means that some minimum amount of rational capacity is required for a person to be morally responsible, and so long as this threshold is met, further facts about a person and her life do not need to be included in the moral responsibility calculus.⁸⁵ Perhaps they do not need to be included, but if Smilansky and Waller are right about the intrinsic unfairness of moral responsibility practices, then they ought to be. Reader's reflections on patiency are a good ice breaker in this context. She explains both that persons are always patients, i.e. subject to factors that they cannot control, that this is important, and that it is difficult to be mindful of this.

⁸³ (Vargas, 2013a) is a good example.

⁸⁴ (Waller, 2011, pg. 223-232) recapitulates and criticizes what he calls 'the moral responsibility plateau.' This idea will be discussed in detail in chapter four.

⁸⁵ Bernard Williams articulates this well: "To the extent that the institution of blame works coherently, it does so because it attempts less than morality would like it to do... [it] takes the agent together with his character, and does not raise questions about his freedom to have chosen some other character" (Williams, 2006, pg. 194)

Reader's summary of her own work is superb. She writes that the task at hand

is that of exploring what persons suffer as well as what they do, what they cannot do or become as much as what they can, the respects in which they are constrained, the necessities they are subject to, the contingencies that structure their responses to the world, and the ways they are connected to and dependent on other things. To make good progress with this, we need to pay attention to things that do not demand our attention ... We have to resist the temptation to tidy personhood up, to present it like a student on graduation day, all neatly turned out to receive a prize for its achievement. It is fine to pay attention to the agential aspects of personhood ... But it is not fine to give agency all the attention, and to pretend that the non-agential aspects of our life are somehow less human, less valuable, less our own.⁸⁶

Although, Reader is concerned with the concept of the person in general, her reflections are particularly poignant in the context of moral responsibility. The next two chapters will explain why individuals' patiency is particularly acutely and inevitably subjugated to concerns about agency in moral responsibility judgments. Before this is accomplished, however, a basic taxonomy of forms of patiency should be in place.⁸⁷

Types of patiency

As mentioned above, there is one type of patiency that moral philosophers typically address: interpersonal patiency. Interpersonal patiency refers to the ways in which persons are subject to the actions of other persons. Mostly, interpersonal patiency has carried the torch of incompatibilist skepticism about moral responsibility in the form of manipulation cases. Typically, manipulation cases are intended to poke holes in compatibilist theories about what legitimizes moral responsibility judgments.

⁸⁶ (Reader, 2007b, pg. 604)

⁸⁷ I will include a table of different types of patiency at the end of this chapter.

For example, the work of Harry Frankfurt⁸⁸ has inspired many to propose *identificationist* accounts of moral responsibility, according to which a person's moral responsibility for an action depends on whether she identified with the action, i.e. considers the action to have resulted from motivations that she actively endorses. Incompatibilists have responded with thought experiments wherein a person's mental life is manipulated, usually by nefarious neuroscientists, so that the person both performs and identifies with an action of the manipulator's choosing.⁸⁹ For many, the presence of the manipulator mitigates or eliminates the person's moral responsibility for performing the relevant action even though the identificationist criteria for moral responsibility are met in such cases.

The incompatibilist reasoning goes: if a person can fail to be morally responsible for an action with which she identifies, then there must be something wrong or at least missing from identificationist accounts of moral responsibility. The first thing to note about these cases is that determinism does not play a major role. The claim that a person can be theoretically manipulated so comprehensively by another person is not equivalent to the claim that persons are deterministic systems. Incompatibilists tend to see them as straightforwardly related, though.⁹⁰

⁸⁸ (Frankfurt, 1971)

⁸⁹ Noteworthy examples of this include Derk Pereboom's four cases argument (Pereboom, 2013), and Joshua Greene and Jonathan Cohen's Mr. Puppet thought experiment (Greene & Cohen, 2004). The latter will be treated extensively in the final chapter of this thesis.

⁹⁰ Greene and Cohen, for example, write about their story of poor Mr. Puppet (manipulated from birth to commit a crime as an adult) that, "The fact that someone could deliberately harness these forces [the influence of genes and formative environment upon a person's conduct] to reliably design criminals is an indication of the strength of these forces, but the fact that these forces are being guided by other minds rather than simply operating on their own seems irrelevant, so far as Mr Puppet's freedom and responsibility are concerned." (Greene & Cohen, 2004, pg.

I think Dennett's response to such arguments is instructive. He developed the principle of default responsibility to neutralize manipulation cases. He writes, "We invoke what I call the Principle of Default Responsibility: if no other agent is responsible for your condition and the acts that flow from it, you are. The buck stops there, if you are competent."⁹¹ There are two important claims that are implicit in this principle. First is the idea that only morally competent persons can be controllers, in the technical sense of 'controllers' defined above. This pre-empts the common incompatibilist move of claiming that manipulation cases vindicate their position by illuminating the way that persons can be controlled by deterministic forces. In order for a person to be controlled, the controller must have desires regarding how the person is to conduct herself, which cannot be said of the universe at large.

The mitigation of moral responsibility that is intuitive within manipulation cases, thus, is plausibly due to the fact that the manipulated person lacks control over her manipulator, as opposed to the fact that the universe is deterministic. This is the second idea implicit in Dennett's principle: a person's responsibility for an action can be mitigated or even eliminated if its performance was due to another agent. The fact that Dennett has this view even though he rejects the idea that determinism can be exculpatory⁹² illustrates the value of the concept of patiency within moral responsibility judgments. According to the principle of default responsibility, the mitigation at work in manipulation cases is due entirely to interpersonal patiency and not determinism.

1780) Similarly, Pereboom's four cases argument is designed to challenge compatibilists to articulate the relevant difference between person-to-person manipulation and existing as a deterministic system.

⁹¹ (Dennett, 2011, pg. 11)

⁹² "Determinism is a red herring." (Dennett, 2011, pg. 18)

Although Dennett does not state it as such, he concedes that when a person's actions result from uncontrollable factors (in this case the manipulators), the uncontrollable factors should be given an opportunity to mitigate her moral responsibility. In other words, it appears that Dennett would agree that a person is not responsible for that which results from her patiency, even though he limits the purview of patientic considerations to cases of intensive interpersonal patiency. The basic idea is there: a person is not responsible for that which they do not control. To get from here to the tectonic theory of moral responsibility, all that is required is recognition that a person's actions can be more or less a result of uncontrolled factors, and that her moral responsibility can thus be more or less mitigated by patientic considerations.⁹³

So far, this section has considered how patiency already plays a role in the moral responsibility literature in order to show that the tectonic theory's fundamental reason for adding patientic considerations to moral responsibility judgments is already in the air, so to speak, and need not be tied to determinism/incompatibilism. For the most part, the only sorts of patiency that have currency in the moral responsibility literature are either inter-personal patiency, or the purportedly agency-undermining patiency of determinism-emphasizing incompatibilists. The latter sort of patiency was rejected in the previous chapter because it insists on the metaphysical illegitimacy of moral responsibility practices, which is untenable from a socio-normative

⁹³ This is not Dennett's position, of course, but I do think it is compatible with his views, which case will be made in chapter 4. Dennett thinks it is important that moral responsibility practices be limited to ascertaining a basic moral competence in persons, which legitimizes their moral responsibility. Dennett's opposition to complicating this is a result of the fact that the only alternative he entertains is moral responsibility eliminativism. Since the tectonic theory advocates for retaining moral responsibility practices while building into them a routine of considering persons as patients after they have already been categorized as agents, it is not clear that Dennett would object to the theory. In chapter 4, I will argue that the problems Dennett has with Waller and Greene and Cohen's eliminativist views should not apply to the tectonic theory.

perspective. The former sort, interpersonal patiency, however, provides a form of patiency whose importance is comparatively agreed upon, and which thus can be used to make further distinctions.

The most important distinction to make between different types of patiency is a standard feature of psychological investigations concerned with personality: the distinction between traits and states. These terms are used with more or less their everyday meaning. Roughly, traits are enduring features of a person across times and situations, while states are temporary changes in a person, typically prompted by a particular event or stimulus.⁹⁴ This distinction is more basic than the difference between other types of patiency. Interpersonal patiency, for example, can manifest either by contributing to the formation of a person's traits, or by altering their states. The familiar cases of one person coercing another into performing an action, say, with a gun to the head, are examples of interpersonal patiency whose relevance for moral responsibility is related to the state of the coerced. For an example of trait related interpersonal patiency, consider a person who has been indoctrinated into a cult, resulting in major alterations to her personality.

It is important to make the state/trait distinction with respect to patiency because there are similar distinctions already extant in both moral philosophy and legal codes. In moral philosophy, a distinction is often made between excusing conditions and exempting conditions.⁹⁵ Excusing conditions relate to the states of a person leading up to and during the performance of an action.

⁹⁴ For a more detailed discussion of the various elements of the trait/state distinction as well as an endorsement of its usefulness, see "Conceptual note on state, trait, and the state-trait distinction," by Bram M Fridhandler (Fridhandler, 1986).

⁹⁵ See (P. F. Strawson, 1962)

If a person has been threatened with severe harm so that she performs an action, she is typically excused from moral responsibility for that action while remaining a candidate for moral responsibility in general. In contrast, if a person is psychologically abnormal, e.g. suffering from severe schizophrenia, she may be considered exempt from moral responsibility altogether. In many legal systems, a distinction is made between temporary insanity, which claims that the state of an actor leading up to and throughout an action constitutes a severe deviation from the actor's usual states which can be expected not to re-occur, and insanity, which alleges that an actor is generally incapable of rational conduct.

Having made the basic distinction between state and trait related patiency, I can now explore different types of patiency that can manifest in either of these two forms. This discussion will be more exploratory than argumentative. I would like to map out some categories for the sake of organization more than the making of any philosophical claims. Further empirical study is required to sharpen up these distinctions, some of which may turn out to be comparatively idle within tectonic moral responsibility judgments. Very little work has been done on the different ways that persons' patiency can mitigate moral responsibility judgments, and so I'm guided here by my own avowedly defeasible intuition.

Returning to the relatively familiar case of interpersonal patiency, I propose that a distinction between person to person patiency and group to individual patiency could be useful. Consider a person who is subject to relentless belittling by a narcissistic boss at work, and eventually finds herself no longer disposed to taking a stand against similar treatment of others in the office. Even

if we consider the indolence of the belittled to be a morally culpable offense, the fact that it results in part from the harsh conditions instituted by her boss might mitigate the degree to which she should be blamed in the eyes of some people. Now consider someone who was raised in a severely racist culture and as an adult proudly espouses the racist views to which she was exposed. This person's degree of culpability might also be mitigated by the patientic consideration of her upbringing, and plausibly to a different degree.

My own intuition is that it is easier to expect an individual to outcompete the patientic factors that result from a single relationship than those that result from being embedded in a particular culture. This is just my personal intuition, though, and others may even have the opposite view. To my knowledge, psychologists have studied neither the relative resistability of personal and cultural influences, nor the extent to which observer's moral responsibility judgments are mitigated by their consideration. Both could be profitably addressed empirically, and I think it is a virtue of the tectonic theory that it allows for at least a defeasible hypothesis about this. In the absence of any data on the matter, I think it is worth distinguishing between person to person patency, for which I will use the term interpersonal patency, and group to person patency, which I will call cultural patency.⁹⁶

In addition to the interpersonal/cultural distinction, I think it is plausible that people's intuition will react differently to psychological and biological patientic factors. Psychological patency

⁹⁶ The state/trait distinction could be useful for empirical study of this distinction. It could be that interpersonal patency more often manifests in state related patency, and cultural patency more often manifests in trait related patency, associating these loosely with excusing and exempting conditions. As above, I stress that this is just my personal intuition, which could profitably be defeated by empirical data.

can be relatively neatly decomposed into state and trait related factors. The distinction between biological and psychological patency is tricky, however. From a philosophical perspective, it is difficult to draw a meaningful line between the two because most contemporary philosophers are naturalists committed to the idea that psychological phenomena ultimately are biological phenomena. Nevertheless, there is some empirical work that shows biological explanations of behavior to be more intuitively mitigating of moral responsibility than psychological explanations.⁹⁷ This sort of work suggests that even though the boundary between psychological and biological phenomena will likely change shape as discoveries about the biological bases of psychological features accrue, people can be expected to assign greater mitigating value to those conditions that have biological explanations. It is thus worth making the distinction for the foreseeable future.

The controversial case of Mr. Oft, who developed pedophilic tendencies seemingly as a result of a brain tumor, is a good way to illustrate this. Neuroscientist David Eagleman presents the case of Mr. Oft as straightforwardly demonstrative of the extent to which persons in general overestimate the degree of control their thinking selves have over their conduct, and drastically underestimate the contribution of uncontrollable neurological factors.⁹⁸ For Eagleman, failing to interpret Oft's tumor as mitigating betrays a problematic bias toward the agentic power of the mind and an equally problematic bias against the contribution that biological factors make to behavior. Although my intuition is to agree with Eagleman, I acknowledge that he might be

⁹⁷ (Monterosso, Royzman, & Schwartz, 2005)

⁹⁸ (Eagleman, 2012)

wrong. Stephen J. Morse is perhaps soberer in insisting that “we do not know whether Oft could not—that is, lacked the ability to—control his sexual behaviour, or whether he simply did not.”⁹⁹

Morse’s view is that Oft may have possessed the psychological fortitude to resist the urges that resulted from his biological misfortune. He is right that we can’t claim to know whether this is the case. Nevertheless, Morse grants that “it is not absurd to think that perhaps Oft deserved mitigation or excuse for his sexual deviance on the ground that he could not control himself.”¹⁰⁰

The tectonic theory’s concern for patiency allows for taking a different angle on this case. Where Morse insists that mitigation or excuse requires a fact about whether Oft was able to control himself, the tectonic theory looks toward mitigation on the basis that Oft was clearly unable to control for the presence of the tumor. The extent to which the tumor should mitigate his responsibility, then, can be determined by individuals for themselves based on their personal speculations as to how irresistible Oft’s urges probably were.

This may seem like a non-result, but I think it is a virtue of the tectonic theory that it does not by itself lead to a certain verdict on such a case. Morse is correct that we do not know what it was like to be Oft as his tumor developed, which entails that idiosyncratic factors in observers’ perspectives play a role in addressing the issue. The tectonic theory is built to be agile in this way, allowing for divergent judgments in particular cases, the friction between which may be useful in bringing to light possibly problematic preconceptions about the relative weights of agentic and patientic considerations. Additionally, the tectonic theory is agile insofar it is

⁹⁹ (Morse, 2011, pg. 32)

¹⁰⁰ (ibid)

susceptible to refinement by empirical results. Fortunately, cases like Oft's are rare, but over the long term, data could accumulate. If in a hundred years' time dozens of such cases were studied, and it turned out that Oft was among a minority who failed to resist the urges resulting from a brain tumor like his, the tectonic theory would be able to take this into account, helping people to make an informed judgment about such cases by reducing, though probably seldom eliminating, the role played by observers' idiosyncratic views about the relative weights of agentic and patientic factors.

Conclusion

This chapter was intended to acquaint readers with the concept of patientcy, and how it can figure in moral responsibility judgments. Mainly, this was accomplished by presenting Reader's work on the impoverishments of agentic analyses of persons. To connect these considerations with moral responsibility, Dennett's analysis of control was used to show that many of the ways in which persons are patients can be relevant to moral responsibility. Finally, a brief taxonomy of types of patientcy was delineated. The taxonomy began with a discussion of manipulation cases, in which interpersonal patientcy was shown to play a central role. The relative ease with which the tectonic theory clarifies why manipulation cases typically provoke mitigation or elimination of the manipulated's moral responsibility was presented as a virtue of the theory.

Additionally, the tectonic theory enables the making of distinctions between a number of patientic factors that the moral responsibility literature typically addresses only as vindicators or vitiators of compatibilism or incompatibilism. The tectonic theory can discriminate between

trait-based and state-based patiency, either of which can result from interpersonal, or cultural patiency. I also made a special case of biological patiency, which has been shown to be particularly significant to observers. These distinctions, as I stressed, constitute a sort of first pass over the range of possibilities. Very little empirical or conceptual work has been done on these issues, and different types of patiency as well as their respective significances should be the focus of further study.

<i>Types of Patiency</i>	Trait-based	State-based
Interpersonal	When one person's behavior causes changes in a second person ...	
	... that constitute character traits of the second person	... that constitute temporary states of the second person
Cultural	When a cultural environment causes changes to an individual ...	
	... that constitute character traits of the individual	... that constitute temporary states of the individual
Biological	When an individual's biological features result in ...	
	... character traits that the individual cannot change	... states that the individual cannot avoid

Chapter three: Empirical Work on Patience

“We need a reliable process with which one part of our psychology can mitigate the ill effects of another part.”

(Paul Sheldon Davies)¹⁰¹

The first chapter of this thesis showed that traditional approaches to moral responsibility (compatibilism and incompatibilism), although not without their merits, each suffer from deficiencies regarding significant details of particular cases of moral responsibility. I introduced my tectonic theory there as an alternative to working with either or both of these. The novel element in the tectonic theory is that it prescribes a concern for the patience of agents, which allows for moral responsibility judgments to incorporate facts about what a person is unable to control without having to invoke determinism. The inclusion of patientic considerations in the tectonic theory simultaneously allows for compatibilist forms of agency to continue scaffolding important social institutions and practices, and for moral responsibility judgments to be sensitive to non-agentic factors in particular cases, which is typically prized by incompatibilists. The previous chapter worked out a definition of patience, and with the help of Reader, showed that it is an important and inevitable feature of persons' lives, thus deserving of attention within moral responsibility judgments. This chapter engages empirical work that involves patience so as to explain why and how patientic considerations can be expected to play an important role in moral responsibility judgments.

There are two currents of scholarship that I address, each centered on the work of a pair of researchers: Kurt Gray and Daniel Wegner, and Philip Robbins and Anthony I. Jack. Gray and

¹⁰¹ (Davies, 2013)

Wegner's work helps to illuminate the features of a person's mental life that either result from or constitute patientic factors, and how these differ from the mental phenomena associated with agency. Robbins and Jack are less concerned with what observers attribute to persons considered in their agency and patiency, and more concerned with the stances that observers take towards persons, so construed.

Gray and Wegner have shown that considerations of agency and patiency make distinctive contributions to moral thinking. Roughly, seeing someone as an agent confers moral responsibility, and seeing someone as a patient confers moral status (or moral concern, or moral rights). More specifically, Gray and Wegner have shown that individuals that are high in agency are typically assigned high degrees of moral responsibility for their actions, while individuals that are high in patiency are typically assigned less. Also, individuals that are high in patiency induce greater concern or compassion than individuals that are high in agency but not patiency. While Gray and Wegner's work is philosophically messy, often overstating its case, it undoubtedly shows that many morally charged events are interpreted through the application of the concepts of agent and patient, and that these concepts make different contributions to moral thinking.¹⁰²

¹⁰² Gray and Wegner's group has also gathered data that suggests patiency is associated with the mind and the body, while agency is more of an exclusively mental phenomenon (K. Gray, Knobe, Sheskin, Bloom, & Barrett, 2011), and that individuals seen as patients are more likely to induce leniency in punitive judgments (K. Gray & Wegner, 2011). These studies will be presented and analyzed in the final chapter. The focus on the moral/legal significance of neuroscience will help to clarify their import.

Robbins and Jack developed the concept of *the phenomenal stance*, which corresponds roughly to seeing a person as an experiencing patient as opposed to a goal-forming/pursuing agent.

Robbins and Jack reviewed large literatures and performed numerous behavioural and neurological studies that illuminate a cognitive rivalry between two basic ways of thinking. The human brain, they contend, includes two distinct systems, one for thinking about inanimate objects, the task-positive network and one for thinking about minded-persons, the default mode network. Jack has proposed the *theory of opposing domains*,¹⁰³ which claims that these two systems mutually inhibit one another.¹⁰⁴ Discerning the import of opposing domains theory for moral responsibility will require some philosophical ablutions. Robbins and Jack contrast the phenomenal stance with what Dennett termed the intentional stance.¹⁰⁵ Taking the intentional stance toward an entity is to think of it as a rational agent capable of forming and pursuing goals. According to Robbins and Jack, taking the intentional stance toward a person, though useful for explaining her behavior, typically results in a failure to appreciate the lived experience of the person, which can be corrected by taking the phenomenal stance.

The way that Robbins and Jack overlay the contrast between the intentional and the phenomenal stance upon the opposing domains mentioned above requires a little more philosophical hygiene than they provide. The intentional stance is described as involving a blend of activity from the two networks, with the balance tilted toward the more analytic and less emotional system, which

¹⁰³ (Jack, 2014)

¹⁰⁴ “there are dissociable brain networks specialized for thinking about an entity’s physical (e.g., causal-mechanical) attributes and for thinking about an entity’s subjective mental life. The first network [the task-positive network] is activated when we take up the “physical stance” toward the entity, the second [the default mode network] when we take up the “phenomenal stance.” The network architecture is configured in such a way that the engagement of one stance inhibits the engagement of the other.” (Jack, Robbins, Friedman, & Myers, 2014)

¹⁰⁵ See (D. C. Dennett, 1989) From the abstract: “The intentional stance is the strategy of prediction and explanation that attributes beliefs, desires, and other “intentional” states to systems - living and nonliving - and predicts future behavior from what it would be rational for an agent to do, given those beliefs and desires.”

is alleged to have evolved to facilitate humans' dealings with the inanimate. The phenomenal stance, in contrast, is characterized as fully constituted by activity in what is known as the brain's *default mode network*, which is implicated in all sorts of social cognition in which an appreciation of others' mental lives plays a role.

Robbins and Jack frequently characterize the phenomenal stance as more significant for moral thinking than the intentional stance, and disparage the intentional stance as a colder, more distanced way of thinking about a person. There is, however, reason to complicate this way of conceiving of the contrast. The so-called *task positive network* that is more analytic and less compassionate is plausibly needed to facilitate reasoned inferences pertaining to the deliverances of the more person-concerned default-mode network.

Taking the issue of blended forms of neurological activity into account, I propose that the tectonic theory should not be understood as simply advocating for the successive application of the intentional and phenomenal stances to a person. Instead, I suggest that the two-step analysis of a person be guided by my more philosophically motivated motifs: first consider what a person is/was capable of controlling, and then consider what the person was unable to control. That which Robbins and Jack describe as the phenomenal stance can be said to provide inputs to the latter process. This schema can mobilize the concepts of agency and patiency within moral responsibility judgments in such a way that opportunities for fairness are maximized.

Gray, Wegner and Colleagues

Kurt Gray and Daniel Wegner have led numerous studies dealing with patiency, beginning with the 2007 paper “Dimensions of Mind Perception,”¹⁰⁶ and culminating with the award winning¹⁰⁷ 2012 paper “Mind Perception is the Essence of Morality,”¹⁰⁸ Gray, Wegner and colleagues have shown that the concept of patiency is highly probative with respect to the nuances of moral thinking. Their sustained interest in patiency began as a way to introduce a finer-resolution on the phenomena of mindedness-attributions.

Gray et al. questioned the traditional view according to which mind-having is attributed along a single dimension, with various things having more or less, but not different types of minds. They queried subjects about the various mental capacities of a wide range of entities,¹⁰⁹ as well as the subjects’ personal judgments about the entities. The mental capacities were grouped into two broad categories: experience and agency. Experiential capacities included the ability to feel hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy, and agentic capacities included self-control, morality, memory, emotion recognition, planning, communication and thought. The results showed that these two sorts of mental capacities can come apart. Although adult humans were ascribed high-agency and high-experience, robots and God were high-agency and low-experience, and frogs and fetuses were low-agency and high-experience. This is the first store of evidence brought to bear on the question of whether mindedness attributions come in different types.

¹⁰⁶ (H. M. Gray, Gray, & Wegner, 2007)

¹⁰⁷ “Mind Perception is the Essence of Morality” was awarded the 2013 Theoretical Innovation Prize from the Society for Personal and Social Psychology

¹⁰⁸ (K. Gray, Young, & Waytz, 2012)

¹⁰⁹ A human baby, a chimp, a dead woman, a dog, a fetus, a frog, a girl, God, a man, a robot, a man in a persistent vegetative state, a woman, and finally the subject herself

The idea that judging an entity to be more agentic or more experiential has some moral content was shown by the subjects' personal judgments about the entities. Some personal judgments correlated with both agentic and experiential mind attributions (liking the entity, wanting to prevent its destruction, wanting to make it happy, and judging it to have a soul), while others correlated with just agency (deserving of punishment) or experience (desire to avoid harming). The authors conclude that their findings “reveal not one dimension of mind perception, but two, and show that these dimensions capture different aspects of morality.”¹¹⁰ Following Aristotle, the authors name the different aspects of morality *agency* (which refers to agency-related mental capacities) and *patience* (which refers to experience-related mental capacities).¹¹¹

From this initial study, a significant literature has emerged. Before presenting the most important results, it is important to clarify the relationship between Gray and Wegner's use of the concept of patience and my own. In their early work on dimensions of mind perception, the word patience occurs just once toward the end of the paper in a reference to Aristotle. They use the term ‘experience’ to mark out the non-agentic content that can be attributed to minds. The connection between experiential mental states and patience is elaborated on in their later work on *dyadic morality*. This is the theory that moral judgments are essentially constituted by a symbiotic pair of elements: an agent that intentionally acts, and a patient that suffers harm as a result of the action. This is interpersonal patience, which was discussed in the previous chapter.

¹¹⁰ (H. M. Gray, Gray, and Wegner 2007, pg. 619)

¹¹¹ “The [agentic and experiential] dimensions [of mind perception] thus relate to Aristotle's classical distinction between moral agents (whose actions can be morally right or wrong) and moral patients (who can have moral right or wrong done to them). Agency is linked to moral agency and hence to responsibility, whereas Experience is linked to moral patience and hence to rights and privileges. Thus, our findings reveal not one dimension of mind perception, but two, and show that these dimensions capture different aspects of morality.” (ibid)

Although it is probably the most intuitively salient form of patiency, I think that dyadic morality neglects other forms of patiency. I share this interpretation with Brent Strickland, Matthew Fisher and Joshua Knobe. Strickland et al. take the moral agent/moral patient template from Gray and Wegner as symptomatic of a more fundamental distinction between agency and patiency in general.¹¹² Before I present the dyadic morality work and this commentary, it is important to compare and contrast Gray and Wegner's concept of experience, and my concept of patiency.

A few of the items included under the experience heading do not fit my definition of patiency as that which a person cannot control. Most of the items making up the experiential aspects of mind in Gray and Wegner's probes are emotional states (fear, rage, pride, embarrassment, joy). Some are physiological states with strong affective elements (hunger, pain, pleasure), and then there are a pair of more complex phenomena (personality and consciousness). The first two groups map well enough onto my definition of patiency as comprised of that which a person cannot control. The emotional states like fear and rage are nearly paradigmatic examples of feelings that arise unbidden in persons. This is not to say that a person's behaviour is determined by such feelings when they occur, but just that their occurrence is not something that can typically be suppressed. Likewise for hunger, pain and pleasure; although these often come about and/or recede due to a person's agentic efforts (or lack thereof), while they are occurring they cannot be suppressed through agentic efforts.

The remaining items of personality and consciousness are trickier. These are more complex ideas that do not clearly constitute uncontrollable phenomena. It is plausible to say that a person has a significant degree of control over her personality; most of us manage to become people that we

¹¹² (Strickland, Fisher, & Knobe, 2012)

wanted to be in at least some respects. However, there is a sense of ‘personality’ that is conventionally taken as a given for persons. “I am the way I am” is a common refrain in morally charged interpersonal interactions. I’d prefer the word ‘temperament’ as a label for the basic affective settings in a person that are commonly thought to be beyond the reach of agentic alteration. A distinction between the temperament of a person, referring to a person’s characteristic responsiveness to certain sorts of stimuli, and personality, referring to the broader phenomenon of a person’s way of interacting with others, preserves a space with which to consider the ways that persons can agentially alter themselves, while at the same time creating a space in which to be alert to the possibility of there being other aspects that a person cannot alter.

The phenomenon of consciousness is, of course, imbued with all sorts of philosophical content, only some of which can be understood as wholly experiential. Philosophers distinguish between many forms of consciousness such as access consciousness and deliberative consciousness.¹¹³ It is, after all, within our conscious minds that moral deliberation occurs, which is a feature of agency par excellence. The type of mindedness known as phenomenal consciousness, which refers to a person’s capacity for incurring particular experiences, however, does seem to have an element of uncontrollability. A person is not able to directly change the experience they have when seeing red, for example.¹¹⁴

This concept of phenomenal consciousness, though, does not distinguish between agentic and patientic phenomena. Persons often consciously experience their own agency, as well as

¹¹³ See (Block, 1996)

¹¹⁴ It is worth noting that this claim about not being able to directly alter perceptual experiences applies only at the time of the experience. Over the long term, it is possible to change some of the ways we experience things through agentic efforts. For example, through some hard work, I eventually developed a taste for beer, which at first I found to be distasteful.

whatever further phenomena it brings about. For example, a paramedic working with an injured person may be very lucid about her agentic efforts to help the injured person while at the same time incurring an experience of the episode which she is not able to directly alter during the episode. The phenomenal content in experiences of one's agency might not be something that persons can change at the time, but attempting to incorporate the way it feels for a person to be the agent that they are into morally responsibility judgments seems overly elaborate to me. If there were an aspect of a person's experience of their own agency that is relevant for moral responsibility judgments, such as a kleptomaniac's overwhelming excitement at planning a theft, this would be better captured as just an uncontrollable feature of the person's behavioural disposition (trait-based patiency) rather than as a feature of the person's experience.

Classifying such phenomena as aspects of a person that the person cannot agentially alter instead of as features of her experience preserves a space to recognize that a person might not have the relevant experience, as when situational factors exert unacknowledged influence on behaviour. Philosophers and psychologists refer to this phenomenon as *situationism*.

Situationism claims that persons' actions are best explained in terms of features of the circumstances in which persons find themselves rather than in terms of character.¹¹⁵ Although situationism is more often taken as a challenge to claims about persons having stable character traits, because the influence of situational factors has been shown to lie outside the awareness of persons in some cases,¹¹⁶ it has also been addressed in the moral responsibility literature.¹¹⁷

¹¹⁵ See (Doris, 1998) for a good discussion of the philosophical implications of situationism

¹¹⁶ See (Harman, 2003)

¹¹⁷ See (Woolfolk, Doris, & Darley, 2006), (Doris & Murphy, 2007), (Brink, 2013), (Schlosser, 2013), and (Vargas, 2013b) for discussions that wrestle with situational influences and constraints in the context of moral responsibility. Vargas' treatment of situationism will be dealt with in detail in chapter five.

Because of the situationism issue, it seems prudent to allow the concept of patiency to range over uncontrollable non-conscious factors that play a role in a person's behaviour as much as uncontrollable conscious factors. In fact, Gray and Wegner's focus on mental states, although perfectly understandable from psychologists, is problematic in the context of philosophical theorizing about moral responsibility. The mental state-focus is particularly out of step with my endorsement of including patientic considerations in moral responsibility judgments because I defined patiency in such a way that it can be partially constituted by non-mental factors.

I will show that Gray and Wegner's main concern is with the mental-states that observers ascribe to persons seen as patients. In contrast, my concept of patiency is intended to target a person in the world, inclusive of her mental states but not exhausted by them.

A natural way of describing Gray and Wegner's work on experience is thus that it illuminates an enabling condition for tectonic judgments. As Reader emphasizes, it must be granted that a person is just a recipient of large parts of her life, and this plausibly entails recognition of a person's experiential dimension. The previous chapter argued that if someone is considered just in her agentic features such as her beliefs, desires, and goals, her personhood has only been partly accounted for. Gray and Wegner's concept of experience provides the mental element that would be left out of such an analysis. The items they use to probe experiential ascriptions such as pain and pleasure represent patientic mental phenomena because they are not subject to agentic alteration while they are occurring. A recognition that such phenomena exist and are patientic is plausibly needed to open the door to more wide-ranging patientic considerations. If it were denied that persons' minds are partially constituted by experiential and patientic states of the sort Gray and Wegner describe, this would severely limit the usefulness of the tectonic theory.

By showing that some patientic phenomena play out within the lived experience of persons, Gray and Wegner's work provides intuitive reason to recognize the importance of patientic phenomena: patientic phenomena can be important because they can be constituted by persons' experience. The more abstract patientic considerations encouraged by the tectonic theory could probably not get off the ground if it is not granted that persons' minds are not only the site of agency, but can also be the site of patientic experience. Gray, Wegner, and colleagues show that a person ought to be understood as capable of incurring experiences that she cannot control. Reviewing their work will illuminate much of what it is to consider a person as a patient, as well as how attitudes toward a person so considered are different from attitudes toward a person considered as an agent. To complete this task, I will present Gray and Wegner's work on dyadic morality, and moral typecasting.

Dyadic Morality

Once their work on the different dimensions of mind perception got the ball rolling, Gray and Wegner continued to investigate the distinct contributions that agency and experience attributions make to people's thinking about moral matters. Two findings in particular inspired them to formulate the ambitious claim that the essence of morality is the dyad of an intending agent and a suffering patient.¹¹⁸ These are the phenomena of dyadic completion and moral typecasting.

Moral typecasting will be dealt with in detail in the next section because it is particularly

¹¹⁸ It is worth mentioning that Gray et al. also appeal to the moral deficits peculiar to autism spectrum disorders and psychopathy to motivate their model. Roughly, autistic people have difficulties attributing agency to others, and psychopaths have difficulties attributing experience/patientic to others. Robbins and Jack also make these appeals, but add consideration of sufferers of Williams' Syndrome to the analysis, and so I will wait to present this material in the section on their work.

important for explaining the tectonic theory's two-step structure. I will therefore just discuss dyadic completion to provide a sense of what motivated the dyadic morality thesis.

Dyadic completion both supports and clarifies Gray and Wegner's claim that the *intending agent* + *suffering patient* dyad constitutes a cognitive template that structures thinking about moral transgressions. Gray and Wegner found evidence for the template claim by showing that "when we see someone blameworthy—an apparent moral agent—we should complete the dyad by inferring the presence of another mind to suffer—a moral patient. Second, when we see a suffering patient, we infer the presence of another mind to take responsibility as a moral agent."¹¹⁹ A familiar example of this is the phenomenon of people blaming gods and spirits for suffering.¹²⁰ The phenomenon also occurs in small settings. Carey K. Morewedge showed that subjects who received poor outcomes in economic games were more likely to see their fate as the result of an agent's doing.¹²¹

The other sort of dyadic completion, the inferring of a victim from the presence of an intending agent, has been shown by Peter DeScioli, Sarah S. Gilbert and Robert Kurzban.¹²² DeScioli et al. argue that the dyadic morality model is a bit oversimplified and suggest a model according to which "the moral faculty can be switched on by a variety of factors that compose cognitive models of moral events ... suffering victims ... menacing perpetrators, righteous punishers, or specific violating actions ... [when some of these are present] moral cognition could seek to fill

¹¹⁹ (K. Gray, Young, et al., 2012, pg. 111)

¹²⁰ Gray and Wegner cite, among others, the work of Pascal Boyer, which documents many incidences of this phenomenon in many different cultures, (Boyer, 2001)

¹²¹ (Morewedge, 2009)

¹²² (DeScioli, Gilbert, & Kurzban, 2012)

the remaining slots of the cognitive template with the best available alternatives."¹²³ They did, however, perform a study that aimed to verify the "victim completion" arm of dyadic completion. Subjects were given 12 moral "offenses" with ambiguous or non-existent victims (e.g. suicide, grave desecration, drug use) and asked, whether the action was wrong, whether someone was wronged, who was wronged, whether the actor should be punished, and who should do the punishing. Wrongness judgments significantly predicted victim perception. Subjects were 43 times more likely to perceive a victim when an action was judged wrong.¹²⁴

The fact that people are disposed to dyadic completion provides significant support for Gray et al.'s dyadic model of morality. It does seem to be the case that moral transgressions are subject to the agent/patient template. However, it remains to be established why this is the case. If it is so because of something specific to cognition related to moral wrongs, the dyadic thesis would be of little import to the tectonic theory, because the tectonic theory is intended to range over many other factors (whatever can or cannot be controlled by a person) aside from actions judged to be moral transgressions. Fortunately, Brent Strickland, Matthew Fisher, and Joshua Knobe argue persuasively that the agent/patient template plays a role in human cognition of events in general.¹²⁵

Strickland et al. put forward the hypothesis "that the structure [Gray et al.] finds in moral cognition actually falls out of a structure one finds in event cognition more generally."¹²⁶ They

¹²³ "In sum, participants who condemned behaviors as wrong perceived a victim 89% of the time, whereas other participants responding to the same stimuli, but judging behaviors as not wrong, perceived a victim only 15% of the time." (DeScioli et al., 2012, pg. 145)

¹²⁴ (DeScioli et al, 2012, pg. 144-145)

¹²⁵ (Strickland et al., 2012)

¹²⁶ (ibid, pg. 199)

appeal to theoretical linguistics to motivate the view. Essentially, Strickland et al. marshal data from linguists about verbs to show that the agent/patient distinction is present in the syntactic structure of all languages. In the first place, they note that languages always contain transitive verbs, which require both a subject (agent that performs an action) and an object (patient that receives the action). This is merely suggestive of the truth of their hypothesis, however. More forceful is their claim that human languages never contain verbs whose subject is the patient.¹²⁷ Given the myriad ways that languages do vary in their structure, the fact that the agentic and patientic roles appear to be robust features of all them is compelling support for the hypothesis the agent/patient distinction is a general feature of human cognition, as opposed to a specifically moral phenomenon as claimed by Gray et al.

Strickland et al. also performed a study to test whether being the subject or object of a verb made a difference to people's agency and patiency ratings. They provided subjects with pairs of sentences from one of two conditions (morally neutral or morally charged). The first sentence introduces two people who were chosen because they can be expected to be taken to vary in their respective degrees of agency and patiency (an older man and a younger woman) and the second sentence described the two people kissing, varying which was presented as the subject.¹²⁸

Subjects were then asked a number of questions designed to probe intuitive reactions to these

¹²⁷ “suppose we define a verb *shmite* whose meaning is simply the converse of *bite* (so that it means something like be bitten by). Then it might seem that we could switch everything around and use a sentence like: 1'. The man *shmit* the dog. [which would have the same meaning as “The dog bit the man.”] Yet research in linguistics suggests otherwise. Sentences like this one do not appear to occur in any human language.” (Strickland et al., 2012, pg. 199)

¹²⁸ “MORALLY CHARGED: 1. Steven is 25 years old and Kate is 15 years old. Steven French-kissed Kate. 2. Steven is 25 years old and Kate is 15 years old. Kate French-kissed Steven; MORALLY NEUTRAL 3. Steven is 35 years old and Kate is 25 years old. Steven French-kissed Kate. 4. Steven is 35 years old and Kate is 25 years old. Kate French-kissed Steven.” (Strickland et al., 2012, pg. 202)

scenarios.¹²⁹ Strickland et al. found that being the subject of the kiss-describing sentence had a major impact on intentionality ascriptions for both the morally charged and morally neutral conditions. Syntactic subjects were rated as having acted significantly more intentionally regardless of which person was the subject or whether the scenario was morally neutral or charged.¹³⁰ Responsibility and punishability ratings followed suit, but were less drastic. The upsetness probe, in contrast, went the other way with the verb's object being consistently rated as more upset. Regression analyses showed that intentionality ascriptions were predictive of responsibility and punishability ratings.¹³¹

These findings indicate that whether a person is presented as an agent (linguistic subject) or a patient (linguistic object) has more of an influence on subjects' judgments than the presumed social cues about age, gender, and moral transgressiveness.¹³² Strickland et al. conclude that "the importance we see of agency and patiency within the moral domain simply "falls out" of the importance of these roles in people's cognition more generally."¹³³ This brings up the question of why the roles of agent and patient contribute structure to human cognition in such a general way. Answering this question is not straightforward. Strickland et al. suggest that the agent/patient

¹²⁹ "How intentionally did Steven act? How responsible was Steven for the act? How likely is Steven to be upset? How harshly should Steven be punished?" (Strickland et al., 2012, pg. 202)

¹³⁰ See table 1 (Strickland et al., 2012, pg. 202)

¹³¹ See the Mediation section (Strickland et al, 2012, pg. 203)

¹³² "Another remarkable aspect of these findings is that gender (older man vs. young woman) did not lead to any meaningful differences between groups. In other words, attributions of agentic properties like intentionality and responsibility did not differ as a function of gender/age. Similarly, attributions of patient-like properties also failed to differ as a function of gender/age. Perhaps this is because the effect of syntactic cues is so large that it simply swamped any influence of social cues." (Ibid, pg. 204)

¹³³ (Ibid)

distinction may constitute an innate template with which humans are equipped and which can be used to understand events generally.¹³⁴

Claims about innate cognitive templates are tricky to confirm, of course, but Strickland et al.'s work forcefully shifts the burden of proof to the doubters. They provide sufficient evidence to proceed with the view that agentic and patientic aspects of a person and her life can be distinguished outside the domain of moral transgressions that Gray et al. focus on. This is important for the tectonic theory because it suggests that a person's agentic and patientic aspects can be assessed independently of moral dyads, which enables more general consideration of each to be counterbalanced by consideration of the other, as the tectonic theory prescribes. The fact that Strickland et al. found that people's thinking about even non-morally charged actions is sensitive to the agent/patient distinction, and that these sensitivities carry the same freight that Gray and Wegner found in the moral domain (agents are more responsible for events than patients, and the effects of events on patients are more concerning than on agents) suggests that the tectonic theory may well provide a way to sophisticate moral responsibility judgments as advertised.

¹³⁴ "Why is it that people show this general tendency to understand events in terms of agents and patients? Our results do not speak directly to this deeper question, but we do want to suggest one possible hypothesis about the answer. Specifically, we propose that people's more general tendency to understand events in terms of these roles may arise from the structure of what is sometimes called 'core knowledge.' That is, it may be that people have an innate preparedness to understand events in terms of these roles and that this innate preparedness then structures their thinking about events in numerous domains. People's core knowledge of role assignments could then explain the patterns we observe in their moral judgments, in their linguistic intuitions, and perhaps in many other areas besides." (Strickland et al., 2012, pg. 205)

In addition, that the agent/patient distinction is so wide-ranging provides partial explanation for the tectonic theory's two step structure. If a person can be seen as either an agent or patient with respect to large numbers of aspects of her character and life, it makes sense to tabulate these one at a time so as to achieve thorough overviews of each. Such a process allows for each's import to accumulate without reference to the other, thus maximizing the potential for both agentic and patientic considerations to contribute nuance to moral responsibility judgments. These reflections, however, speak only in favour of the view that it would be effective to carry out moral responsibility judgments in a tectonic way. That the tectonic structure is necessary for injecting a fairness-enhancing degree of nuance into moral responsibility judgments requires further argumentation. To make this case, I turn to Gray and Wegner's moral typecasting studies, which will provide a sort of bridge to the neurologically informed work of Robbins and Jack. Altogether, I think this work does suggest that a tectonic structure is *required* to enhance the fairness of moral responsibility judgments.

Moral typecasting

Gray and Wegner's *moral typecasting thesis*¹³⁵ claims that perceptions of agency and patiency are inversely related such that perceiving an individual as an agent makes it difficult or impossible to perceive her patiency-related characteristics and vice-versa. They suggest that mind perception may be a two-step process, with mind-having as a general first pass which affirms the presence of experience and agency followed by a closer look which selects just one of these as distinctive of a person. They perform seven separate studies to investigate whether

¹³⁵ (K. Gray & Wegner, 2009)

and when moral agency and moral patiency perception are inversely related. All of these studies provide some support for the moral typecasting thesis. However, Gray and Wegner's claims are about how people perceive the mental aspects of others. The typecasting thesis has little direct relevance for moral philosophy, wherein cultivated reflectiveness is more of a factor than perception. Indirectly though, as a claim about the different sorts of thoughts that can be included in moral responsibility judgments, moral typecasting contributes part of the explanation for why the tectonic theory has its two step appraisal.

Gray and Wegner developed the moral typecasting thesis as an interpretation of the apparent independence of agentic and experiential attributions established in their work on dimensions of mind perception. As described above, this work showed that different entities will be ascribed different blends of agentic and experiential capacities. The moral typecasting studies, in contrast deal just with people. In the initial study on dimensions of mind perception, it was found that adult humans are typically ascribed high degrees of agency and experience, which speaks clearly against the moral typecasting thesis.¹³⁶

Although Gray and Wegner do not do so, this can be explained. The dimensions of mind perception study measured the variation in agentic and experiential attributions across a wide range of entities. In comparison to robots, frogs, and children, adult humans are perceived to be comparatively highly agentic and highly experiential. The moral typecasting studies, in contrast, probe for variation in the ways that different people are construed as more or less morally agentic or patientic. The studies are designed to test whether seeing someone as a moral agent reduces the likelihood of that person being ascribed experiential capacities, and/or whether seeing

¹³⁶ This is among the problems with the typecasting thesis identified by Adam Arico (Arico, 2012).

someone as a moral patient reduces the likelihood of that person being ascribed agentic capacities.

Unfortunately, the patiency measures involved in the typecasting studies are very limited.

Whenever Gray and Wegner investigate the degree of patiency ascribed to a person, the query concerns the person's sensitivity to pain and/or pleasure. Such measures presume that when a person is seen as more of a patient, she will be judged as more sensitive to pleasure and pain, and when person is judged to be more of an agent, she will be taken to be less sensitive to pleasure and pain. The root of this problem is the way that Gray and Wegner set up their experiments to probe observations about others' *capacity* for patiency. It is odd, if not incoherent, to think of a person's sensitivity to pain as a capacity. A capacity to feel pain is at least not an agentic capacity insofar as it cannot be exerted according to a person's desires. The limitations of this approach can be shown through an analysis of one of the more provocative results from Gray and Wegner's studies. They showed that subjects were more willing to harm individuals high in agency than average people, even when the agent in question is supposed to be highly admirable, like the Dalai Lama or Mother Theresa.¹³⁷ Gray and Wegner designed a study in which subjects were given a number of pain pills and pleasure pills that needed to be distributed between two people.¹³⁸ When either the Dalai Lama or Mother Theresa was paired with an average person, these putative saints were given a greater number of pain pills, with the average person being given a greater number of pleasure pills.¹³⁹ This result is alleged to support moral typecasting

¹³⁷ Study 7: Harming the Saints (K. Gray & Wegner, 2009, pg. 515-516).

¹³⁸ "There were seven targets: two good agents (Dalai Lama, Mother Theresa), two neutral targets (high school teacher, radiology technician), two patients (orphan, date rape victim), and one bad agent (serial killer Ted Bundy). Each target was paired with every other target, yielding a total of 21 pairs." (K. Gray & Wegner, 2009, pg. 516)

¹³⁹ Gray and Wegner describe the results thusly: "people assign both positive and negative outcomes on the basis of perceived moral patiency, giving the most pain to the agents, the least to the patients, and an intermediate amount to neutral targets. These results reveal a paradoxical effect of moral typecasting; namely, doing good deeds can lead

insofar as being seen as a highly effective agent correlates with being seen as less sensitive to pleasure and pain. However, the finding is highly interpretable. It could also be that when construing someone as an agent, one is less concerned with their experiential states, rather than having judged such states to be less intense. Correspondingly, one might be more concerned with the experiential states of someone who is being construed as a patient, as opposed to having judged such states to be more intense.

In other words, it could be that agent and patient construals vary the extent to which their beholder is concerned with a person's pain and pleasure. If this were so, the fact that subjects were more willing to harm the Dalai Lama than an ordinary person would indicate that subjects felt that it was less important for them to protect him from harm than it was for them to protect an ordinary person from harm. The fact that seeing a person as an agent seems to reduce subjects' judgments about her sensitivity to pain and pleasure¹⁴⁰ can be accounted for by the way that a person construed in her agentic aspect is a person whose degree of control over herself is emphasized. Awareness of a person's capacity to control herself and her life plausibly reduces concern for her experiential sensitivities because in attending to a person's agency one is in part attending to the person's capacity to bring about pains and avoid pleasures through her agentic efforts. Since a person seen as an agent is already seen as having a normal degree of control over herself and life, there is less reason for an observer to take on these concerns.

others to perceive you as less sensitive to pain, thereby making them more likely to hand out pain to the good-doer. Comments from participants reflect this idea; as one person said of the Dalai Lama, 'With all he's done, I figure he can take the pain.'" (ibid)

¹⁴⁰ Gray and Wegner took the fact that agents are seen as less sensitive to pain and pleasure to be established by studies 1a, 1b, and 2. Study 1a showed that adults are judged to be less sensitive to pain than less agentic children; 1b showed that normal agentic adults are seen as less sensitive to pain than less agentic mentally handicapped adults. Study 2 looked at a wide variety of adults (Mother Theresa, Dalai Lama, and Martin Luther King Jr., Osama bin Laden, Hitler, and Ted Bundy, an orphan, a victim of date rape, and a man with Down's Syndrome, a radiology technician, Britney Spears, a network administrator, Prince Harry, and a high school teacher) and found that the more agentic persons were judged to be less sensitive to pain.

The moral typecasting thesis was criticized in a similar way by Arico,¹⁴¹ who finds Gray and Wegner's support for the inverse-relatedness of agency and patiency ascriptions/perceptions to be insufficient. Arico accepts that Gray and Wegner have shown that ascribing a high degree of agency to an entity correlates with ascriptions of low patiency (and vice-versa), but denies that a causal relationship has been shown. It is therefore still possible, according to Arico, for the same individual to be assessed in both her agentic and patientic aspects. I eagerly accept this, of course, but argue that a very weakened reading of the moral typecasting thesis still holds: it is difficult for people to consider the same individual's agentic and patientic features at the same time.¹⁴²

I do not claim that there are any strong impediments to philosophical theorizing that attunes to both agentic and patientic features of the same individual, but only that philosophers who wish to do so will need to somehow overcome the tendency to focus on just one of these at any one time. The tectonic features of my theory of moral responsibility are intended to address this difficulty by suggesting that an individual's agency be ascertained in the usual way, with investigation of her patiency and whether it ought to mitigate her responsibility happening subsequently. Some such structure appears to be needed given the distinct contributions that agentic and patientic considerations make to moral thinking.

Gray and Wegner's work has been commonly interpreted as showing that seeing a person as an agent is what makes them morally responsible, and seeing a person as a patient is what grants

¹⁴¹ (Arico, 2012)

¹⁴² I should note that this view is partially motivated by the work of Robbins and Jack, and can't be effectively supported just on the basis of Gray and Wegner's typecasting studies.

them moral status, which is roughly to be deserving of moral consideration. A recent article by Jordan Theriault and Liane Young¹⁴³ summarizes this division of labour hypothesis well.

Theriault and Young review work from Gray and Wegner as well as that of Robbins and Jack, which they take to provide sufficient support to claim that features of a person's mental life really do fall along one of two dimensions, either agency or experience.¹⁴⁴

Regarding the alleged division of labour between these attributions, Theriault and Young are less certain. They present work from Justin Sytsma and Edouard Machery which claims that moral status can be conferred as a result of discerning agency in others, even if it is more usually a function of attributing a capacity for experience.¹⁴⁵ I will not address this complication because for the purposes of the tectonic theory it doesn't matter how moral status is conferred upon persons whose moral responsibility needs to be ascertained. It is, of course, crucial that the subject of tectonic moral responsibility judgments be granted moral status, i.e. it is crucial that the performer of a tectonic judgment believes that what happens to the subject of the judgment is important. However, whether the treatment of tectonic judgment's subject becomes a concern as a result of the judger's appreciation of the judgee's agency or patiency is not important. It is plausible, given the findings of Gray and Wegner and Robbins and Jack (and their respective colleagues), that the tectonic theory's prescribed concern for patiency will often serve to emphasize or enhance the moral status of judges. This is because the intentional stance, which construes persons as agents, less typically generates concern for persons than does the phenomenal stance with its interest in a person's experiential aspect. Thus, even if attributing

¹⁴³ (Theriault & Young, 2014)

¹⁴⁴ "Paralleling the distinction between an intentional and phenomenological stance are the dimensions of Agency and Experience ... which pair moral responsibility with the attribution of Agency, and moral rights with the attribution of Experience." (Theriault & Young, 2014, pg. 103)

¹⁴⁵ (Sytsma & Machery, 2012)

agency to persons can sometimes serve as a generator of concern for their experiences, the tectonic theory's concern for persons' patiency will be there to generate this concern whenever it fails to be the passenger of agency attributions.

The way that concern for a person's patiency can emphasize or enhance the moral status of persons makes a major contribution to the tectonic theory's claim to fostering less punitive moral responsibility judgments, but it is not the whole story. To the extent that patientic considerations enhance generalized moral concern, they plausibly dampen retributive impulses by making observers alert to the consequences of retributive action. More simply put, when one has some degree of care for the future experiences of a person, it is probably more difficult to endorse harsh punitive treatment of that person. But the tectonic theory's prescribed concern for patiency is not a simple instruction to care about the subjects of moral responsibility judgments. The tectonic theory asks that moral judgments consider behavioural factors that a person does not control because this is fairer than not considering them. That such consideration enhances concern for a judgee as a potential recipient of punishment is actually a product of human neurology. As we will see in the next section, Robbins and Jack have shown that taking a more empathic stance toward a person involves more neural activity in the default-mode network, which happens not to specialize in discerning persons' agentic aspects.

The human brain appears to be configured in such a way that whenever a person is being considered in their non-agentic aspects, which by my definition constitute her patiency, empathy and compassion toward the person flow more easily. But this does not mean that empathy and compassion do all the tectonic work of mitigating the punitiveness of moral responsibility

judgments that are inaugurated by agentic considerations. In contrast, I suggest that a potentially minimal degree of empathic concern is all that's required for the tectonic theory to function as advertised, and that some quantity of empathic concern is an inevitable upshot of considering a person as a patient. The explanation for this comes from Robbins and Jack's work, to which I now turn. We will see that the empathy generated by the phenomenal stance's attunement to a person's experiential or patientic aspects, although important insofar as it creates a context of basic concern towards the person, is just one aspect of how the tectonic theory can work. More significant is the simple fact that directing focus toward a person's patientcy must as matter of neurology funnel attention away from her agency, which enables consideration of uncontrolled factors in a person's life to compensate for the impoverishments of a purely agentic construal.

Robbins and Jack

Robbins and Jack developed the idea of the phenomenal stance as a way of explaining the intuitions that give rise to what Joseph Levine describes as the explanatory gap¹⁴⁶ between materialist and psychological explanation.¹⁴⁷ They argue that difficulties within the project of naturalizing the mind are related to moral thinking. Roughly, Robbins and Jack explain that different cognitive sub-systems evolved in humans for dealing with the different sorts of things in the world. Taking their cues from Dennett's nomenclature, Robbins and Jack delineate three stances that a person can take toward items in the world. There is the physical stance/folk physics that engages with inanimate objects. There is the intentional stance/folk psychology that engages with agents, i.e. goal pursuing entities. Finally, there is the phenomenal stance that engages with

¹⁴⁶ (Levine, 2001)

¹⁴⁷ (Robbins & Jack, 2006)

patients, i.e. experience-prone entities.

Robbins and Jack motivate this threefold distinction by considering the deficits involved in various psychological disorders. People who suffer from autism have been shown to have trouble ascertaining the agentic aspects of others, while their sensitivity to others' experiences remains intact.¹⁴⁸ Sufferers of Williams' Syndrome perform poorly on tasks that probe the efficiency of reasoning about physical objects, but appear to have intact sociality.¹⁴⁹ Finally, psychopaths have been shown to perform normally or better than average on mindreading tasks,¹⁵⁰ but well below average on empathic tasks.¹⁵¹ These clinical studies, the data from Gray and Wegner's work, as well as a number of studies performed by Robbins and Jack themselves,¹⁵² provide strong support for these three stances being relatively autonomous within human cognition.

Robbins and Jack's major claim is that the human brain evolved in such a way that makes information wrought through application of the phenomenal stance particularly difficult to integrate with information arrived at by taking the physical stance. They claim that this is why so many people have intuitions about an explanatory gap afflicting attempts to explain consciousness/phenomenal experiences in physical terms. The fact that different things in the world recruit different neural networks to provide understanding, applying the physical stance to objects and the phenomenal stance to persons, predisposes humans to believe that these are fundamentally different kinds.

¹⁴⁸ (Frith, 2003), (Baron-Cohen et al., 1997)

¹⁴⁹ (Bellugi, Lichtenberger, Jones, Lai, & St. George, 2000)

¹⁵⁰ (Blair, 1996)

¹⁵¹ (Blair, 1995)

¹⁵² (Jack & Robbins, 2012)

Robbins and Jack's innovation in connection with the mind/body problem is the claim that gap intuitions do not result from a conflict between neural networks/cognitive sub-systems keyed toward entities with and without minds. Instead, Robbins and Jack claim that a more specific conflict between the networks/systems that are respectively sensitive to mindless objects and experiencing subjects is the culprit. They claim that concerns about an entity's agency are less recalcitrant to physical construals than are concerns about an entity's patiency. This explains why persons feel moral concern toward other persons but not toward physical objects even when the physical objects are construed through the intentional stance.¹⁵³ An object being susceptible to agentic analysis does not distinguish it as a person deserving of full-fledged moral concern, but when an entity is taken to harbour experiential states that recruit the phenomenal stance, it is thereby endowed with moral status. The gap intuitions arise, thus, in order to alleviate the dissonance that results from the de-emphasis of a person's moral status that comes with considering a person as a physical system.

With respect to the tectonic theory, this is important because it shows that considering a person in her non-agentic aspects emphasizes her moral status, which partially explains why patientic considerations can have a mitigating effect on moral responsibility judgments as discussed above. This still leaves the question of why consideration of a person as an agent fails to deter such harsh decisions, which is important for explaining why there couldn't just be a more sophisticated form of agentic assessments in place of the two step structure in the tectonic theory. Though this question isn't specifically addressed by Robbins and Jack, their work provides an

¹⁵³ "Even when we attribute intentions to our chess computer, if we got bored with it then it wouldn't bother us to disassemble it and turn it into a toaster. Mental state attribution might therefore support action understanding on some level as we deal with computers or other inanimate entities, but simply taking an intentional stance or anthropomorphizing an entity does not necessarily imbue it with moral rights." (Therault & Young, 2014, pg. 107)

explanation in the way that they distinguish between the intentional and phenomenal stances. The intentional stance, they claim, constitutes a more distanced and explanatory form of social engagement, while the phenomenal stance constitutes a more empathic and affiliated form of social engagement.¹⁵⁴ This in turn is explained by the neural rivalry between the task-positive and default-mode networks in the brain.¹⁵⁵ According to Robbins and Jack, the intentional stance recruits both, which results in an outlook towards persons that is essentially a blend of two forms of neural activity that are primarily respectively associated with objects and persons. In contrast, the phenomenal stance is characterized as unblended, consisting just of neural activity in the default mode network.

To elaborate, Robbins and Jack's view is that while the intentional stance does involve activity in the default-mode network, which is associated with social cognition, the usual upshot of such activity (empathic engagement) is significantly adulterated by the way it is blended with activity in the task-positive network, which is highly analytic in comparison. In more everyday terms, one could say that the both networks are recruited when an explanation of a person's behaviour is being sought, while the default-mode network alone is recruited when empathic understanding of a person's personal circumstances is being sought. This explains why when trying to explain or

¹⁵⁴ "This second cognitive mode [taking the phenomenal stance] is the default mode for unguarded social interactions, in particular between a parent and child but also more generally for in-group members—in other words, for anyone whom we humanize ... However, it is not the only mode for social interaction. A more emotionally distanced mode of social interaction, corresponding to the intentional stance, involves a blend of these two cognitive modes which are naturally opposed to one another." (Jack, Robbins, Friedman, & Myers, 2014, pg. 130-131)

¹⁵⁵ "There are dissociable brain networks specialized for thinking about an entity's physical (e.g., causal-mechanical) attributes and for thinking about an entity's subjective mental life. The first network is activated when we take up the "physical stance" toward the entity, the second when we take up the "phenomenal stance." The network architecture is configured in such a way that the engagement of one stance inhibits the engagement of the other. Alongside these mutually antagonistic stances is a third stance that is compatible with both, namely, the "intentional stance," which is engaged when reasoning about an entity's goal-directed behaviors. Of these three stances, we submit that one of them in particular—the phenomenal stance—plays a distinctive role in moral reasoning, as this is the stance we step into when thinking about an entity as a target of moral concern." (ibid, pg. 129)

predict a person's behaviour using the intentional stance, there is little to no need for invoking any factors beyond her control. Robbins and Jack's work, thus, provides a neurologically informed explanation for why what Reader calls the agentic bias is particularly acute in the context of moral responsibility.¹⁵⁶ Since in order to categorize a person as a suitable candidate for moral responsibility she must be recognized as an agent, moral responsibility judgments typically begin with the intentional stance turning attention away from the patientic features of a person and her life. The tectonic theory's concern for patiency, thus, is compensatory with respect to this anatomically rooted shortcoming of the intentional stance.

However, compensating for the agentic bias is not as simple as just following up the intentional stance by taking the phenomenal stance. This qualification is required because of the way that Robbins and Jack characterize the phenomenal stance as unblended neural activity, transpiring wholly in the default mode network. I think it would have been more perspicuous for Robbins and Jack to grant that the phenomenal stance can involve blended activity from the task-positive and default mode networks just like the intentional stance, and then to use their concept of primary cognitive mode to distinguish between the two. Instead, they appear committed to drawing the distinction on the basis that the phenomenal stance is unblended. As will be discussed below, this can be little more than a rhetorical strategy because as they have it, the phenomenal stance only describes the reception of certain stimuli that generate empathic engagement, and fails to countenance any reflective thought or judgment about the stimuli.

¹⁵⁶ There remains an interesting question as to why the agentic bias would be so severe outside the realm of moral responsibility judgments, when philosophers just engage the concept of the person. An interesting answer to this question is suggested by Sneddon's work linking the nature of an action to moral responsibility (Sneddon, 2006). If the concept of the person is aptly characterized as the sort of being capable of actions, as opposed to just doings, and Sneddon is right about moral responsibility being partially constitutive of the concept of action, this would explain why the agentic bias is present in theorizing about persons in general, as opposed to just moral responsibility.

Rather than police Robbins and Jack's use of their own term, I proceed by clarifying how considering patiency differs from taking the phenomenal stance.

A close look at the empirical studies from which Robbins and Jack draw their contrast shows that just taking the phenomenal stance cannot be sufficient for carefully assessing whether a person's patientic features mitigate her moral responsibility. Instead, it is clear that if adopting the phenomenal stance is to have any import in moral responsibility judgments, it will need to involve some degree of cognitive alliance between the neural rivals of the default mode and task-positive networks. In other words, attending to a person's patiency will require a blended cognitive mode, akin to the intentional stance, though plausibly blended differently.

Reviewing evidence from a number of studies on the distinct neural underpinnings of intuitive physics and intuitive psychology, Jack and colleagues conclude that even when these usually antagonistic neural networks are both recruited by a particular task, one or the other will typically dominate.¹⁵⁷ Jack et al. appeal to the cognitive specialty of whichever network is dominant to describe the "primary cognitive mode" induced by a given task. Tasks are thus characterizable as primarily analytic or primarily empathic. Primarily analytic tasks can involve blends of neural activity within the task positive and default mode network with the former being dominant, while primarily empathic tasks are dominated by activity in the latter.

Robbins and Jack provide evidence for the claim that taking the intentional stance typically

¹⁵⁷ "These findings suggest a model of functional organization in which cognitive domain also plays a major role in determining how brain areas are recruited. It appears that there are fundamental differences in the information processing resources that are recruited, depending on *the primary cognitive mode* engaged by the task context." (Jack et al., 2013. pg. 11, italics added)

involves an analytic as opposed to empathic primary cognitive mode and vice versa.¹⁵⁸ This is highly relevant for explaining the value of the two step structure in the tectonic theory. It explains why patientic considerations are not suited to being taken into account when deciding whether a person is sufficiently agentic to be held morally responsible (the first step of tectonic moral responsibility judgments), which in turn explains why the compensatory second step is required to enhance the fairness of moral responsibility judgments. To determine whether a person is indeed an agent, one applies the intentional stance, which recruits the task-positive network, which makes the primary cognitive mode of agent-categorizations analytic as opposed to empathic. The human nervous system apparently allows for the demarcation of agents from non-agents without having to grant moral significance to the agents.¹⁵⁹

The tectonic theory's plea for concern with persons' patiency provides a principle that can countervail the tendency of purely agency-focused moral responsibility judgments to neglect non-agentic aspects of a person. Robbins and Jack's work suggests that this involves taking the phenomenal stance toward subjects of moral responsibility judgments, but their characterization of this as a neurally unblended form of cognition is problematic. As already mentioned, the process of considering how patientic factors should or should not mitigate a person's moral responsibility is not just a matter of taking an empathic perspective towards her. The second step of the tectonic theory is intended to be a deliberative philosophical judgment with idiosyncrasies

¹⁵⁸ The relevant studies concern dehumanization: "Our recent findings suggest that humanizing depictions evoke the neural signature of the phenomenal stance, whereas animalistic dehumanizing depictions, which involve a socially distanced representation of the emotional states of others, evoke the intentional stance." (Jack et al. 2014, pg. 160-161)

¹⁵⁹ Although this configuration appears to underlie the regrettable agentic bias discussed by Reader, it is worth noting that it likely has (or at least had) survival value for humans and thus would have been selected for. It is plausibly very useful for humans to be able to think of various things as agents without thereby inducing feelings of affiliation or concern towards them. Such a capacity would allow for ancient hominids to develop rudimentary understandings of animals, plants and other environmental features without either having to explicitly reason about their regularities or feeling hesitation about destroying or manipulating them.

intruding where needed. It is highly implausible that this could be done without contributions from the more analytically capable task positive network, whose activation appears to be implicated in abstract thinking of this kind. It is thus necessary to distinguish tectonic considerations of patiency from taking the phenomenal stance towards a person.

Robbins and Jack distinguish the phenomenal stance from the intentional stance in that it does not recruit (but rather diminishes) activity in the task-positive network. It needs to be emphasized that the suppression of task-positive activity is partial. Robbins and Jack never claim that the task-positive network is shut down during the phenomenal stance, but just that its activity is reduced. Additionally, the inverse relatedness of activity in the two networks is short-lived whenever the investigation involves a judgment or task, as opposed to just the forming of representations. These considerations suggest that Robbins and Jack's characterization of the phenomenal stance as usefully unconcerned with analytic reasoning appears to be an essentially rhetorical move keyed toward emphasizing the relative moral importance of empathic as opposed to analytic cognition. However, if it is important that the empathic engagement and affiliated concern toward a person yielded by the phenomenal stance play a role in moral thinking or rational action, as opposed to just providing a context of concern, then it will be necessary that the analysis oriented task-positive network gets involved.

Jack et al. essentially make this concession in their more empirically focused work. They note that when it comes time for fMRI subjects to judge or respond to a stimulus, the activation patterns that indicate antagonism between the task positive and default mode networks cease and

become more like the blend of the intentional stance.¹⁶⁰ Their explanation for this echoes my speculations about the task positive network being required to operationalize the contents wrought by the phenomenal stance within a moral responsibility judgment. Jack et al. write, “co-activation of DMN and TPN regions ... may reflect the communication of information represented in the DMN to areas involved in decision making and response. Mutual suppression is seen when the only task required of participants is to use perceptual information to build a representation.”¹⁶¹

While I agree with Robbins and Jack that the phenomenal stance is particularly important for moral thinking, I do not think that this is due simply to the way it enhances empathic engagement. In contrast, I propose that a degree of empathic concern and consideration of a person’s patiency happen to come as a neurological package. The crucial claims from Robbins and Jack for the tectonic theory are that (1) the intentional stance can easily fall short of generating moral concern for its targets, which explains why agency-focused moral responsibility judgments can be profitably supplemented with patientic considerations, and (2) in order to consider a person as a patient, one must step outside of the intentional stance. The resultant perspective is akin to, but not identical with Robbins and Jack’s concept of the phenomenal stance. Because Robbins and Jack define the phenomenal stance as consisting just of neural activity in the empathic default-mode network, it is necessary to characterize the cognition involved in phenomenal stance-taking as something like an input to patientic-considerations. This allows for the enhanced empathy of phenomenal stance-taking to figure in

¹⁶⁰ “There is a clear pattern of activation [of the default mode network] and deactivation [of the task positive network] throughout the stimulus presentation period. However, this pattern is disrupted during the question/response period. During this time period, we observe a pattern which is closer to what has typically been observed i.e. the TPN is not deactivated, even for conditions which activate the DMN.” (Jack et al., 2013, pg. 12)

¹⁶¹ (Ibid)

but not exhaust patientcy considerations.

Conclusion

This chapter presents work from two research groups, Gray, Wegner, and colleagues and Robbins, Jack, and colleagues. I incorporate aspects of their behavioural and neuroscientific findings into my account of how patientcy can figure in moral responsibility, as well as why moral responsibility judgments in the wild, so to speak, tend to neglect persons' patientcy. I drew from Gray and Wegner an account of the different contributions that agentic and patientic considerations make to thinking about a person. Their work strongly supports the division of labour hypothesis, which claims that moral responsibility is conferred when a person is appraised in her agentic features, and moral status or concern results from considering a person's patientic features. Gray and Wegner's early work showed that these sorts of considerations are cognitively independent from one another, meaning that thinking about a person's agentic features neither requires nor typically compels consideration of her patientic features.

I paid particular attention to Gray and Wegner's thesis of dyadic morality, which alleges that the basic cognitive template according to which moral transgressions are discerned includes an agent intending to harm another person, who must be seen as patientic insofar as she is capable of experiencing the intended harm. Finding Strickland et al.'s response to this work persuasive, I concluded that the agent/patient distinction plays a much larger role in human cognition than is captured by the dyadic morality thesis. This is important for the tectonic theory because it allows the patientic considerations recommended therein to range over factors aside from the moral

wrongs a person has experienced. I also presented Gray and Wegner's work on moral typecasting, which describes an inverse correlation between agentic and patientic considerations of people. Taking some cues from Arico's criticism of these studies, I concluded that the most that can be said of moral typecasting is that looking into a person's agentic and patientic features should occur one at a time, which explains the value of the tectonic theory's two step structure for assessing a person's moral responsibility.

I then looked at Robbins and Jack's similar work, focusing on their neurological claims. I presented their work as an explanation for why the tectonic theory's two step structure is needed to sensitize moral responsibility judgments to patientic considerations. Robbins and Jack theorize that the human brain includes disparate systems for handling different sorts of cognition. There is the object-oriented and thus primarily analytic task positive network, and the person-focused and thus primarily empathic default mode network, and most of the time high levels of neural activity in either of these correlates with low levels in the other. This distinction allowed Robbins and Jack to distinguish between the different sorts of cognitive content that result from looking into persons' agency and patiency. Assessing a person as an agent involves taking what Dennett called the intentional stance, which Robbins and Jack contrast to what they call the phenomenal stance. They characterize the intentional stance as a colder, more abstract and distanced form of social cognition, which they take to be explained by the way that it recruits both the task-positive and default-mode network. The phenomenal stance, in contrast, is characterized as the crucial cognitive element of moral cognition because it engenders empathic concern towards its targets in virtue of the fact that it processes other persons' experiences.

Roughly, taking the intentional stance and the phenomenal stance correspond respectively to considering a person as an agent or a patient. However, some qualifications were necessary. First, I suggested that in the context of moral responsibility judgments, the upshot of taking the phenomenal stance toward a person is not just a matter of inducing concern toward her. Although this is important, I don't think it is the central factor. The main reason that taking the phenomenal stance toward a person is useful in the context of moral responsibility judgments is that it directs focus away from their agency, thus allowing for consideration of their patiency. Second, given the way that Robbins and Jack defined the neural correlate of the phenomenal stance to be enhanced activity in the default-mode network concurrent with suppressed activity in the task-positive network, it was necessary to qualify phenomenal stance-taking as crucial just to providing the inputs to considering a person's patiency, and not ultimately identical with such considerations. This is due to the fact that the analytically inclined task positive network will need to be recruited if patientic features of a person discerned by taking the phenomenal stance are to figure in the effortful reflection about such features that the tectonic theory prescribes.

Chapter four: The Homogenization Problem

“Unfortunately, activating the fairness intuition is not as easy as firing up the retributive impulse.”

(Bruce Waller)¹⁶²

In this chapter, I present Waller’s critique of the moral responsibility system, explain how to adapt his criticisms to a non-eliminativist context, and argue that a concern for patiency can serve to address the relevant problems. Waller’s case against moral responsibility is unique insofar as he accepts that a naturalistic account of free will can illuminate the way people are free to make choices from arrays of alternatives despite not having anything like a libertarian power to make such choices in an uncaused way. However, Waller still maintains that attributions of moral responsibility are fundamentally unfair so long as a naturalistic world-view is accepted.

The unfairness at issue, then, is not the usual sort. Waller does not present the typical incompatibilist claim that since persons lack control over themselves on account of being physical entities, it would be unfair to reward or punish them. Instead, Waller argues that moral responsibility judgments are inherently insensitive to important nuances of particular cases. As I read Waller, the unfairness at issue consists of a lack of accountability to detail. Of course, compatibilist theories of moral responsibility are not insensitive to all the details of particular cases. By emphasizing things like reasons-responsiveness and rationality, contemporary theories of moral responsibility incorporate relatively nuanced accounts of how it is that people enjoy a measure of control over themselves. However, Waller insists that acceptance of a naturalistic world-view entails commitment to a mass of uncontrolled factors serving to shape any given

¹⁶² (Waller, 2011, pg. 313)

individual's capacity for rationality or reasons-responsiveness, and these are the details about particular individuals that compatibilist theories neglect. Waller believes that the unfairness of moral responsibility becomes clear when sufficient attention is paid to them.

There remains the question of how to get people to pay attention to these factors. As we saw in the previous chapter, there appear to be neurological facts that limit a human's ability to focus on a person's agency and patiency at the same time. Waller is thus in need of some machinery to motivate consideration of the uncontrolled factors in a person's life. To a large extent, this is supplied by Waller's over-arching eliminativism regarding moral responsibility.¹⁶³ In the context of aiming to provide as strong a case as he can against moral responsibility, Waller's call to attend to the uncontrolled factors that shape individuals' capacity for control over themselves is well motivated and intelligible. However, there is a major problem with the case against moral responsibility, one which Waller acknowledges directly in his response to Dennett in a discussion at naturalism.org.¹⁶⁴ The problem is that moral responsibility practices secure many goods for society that do not seem to be otherwise available such as rich communities and social order. In other words, Waller concedes both that a viable eliminativism would require some alternative to moral responsibility and that none is forthcoming.¹⁶⁵

Even though Waller is an unwavering eliminativist, there are local details in his argumentation that hint at a non-eliminativist way to incentivize looking more closely at the uncontrolled aspects of individuals' lives. Because he accepts that a compatibilist form of free will exists,

¹⁶³ "My goal is to kill [moral responsibility] and drive a stake through its heart." (Waller 2011, pg 2)

¹⁶⁴ (Waller, 2012)

¹⁶⁵ Waller writes, "Dennett saves his toughest challenge for last. So you don't think this "fair enough" system is acceptable? Fine; show us a better one. ... I should be honest: I don't have such a system to offer." (Ibid)

Waller's eliminativism does not hinge on the claim that nobody has sufficient control over themselves to be held morally responsible. Instead, Waller develops *the comparative unfairness* argument. He describes the argument as a 'variation on a very old theme,'¹⁶⁶ but I think there is important novelty in it. Waller's goal is to motivate people to be mindful of the contexts that shape individual's agentic capacities. His concern, thus, is mainly with trait-based patiency, which betrays a degree of residual inheritance of conventionally incompatibilist motifs.

As we will see, Waller's appeals to uncontrolled factors in people's lives are all cases in which a person's natural endowments, upbringing, or acculturation have significant downstream effects on their agency. Because he is an eliminativist, Waller aims to trouble the moral responsibility advocate's reliance on agency by illuminating the ways that a person's agency is inevitably developed in consultation with a massive collection of uncontrollable factors. The idea seems to be to show that the forms of agency commonly taken to legitimize moral responsibility are not as pure as they need to be to ground moral responsibility practices. Rather like Smilansky, Waller's argumentation is in part derived from the idea that libertarian free will would legitimize moral responsibility if it were real. His emphasis on the uncontrolled influences on people's developing agency thus functions as something like a bridge from scepticism about libertarian free will to scepticism about forms of moral responsibility tied to compatibilist accounts of agency.

Waller is interested in ensuring that his readers don't slip back into libertarian intuitions, and as I read him, his discussion of uncontrolled factors left out of account by agency-focused moral responsibility practices is generally keyed to prevent this. However, even though Waller is abidingly concerned to keep libertarian intuitions from returning to the table, his way of

¹⁶⁶ (Waller 2011, pg. 24)

preventing this contains interesting innovations, whose import can be decoupled from the goal of staving off libertarian beliefs. Waller's main way of exposing the unfairness of moral responsibility practices is to emphasize the way that agentic capacities are unevenly distributed across the class of agents, and then show how existing theories of moral responsibility are unable to take this diversity into account. I call this his *counterpart tactic*.

I part from Waller's view insofar as I do not insist that the details about persons neglected by moral responsibility judgments are significant because they indicate the attenuatedness of persons' agency. My view is that these details are important because they illuminate what people do not control, which when construed through the concept of patiency and allied with Reader's reflections about the nature of persons, acquires a value of its own that is not dependent on undermining agentic considerations. This allows me to preserve the basic institution of moral responsibility on the basis that people have sufficient control over their behaviour to be held morally responsible, while simultaneously claiming that moral responsibility practices unfairly neglect nuances about agents. I think this preserves the most interesting element of Waller's argumentation, the claim that moral responsibility practices are unfair in failing to take into account important differences between agents, but allows for this insight to motivate the more sophisticated moral responsibility judgments prescribed by the tectonic theory instead of moral responsibility eliminativism.

The Counterpart Tactic

Waller's way of compelling his readers to be lucid about the details that moral responsibility

judgments typically leave out of account is to fabricate a counterpart to any proposed example of morally responsible behaviour with the counterpart always being subject to different uncontrolled factors. Here's a brief summary of one pair of counterparts: Karen and Louise are both facing an opportunity to oppose a racist hiring practice at work. Karen succeeds and Louise fails. Waller comments, "If Karen tries harder, thinks more effectively, deliberates more thoroughly, or empathizes more deeply than Louise, then Karen's superior powers (like Louise's inferior qualities) resulted from causes far beyond her control ... Likewise, the inferior fortitude and commitment of Louise (which led to meager or abortive efforts toward self-improvement) were due to forces that were not under her control."¹⁶⁷ Waller's counterparts are always comparable as agents in that both are reasons-responsive and rational etc., but Waller insists that there are significant differences between them. These differences consistently involve patientic factors, though Waller does not use the concept of patienty, instead describing the relevant factors as beyond a person's control.

Waller does not explain why his counterpart-tactic functions as a reminder of the uncontrolled forces that shape individuals as agents, and he occasionally presents single cases that seem to serve the same purpose adequately.¹⁶⁸ Throughout, Waller's expectation is that the examples will

¹⁶⁷ (Waller 2011, pg. 25) In addition to Karen and Louise, there are: Betty and Benji pg. 29 & 282, Betty and Barbara, pg. 34, Ann and Barbara, pg. 89 & 228, Jane and Judith, pg. 90, Jane and Joe, pg. 93, Ed and Ann, pg. 105, Jan and Kate pg. 117, George and Harsham, pg. 119, Arnold and Bryan, pg. 121, Louise and Martha, pg. 122, Isaac and Jeremiah, pg. 125, Robert and Sean, pg. 128, Tamara and Veronica, pg. 130, Luke and Matthew, pg. 131, Jill and Laura, pg. 136, Julia and Kate pg. 147. Beth and Carolyn, pg. 156, Carl and David, pg. 160 & 168, Jake and Joan, pg. 165, David and Robert, pg. 174, Donna and Matthew pg. 200, Monica and Patricia, 214, and Maria and Wanda 223.

¹⁶⁸ See Jamal, pg. 60, Eve, pg. 61 & 154, Robert, pg. 153. There are a few more cases of single-case examples in Waller's book: Rita, pg. 185, Sam, pg. 186 Alice, pg. 193, and Rafia, pg. 216, but these are all presented by Waller to head-off worries about what else might be lost if we eliminate moral responsibility. Rita and Sam are supposed to show that judgments about character flaws can be retained even without moral responsibility. The case of Alice is keyed to show that moral responsibility eliminativism does not undermine the practice of apologizing, and Rafia is presented as a virtuous person whose character we can continue to judge as morally good so long as excuse-

stimulate judgments of unfairness. More specifically, Waller expects his readers to see the unfairness involved in doling out praise/blame or reward/punishment to any of his examples, whether the paired contrasts or the singular examples. The precise mechanics of the unfairness judgments are not important for Waller because of his over-arching eliminativism. In order to see how patency figures in Waller's argumentation, though, a closer look at his counterpart tactic is required. A clear sense of how patient concerns give rise to the unfairness judgments Waller pursues will in turn begin to chart a non-eliminativist adaptation of his criticisms of moral responsibility.

To see what Waller gains from his counterpart strategy, it is useful to look at the single cases to which he does not add a counterpart. There are three examples: Jamal, Eve, and Robert, all of which figure in Waller's criticism of Frankfurt's *hierarchical authenticity*, previously discussed under the name of identificationism. It is important to note that Waller takes Frankfurt's reflections about hierarchical desires and the importance of identifying with the springs of one's actions to serve as a cogent analysis of a real and significant phenomenon.¹⁶⁹ Waller's critique of Frankfurt's view is concerned specifically with the prospects of legitimizing moral responsibility on an identificationist basis. The examples of Jamal, Eve and Robert are presented to show that identificationism is problematic in this respect. I will discuss Jamal and Eve in some detail to provide a preliminary sense of what the counterpart tactic accomplishes.

extensionism is avoided, as Waller counsels. None of these latter four examples are designed to stimulate the reader's sensitivity to uncontrolled factors, and so I will not engage them here.

¹⁶⁹ More specifically, Waller sees hierarchical identification as a natural bulwark against the possibility of an organism behaving in such a way that varies too much to be advantageous. He writes, "Why then does the concept of an "authentic true path" have such undeniable appeal as an account of human freedom, ... It is beneficial to occasionally pursue new paths, even when we know that our currently advantageous path offers more consistent rewards. But it is also important not to abandon useful paths too easily. Our commitment to authenticity keeps old paths open and available, and thus enlarges our options." (Waller 2011, pg. 67)

Jamal is a formerly independent slave who after a period of resistance begins to acquiesce to his plight. Eventually, Jamal ceases to resist and “loses any desire for freedom: now he wishes only to serve his master faithfully, is glad that he is a slave, and reflectively approves of his enslavement. And now Jamal is free; he has free will and is morally responsible for his own slavery. Or so we must conclude, if we accept Frankfurt’s model.”¹⁷⁰ The case of Eve is parallel. She is born into a deeply religious society that fiercely subjugates women and lives in opposition to these norms for a time. But “Eve eventually succumbs, loses her desire for freedom and independence, and embraces her role as meek and subservient and her status as inferior. She now reflectively approves of her subservient role.”¹⁷¹ Again, if we agree with Frankfurt that hierarchical identification is sufficient for moral responsibility, we should find Eve to be morally responsible for her own subjugation.¹⁷²

Waller sees these cases as straightforward counter-examples to the viability of basing moral responsibility on a hierarchical model. I will not adjudicate this claim. My interest is just in the nature of the patientic factors at play and why neither Jamal nor Eve is paired with a contrasting counterpart as appears to be Waller’s custom. Both Jamal and Eve give up their commendable opposition to prevailing social practices as a result of efforts made by other people in their lives. Waller writes of Jamal that “every attempt at escape results in heavier chains and greater suffering, every act of defiance is severely punished,”¹⁷³ and of Eve, “years of struggle bring a

¹⁷⁰ (Waller 2011, pg. 60)

¹⁷¹ (Ibid, pg. 61)

¹⁷² The case of Robert is similar to those of Jamal and Eve. Robert is a formerly unwilling drug addict who eventually wears down and begins to reflectively endorse his addiction. This case can be treated more or less like those of Jamal and Eve, making allowances for the difference between Jamal and Eve’s acquiescence coming at the hands of other people, and Robert’s being due to his drug. Of course, it would be worth making a special case of addiction, but I leave this for future work because it is not essential to the claims I need to make in this thesis.

¹⁷³ (Waller 2011, pg. 60)

steady stream of failure, condemnation, and psychological and physical abuse.”¹⁷⁴ It is clear that Waller intends that Jamal and Eve’s respective transitions from opposing to endorsing their conditions are caused by the violent disapprobation they faced.

In this respect, Jamal and Eve are the clearest possible examples of patients because they are obviously victims of others’ deplorable behaviours. As discussed in chapter two, there is little difficulty in stimulating sensitivity to varieties of patiency involving clear victims of other agents. The fates of Jamal and Eve are on a par with cases of brainwashing and manipulation at the hands of evil neuroscientists, which are routinely presented as clear-cut defeaters of moral responsibility.¹⁷⁵ Cases of extreme manipulation like this do not need to be explicitly presented as examples of patiency because the mitigation of moral responsibility at work can easily be captured in terms of agency, namely, the agency of the manipulators. However, as was shown with Reader’s work, the concept of patiency ranges over a much greater range of phenomena than is present in such examples. One could even say that the agentic bias accounts for the way that when patiency does show up in discussions of moral responsibility, it is typically a form of interpersonal patiency.

The reason Waller usually adds a contrasting counterpart to his examples is that the subtler aspects of patiency are not as easily attuned to. Factors such as an individual’s sense of self-efficacy or the extent to which deliberative thinking is pleasurable can also make a difference to judgments of moral responsibility but they are not visible on the surface of examples in the same way that the blatant patiency of being subjugated to others is in manipulation and indoctrination

¹⁷⁴ (Waller, 2011, pg. 61)

¹⁷⁵ e.g. (Greene & Cohen, 2004), (Pereboom, 2013)

cases. To adapt Waller's critique of moral responsibility to a non-eliminativist context, some principle that motivates a concern for these subtler aspects of patiency is required and I believe that such a principle is already implied by Waller's counterpart tactic.

To extract a principle that motivates concern for a fuller suite of patientic factors, I will focus on Waller's discussion of Betty and Benji, and the way that Waller construes his comparative unfairness argument as an update of Galen Strawson's infinite regress argument against moral responsibility.¹⁷⁶ Betty is actually an example Alfred Mele constructed in response to G. Strawson.¹⁷⁷ Mele's view is that G. Strawson sets a needlessly high bar for moral responsibility. It is easy enough to construct what appear to be clear examples of intentional self-modification that intuitively merit moral responsibility even if they do not meet Strawson's criteria. As we will see, Betty is one such example, and Mele argues that if subscribing to a given notion of moral responsibility results in a failure to classify an example like Betty as morally responsible, this just speaks against the plausibility of the notion of moral responsibility in play. Waller's comparative unfairness argument is presented as a naturalistic update of Strawson's case that is supposed to eliminate its vulnerability to counter-examples like Mele's Betty.

There are two ways in which Waller modifies Strawson's argument. First, Waller swaps Strawson's regress for a naturalized account of how individuals develop agency. Instead of emphasizing the conceptual impossibility of actions originating entirely from an individual,

¹⁷⁶ Strawson's regress argument is roughly as follows: since what one does is a product of the way one is, moral responsibility for what one does requires moral responsibility for the way one is, but in order to be responsible for the way one is, one must have chosen to be that way, which requires that one possessed principles of choice, which must have themselves been chosen according to some principles, which must also have been chosen according to some principles ... and so on. (G. Strawson 1986, pg. 28)

¹⁷⁷ (Mele, 1995)

Waller discusses a number of results from psychological studies which speak to the way that individuals naturally vary in the resources from which they can draw in developing their agency.¹⁷⁸ The second modification is Waller's counterpart strategy, which is meant to focus attention on the uncontrolled factors that Waller provides in place of Strawson's conceptual regress. Waller does not make clear, though, how or why this works.

As I read Waller, the first modification of Strawson's argument is a burden shifting manoeuvre. He aims to put proponents of moral responsibility on their heels by calling attention to the fact that each of us develops our share of self-control in consultation with a host of uncontrolled factors, such as genetic endowment and early formative experiences. This move is aimed at something of a straw man, however. Most contemporary proponents of moral responsibility would accept that an individual's moral capacities do not originate wholly from her efforts.¹⁷⁹ The counterpart strategy, which emphasizes the downstream consequences of variation in resources related to the development of agency, on the other hand, is a direct challenge to the fairness (as opposed to the metaphysical legitimacy) of extant accounts of moral responsibility.

Mele's Betty example is meant to stimulate an intuitive response to the effect that Betty merits moral responsibility for her efforts at self-modification. Indeed, it is deeply counter-intuitive to withhold praise from Betty. Betty is a six year old who resolves to overcome her childish fear of

¹⁷⁸ Waller mentions the distinctions between individuals that are chronic cognizers and those that are cognitive misers, and individuals with either a strong or weak sense of internal locus of control. Then Waller explains how these variations can give rise to significant differences in an individual's sense of cognitive fortitude and self-efficacy, which in turn explain how individuals vary in their capacity for moral behavior in ways that are invisible if we remain just concerned with compatibilist criteria for moral responsibility such as reasons-responsiveness, or rationality.

¹⁷⁹ It is plausible that Waller has the views of non-specialists in mind when he emphasizes that uncontrolled factors figure in the development of individuals' self-control. His appeal to the fact that people vary in their capacities for moral behaviour, on the other hand, seems aimed squarely at contemporary philosophical views such as Mele's.

the basement, gradually desensitizing herself to its dark ambiance by forcing herself to visit. Like P.F. Strawson, Mele contends that considerations about whether Betty's efforts are uncaused or due entirely to efforts original to her just do not figure in our inclination to assign credit to her.

It is this sort of broadly (P.F.) Strawsonian inclination that Waller's counterpart strategy is supposed to frustrate. Waller writes,

Place alongside Betty her six-year-old twin brother, Benji, who also suffers from fear of his basement (and who, like Betty, knows that no harm has befallen those who venture there). Benji also regards this fear as "childish" and wishes to get beyond it. But Benji is a little — just a little — less self-confident than his sister. Rather than taking bold steps to deal with his fear, Benji decides to wait it out: maybe I'll grow a bit bolder as I grow older, Benji thinks; ... there's no need for me to make an effort that might well fail ... Benji is not quite as strong as his sister, in some very significant respects. He does not have her high level of self-confidence (or sense of self-efficacy); his sister has a strong internal locus-of-control, but Benji is inclined to see the locus-of-control residing in powerful others.¹⁸⁰

Waller explains that Benji "chooses and acts freely, though without the rich freedom resources enjoyed by Betty. And ... those differing resources — though to a significant degree self-made — were shaped by initial resources that [they] did not choose to have."¹⁸¹ Waller elaborates the tale of Betty and Benji, imagining them as adults living in a racist society. Betty is now a civil rights activist who is deeply opposed to the prevailing orthodoxy while Benji is an acquiescent racist. Waller writes, "When we look carefully at the differences in their developed characters, we recognize ... that those differences were the product of early differences in capacities or in circumstances which they did not control and for which they are not morally responsible ... It is unfair to reward one and punish the other, or praise one and blame the other; that is, it is unfair to treat them in dramatically different ways. Their characters and behavior are their own, but that

¹⁸⁰ (Waller 2011, pg. 30)

¹⁸¹ (Waller 2011, pg. 33)

does not make Betty and Benji morally responsible.”¹⁸²

Waller’s counterpart strategy involves two ideas that can come apart. First, Waller emphasizes that Betty and Benji are shaped by uncontrolled forces, but this claim can be made of either in isolation. Second, Waller emphasizes the differences in capacity that result, which requires the contrast to be made clear. I think that the first of these (an emphasis on uncontrolled factors) is meant to induce an abstract skepticism toward moral responsibility. However, following (P.F.) Strawson and Mele, our actual inclinations to attribute moral responsibility to individuals are not troubled by this sort of reflection. In contrast, the second element of Waller’s counterpart strategy (the emphasis on diversity of capacity across agents) does seem to make contact with intuitive thinking about moral responsibility. Waller says of Betty and Benji that is unfair to treat them in dramatically different ways, and I argue that this claim can survive on its own without the support of a generalized skepticism regarding moral responsibility.

Instead of putting patientic considerations (uncontrolled factors) into service casting general doubt on the enterprise of moral responsibility as Waller does, one can appeal to patientic just as a way to illustrate significant diversity across individuals that are comparable in their agency. Although I would like to think that just encountering Reader’s arguments that personhood is as much comprised by patientic as it is agency would be sufficient reason to include patientic considerations in moral responsibility judgments, I have my doubts because of the ultimately neurologically caused agentic bias. I think it is therefore important to motivate the importance of patientic factors based on the features of moral responsibility judgments as well. The idea is to open up a second front in the assault on the agentic bias, arguing for patientic considerations in

¹⁸² (Waller, 2011, pg. 35)

both a bottom up (based on reflections about persons) and top down way (based on reflections about moral responsibility judgments). I therefore claim that the unfairness of moral responsibility judgments results from the way that they inevitably elide patientic details from particular cases in favour of a one size fits all agency. I call this *the homogenization problem*.

The Homogenization Problem as a non-eliminativist critique of moral responsibility judgments

Agency is something that morally responsible individuals share by definition. A purely agentic focus, consequently, is an impediment to accessing the distinguishing details of particular individuals, which Waller's counterpart tactic seems calibrated to overcome. If we remain just concerned with whether the person we are morally judging is a legitimate agent, many of the details by which she may be distinguished from other agents are obscured. However, if we first concern ourselves with whether or not the individual being judged is an agent, and subsequently consider her as a patient, we have gained resources with which to make moral thinking sensitive to both what an individual has in common with other individuals as well as what might be distinctive about her. This is how the tectonic theory can enhance the fairness of moral responsibility judgments.

According to Waller, the homogenization problem results from the threshold-like structure of agent determinations, which he describes as the 'moral responsibility plateau.' In a chapter called 'Begging the Question for Moral Responsibility,'¹⁸³ Waller sets out to reject a number of arguments that purport to legitimize the moral responsibility system. The most important sections concern plateau arguments, which claim that "there is a general level of moral responsibility, and

¹⁸³ (Waller, 2011, ch. 12)

beyond that level, the individual differences have no effect on moral responsibility.”¹⁸⁴ Waller thinks that the moral responsibility system really does have a plateau structure along these lines (with which I agree) but insists that plateau arguments provide “a description of how the moral responsibility system works, not an argument to show that it is fair. And its fairness — which is the question at issue — is very doubtful.”¹⁸⁵

Appropriating the voice of a plateau theorist, Waller writes, “We don’t care what your history is: right now the playing field is level,” which he is quick to condemn (in his own voice): “That is indeed how moral responsibility works in the moral responsibility system, but that marks the deep flaws in that system, not its justification.”¹⁸⁶ I agree with all of this, but differ from Waller in thinking that the moral responsibility system is justified based on the socio-normative benefits it provides for societies even though it is unfair in this way. The tectonic theory is designed to address the homogenization problem while preserving the system as a whole.

Although Waller often emphasizes the importance of the fact that persons lack libertarian free will, and characterizes moral responsibility as dependent on such a notion,¹⁸⁷ he will eventually appear to rest his case on the unfairness due to the homogenization problem. Earlier in his book, Waller distinguishes his way of criticizing moral responsibility from *excuse-extensionism*, by presenting what he calls the *miracle-requisite* view of moral responsibility. Excuse-extensionism refers to an argumentative strategy that aims to show a lack of relevant differences between

¹⁸⁴ (Waller, 2011, pg. 228)

¹⁸⁵ (Waller, 2011, pg. 229)

¹⁸⁶ (Waller, 2011, pg. 230)

¹⁸⁷ “What is the basis for the universal denial of moral responsibility? It is basically that in order to establish moral responsibility, we need miracles: the power to be an unmoved mover, the power of contra-causal free will, the miraculous ability to make yourself ab initio.” (Waller, 2011, pg. 217) Waller also “rejects moral responsibility as a remnant of a shallow prescientific understanding of human psychology that endeavors to keep humans in a special unique moral status, apart from the natural world.”(ibid, pg. 215)

persons that intuition excuses from moral responsibility, and persons who are typically ascribed moral responsibility.¹⁸⁸ Dennett refers to this as “creeping exculpation,” which Waller concedes to be problematic:

If we start from the assumption that everyone is morally responsible and excuse them only when there is clear indication of debilitating flaws or someone else to blame, then “creeping exculpation” does indeed look scary: complete denial of our moral responsibility must be based on our utter incompetence. But that is a false model of the naturalistic denial of moral responsibility: such denial is not an excuse-extensionist model in which we start from moral responsibility and gradually erode it away through excuses; rather, it starts from the claim that we cannot make sense of moral responsibility within a naturalistic framework.¹⁸⁹

Waller’s alternative to the excuse-extensionist approach gives the absence of libertarian free will a central role. He writes, “In contrast to the excuse-extensionist model for universal denial of moral responsibility, consider (what we might call) the miracle-requisite model: moral responsibility requires a miraculous self-creation that is not compatible with naturalism.”¹⁹⁰

There is thus ample evidence that Waller takes the illegitimacy of moral responsibility to be due in part to the way that it insists on persons being miraculous in the ways supposed by libertarianism.

As discussed in chapter one in connection with Smilansky, the thinking behind this claim is that if persons did possess libertarian free will, then they would always be theoretically capable of outcompeting influences that are external to their agency. It is a corollary to this that since persons do not have libertarian free will, they are subject to such influences, i.e. persons’ patency does play enough of a role in their lives to merit consideration in the context of moral responsibility. Without the concept of patency, it is easy to interpret the homogenization problem

¹⁸⁸ (Greene & Cohen, 2004) is a good example of this approach, and will be covered in detail in chapter six.

¹⁸⁹ (Waller, 2011, pg. 145)

¹⁹⁰ (Waller, 2011, pg. 213)

as result of pervasive libertarian intuitions of the sort that naturalistic reflection ought to dispel. But the reflections of Reader, together with the work of Gray and Wegner and Robbins and Jack, provide an alternative explanation.

Moral responsibility judgments are primarily interested in persons' agentic features, which as a matter of human neurology, happen to be cognized independently from patientic considerations. Thus, incorporating patientic considerations into moral responsibility judgments requires a degree of effortful restructuring, as endorsed by the tectonic theory. This alternative explanation for the homogenization problem, I hope, makes it seem more tractable. Instead of having to combat the view that persons are usually sufficiently agentic to be held morally responsible for their behaviour in order to sensitize moral responsibility judgments to patientic considerations, all that is needed are reasons to attend to such factors. Reader's work provided one such motivation by explaining that doing justice to what persons are should involve patientic considerations. I think Waller's work on what I'm calling the homogenization problem provides another.

Additionally, my rejection of Waller's claim that moral responsibility requires miracles allows for a different interpretation of the claim that the unfairness of the moral responsibility system derives from its insensitivity to details about particular persons. The homogenization problem can serve as an effective criticism of the moral responsibility system, but no longer as an effective premise for moral responsibility eliminativism. Since I do not agree that libertarian justifications of moral responsibility practices are the only option (for reasons presented in chapter one), I am free to take the homogenization problem as a regrettable feature of moral

responsibility practices, but not sufficient reason for eliminativism.

The tectonic theory's inbuilt concern for patiency, defined as uncontrolled factors relevant to a person's behaviour, already distinguishes the tectonic theory of moral responsibility from libertarianism-requiring views that are Waller's main target. The question remaining is just whether it can help as a treatment for the homogenization problem.¹⁹¹ To make the case, I need to address Waller's view that advocating for the preservation of moral responsibility entails discouraging looking more closely at individuals in the ways that the tectonic theory prescribes.

Waller claims many times that the moral responsibility system discourages looking any more closely at individuals than is necessary for determining that they reach the moral responsibility plateau. He writes that "the moral responsibility plateau ... blocks any scrutiny of the actual capacities and abilities of those judged morally responsible and of the historical circumstances that shaped their differences."¹⁹² The adaptation of this criticism to a non-eliminativist position requires weakening the claim about the moral responsibility system *blocking* initiatives to look more closely at individuals' particularities. Given the discussions from the previous chapters, I think it is more accurate to say that moral responsibility judgments fail to incentivize looking more closely at persons, rather than blocking such queries.

This view is supported by the fact that it is no coincidence that the features of persons that Waller criticizes moral responsibility judgments for failing to take account of are patientic features, i.e. elements of a person and her life that are not subject to her control. The moral responsibility

¹⁹¹ I would like to emphasize the choice of the word 'treatment.' I do not contend that the tectonic theory is anything like a cure for the homogenization problem.

¹⁹² (Waller, 2011, pg. 231)

system makes essential use of the concept of the person, and the homogenization problem results from the way that no more is needed from the concept of a person aside from its agentic content in order for moral responsibility judgments to proceed. Since moral responsibility thinking is already deeply concerned with agentic facts about persons, the homogenization problem can be most effectively addressed by insisting on the consideration of non-agentic facts about a person. Referring to such factors as a person's patiency is important for two reasons. First, it motivates sensitivity to such factors on the basis of the claim that such factors constitute an important aspect of being a person. Second, it focuses attention squarely upon the factors that are most typically neglected in moral responsibility judgments.

My view is that the apparent tension between the relatively minimal content of granting that a person is an agent suitable for moral responsibility and the comparatively curious and compassionate act of looking more closely at persons' distinctive features results from the factors discussed in chapters two and three. In contrast, Waller thinks that it is a result of people having a vested interest in the continuation of the moral responsibility system and that the system would be imperilled to the extent that people are motivated to look more closely at individuals.¹⁹³ This aspect of Waller's view can be rejected along with his miracle-requisite account of moral responsibility because it assumes that libertarian intuitions are necessary for widespread acceptance of the moral responsibility system.

If Waller were right about libertarianism necessarily being the basis of people's beliefs in moral responsibility, then looking more closely at persons probably would spell death for the moral

¹⁹³ "In order to retain the retributive system, we cannot scrutinize the real problems that need reform because careful scrutiny is death to the moral responsibility beliefs that prop up retribution." (Waller 2011, pg. 297)

responsibility system because an honest assessment of persons in this day and age ought inevitably to reach the conclusion that persons are not miraculous in the ways supposed by libertarianism. Within a post-libertarian discussion, however, the tectonic theory's concern for patiency can be taken as a plausible route to enhancing the fairness of moral responsibility judgments by mitigating the homogenization problem, without being aligned with ambitions of undermining the whole moral responsibility system.

There is, however, another view worth addressing regarding the tension between the plateau-like structure of moral responsibility judgments and the tectonic theory's compensatory interest in looking more closely into the patientic particularities of individuals. In an exchange with Waller and Tom Clark at naturalism.org, Dennett argues that a degree of nuance-eroding homogenization is necessary for the moral responsibility system. However, Dennett makes these claims in response to Waller's eliminativist presentation of the importance of looking more closely at individuals than is strictly speaking required by moral responsibility judgments. Dennett's criticisms do not apply to my tectonic proposal.

As mentioned toward the beginning of this chapter, Dennett's major grievance with Waller's moral responsibility eliminativism is that it fails to provide a viable alternative to moral responsibility practices for securing social order. He does, however, for the most part agree with Waller regarding the unfairness of present-day justice systems,¹⁹⁴ but Dennett is nevertheless compelled to adopt an unpersuaded and somewhat dismissive tone in response to the worries Waller articulates concerning what I've called the homogenization problem. His commentary on

¹⁹⁴ While Dennett is describing a passage from Waller's book, he makes it very clear that extant justice systems (which Dennett agrees are dependent on the concept of moral responsibility) are problematic: "what follows is first four pages of criticism of the oh-so-unfair existing punitive system (everybody agrees about this)." (Dennett, 2012)

the counterpart tactic is a good example:

Waller gives us example after example in which he contrasts two people, one of whom had a salubrious upbringing while the other was deprived of most benefits. Is it fair, he keeps asking, to hold both of them responsible? Life isn't fair. Some folks get a pretty raw deal through no fault of their own, and others thrive thanks to a great head start. Let's face it: physical beauty gives a huge undeserved advantage to those who have it, and it is just one of a variety of unequal endowments that challenge any account of moral responsibility. None of this is fair.¹⁹⁵

Dennett's goal seems to be to alleviate concern about the homogenization problem on the basis that moral responsibility practices expectably inherit the unfairness of life in general. He thus defends the plateau conception of moral responsibility from a practical perspective: "there is an artificial and somewhat arbitrary plateau or threshold [for moral responsibility], and we don't waste effort examining differences between those who make the grade."¹⁹⁶ He compares Waller's plea to look more closely into the distinctive features of individuals to baseball policy that would "look at each struck ball very closely and stop this arbitrary distinction between hits and fly balls, dealing with each feat in all its natural, real properties."¹⁹⁷ Dennett also compares moral responsibility to drivers' licensing.¹⁹⁸ The point appears to be that even if there were some very mature fifteen year olds to whom it would not be problematic to offer a drivers' licence, in order for the drivers licensing system to work well, these sorts of exceptions must be off the table.

I think that Dennett's criticisms are a sensible response to Waller's claim that moral responsibility practices should be abolished because libertarian free will does not exist. However, the homogenization problem has been severed from this issue. I hope to have shown that moral responsibility judgments are systematically insensitive to uncontrolled factors, and that this

¹⁹⁵ (D. Dennett, 2012)

¹⁹⁶ (Ibid)

¹⁹⁷ (Ibid)

¹⁹⁸ "One of the few favorite points of mine that Waller doesn't discuss, is the rule that says you can't drive till you're sixteen years old in most jurisdictions. It's arbitrary, artifactual, and it deliberately ignores minor differences about the abilities and maturity of individual applicants for a driver's license. It's fair." (ibid)

characteristic can be explained on the basis of the agency/patiency distinction together with the empirical claim that these need to be considered separately to be effectively considered at all. Even if moral responsibility judgments are here to stay, about which Dennett and I agree, there is both room and reason for reconceptualising their basis.

Once we take Waller's criticisms out of their eliminativist context, his arguments for looking more closely at individuals can be construed as a call to make moral thinking sensitive to patiency – as a partial antidote to the homogenization problem instead of as an underminer of the very idea of moral responsibility-conferring agency. If we are faced with a choice between looking more closely at individual agents and keeping the moral responsibility system, the latter seems preferable. However, none of Dennett's criticisms of Waller speak against trying to make moral (and particularly legal) reasoning more responsive to individuals' patientic characteristics. Dennett's claim that moral responsibility practices cannot look ever more closely at each case without undermining the efficiency of the system is well taken, but shorn of its optional eliminativist implications, Waller's arguments remain persuasive that there is still a lot of room to look more closely at individual agents than is typical. I hope to have been as persuasive regarding the prospects of the tectonic theory as a way of motivating an interest in looking more closely at subjects of moral responsibility judgments. As Reader wrote, "To make good progress with this, we need to pay attention to things that do not demand our attention."¹⁹⁹

¹⁹⁹ (Reader, 2007, pg. 604)

Chapter five: The Tectonic Theory of Moral Responsibility

In this chapter, I present what I call the basic argument for the tectonic theory of moral responsibility. This argument relies on a small number of what I take to be uncontroversial claims about persons' actions. The basic argument does not invoke the homogenization problem because I think it is important to see how the tectonic theory can be motivated just on the basis of claims about persons' actions. The homogenization problem, thus, becomes a description of how moral judgments can far too easily be neglectful of patientic considerations, while the basic argument locates the reason for taking them into account in the first place. I then flesh out the geological motif used to characterize the ways that agency and patienty can be taken to interact, suggesting that such dynamics can profitably be construed in terms of subduction of one to the other. Subduction is a term for the process whereby one tectonic plate sinks under another and melts back into the earth's mantle. The concept of subduction allows the tectonic theory to recognize that particular cases of moral responsibility can be construed as dominated by either agentic or patientic factors, without having to generalize from either sort of case. These different tectonic configurations make the tectonic theory unusually agile with respect to the particularities of different cases of moral responsibility. I will show this by explaining how a number of popular theories of moral responsibility can be construed as implicitly advocating for taking either agency or patienty as generally subductive of the other.

The basic argument for the tectonic theory

The basic argument for the tectonic theory of moral responsibility is as follows:

- 1- At least some people are responsible for some of their actions
- 2- A person is responsible for that which they control
- 3- A person is not responsible for that which they do not control
- 4- A person's actions are always a result of both controlled and uncontrolled factors
- 5- With respect to any given action then, a person is responsible to the extent that the action was under her control (from 1, 2, 3, and 4)
- 6- Determining an individual's extent of control over a particular action should involve accounting for relevant uncontrolled factors (from 4)
- 7- An individual's moral responsibility for an action, therefore, is a function of both controlled and uncontrolled factors (from 5 and 6)
- 8- A theory of moral responsibility should take both controlled and uncontrolled factors into account (from 7)

I will defend each of these claims in turn.

The first claim reflects my (P.F.) Strawsonian approach, which takes moral responsibility practices to be inevitable for human communities. In chapters one and four, I argued in favour of preserving moral responsibility practices even though they unfairly homogenize across the class of agents. I took this position mainly because of the apparent futility of moral responsibility eliminativism, but the point can also be supported by more positive claims. Ironically, Greene and Cohen's eliminativist paper includes a good discussion of research that points toward the inevitability of moral responsibility practices. Greene and Cohen mention three streams of evidence for the durability of moral responsibility: philosophical, evolutionary, and neuroscientific. Greene and Cohen write,

Regarding responsibility and punishment, one might wonder if it is humanly possible to deny our retributive impulses.²⁰⁰ This challenge is bolstered by recent work in the behavioural sciences suggesting that an intuitive sense of fairness runs deep in our primate lineage and that an adaptive tendency towards retributive punishment may have been a crucial development in the biological and cultural evolution of human sociality.²⁰¹ Recent neuroscientific findings have added further support to this view, suggesting that the impulse to exact punishment may be driven by phylogenetically old mechanisms in the brain.²⁰² These mechanisms may be an efficient and perhaps essential, device for maintaining social stability.²⁰³

As we will see in the next chapter, Greene and Cohen fail to explain how all of these factors can be overcome. They can be applauded, however, for making such a strong case for the opposition, which provides further support for the first claim of the basic argument for the tectonic theory. These considerations speak strongly for the view that moral responsibility cannot be surgically excised from societies.

The second claim, that persons are responsible for that which they control, is an idea that appears to be common to all theorizing about moral responsibility. Even those who deny that moral responsibility is legitimate base their views on claims about persons not having the right sort of control over their behavior, as opposed to challenging the inference from the presence of control to the legitimacy of moral responsibility.²⁰⁴ The third claim is a little trickier, but only because it seldom needs to be said explicitly. For the most part the notion that a person is not responsible for that which they do not control is just appealed to by opponents of moral responsibility who claim that the sense of ‘control’ that is relevant for moral responsibility is never achieved by

²⁰⁰ Greene and Cohen cite (Strawson 1962; Pettit 2002)

²⁰¹ Here, they cite (Brosnan & De Waal, 2003), (Fehr & Gächter, 2002), (Boyd, Gintis, Bowles, & Richerson, 2003), and (Bowles & Gintis, 2004)

²⁰² Greene and Cohen appeal to (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003)

²⁰³ (Greene & Cohen, 2004, pg. 10)

²⁰⁴ Waller is, as far as I’m aware, the only exception to this. He thinks that a naturalistic form of free will does exist but that moral responsibility practices are deeply problematic anyway. This is why I use his work to develop the homogenization problem in the previous chapter.

persons. Nevertheless, even steadfast proponents of moral responsibility typically endorse it at least implicitly by making exceptions of coercion or manipulation cases.

The discussion of interpersonal patiency in chapter two provides an argument to the effect that even Dennett's way of curtailing the importance of patiency within moral responsibility judgments, the principle of default responsibility, endorses the view that in some cases at least, moral responsibility can be unraveled by consideration of uncontrolled factors, i.e. patiency. Thus, even though it is not as commonly endorsed, the third claim in the basic argument is basically a corollary of the second. Although people may often disagree about when someone lacks control over something, this does not impugn the idea that when someone does clearly lack control over something, she is not responsible for its effects. Additionally, the tectonic theory promises to make these disagreements more tractable by conceptualizing different ways in which a person's lack of control over something can be important.

The fourth claim in the basic argument states that a person's actions are always a product of both controlled and uncontrolled factors. Most naturalistic philosophers accept this whether compatibilist or incompatibilist. Fischer explicitly states it: "It is better to think of compatibilism as conceding from the beginning that we are thoroughly subject to factors entirely outside our control. Nevertheless, according to the compatibilist, we can still exhibit a meaningful and robust sort of control."²⁰⁵ Waller is another proponent of the view: "Obviously, we do make choices that influence our subsequent development, but those choices are ultimately traced to influences that were a matter of good or bad fortune, not something we controlled."²⁰⁶ The idea that

²⁰⁵ quoted in (Smilansky 2012, pg. 220)

²⁰⁶ (Waller 2011, pg. 93)

uncontrolled factors never make a difference to a person's behaviour cannot seriously be maintained so long as we accept a non-miraculous or naturalistic view of persons.

The remaining claims in the basic argument are all inferences based on the earlier claims. The fifth claim, that a person is responsible for an action to the extent that the action was under her control, contains two elements. First, it inherits from claim 1 the Strawsonian view that moral responsibility practices are legitimate. Second, it broaches the idea that moral responsibility comes in degrees. This is a consequence of claims 2, 3, and 4. Claim 2 suggests that a person's extent of control over the occurrence of a particular action is a function of the extent to which factors relevant to bringing about the action were under her control. Claim 3 is a reminder that there are inevitably also uncontrolled factors relevant to an action's occurrence. Claim 4 asserts that both controlled and uncontrolled factors are involved in any action, which together with claims 2, and 3, suggests that a person's responsibility for an action accrues according to her extent of control over relevant factors, and recedes according to the extent to which uncontrolled factors figured in bringing about the action. It follows from these, that a person's moral responsibility for an action ought to be tabulated according to both types of factors. Claim 6 explicitly states the part of this idea that is novel to the tectonic theory: judging the extent to which an action can be said to have been under a person's control should involve considering uncontrolled or patientic factors.

Claims 7 and 8 unpack and clarify the import of preceding claims for theorizing about moral responsibility. 7 basically recapitulates claims 5 and 6. Claim 5 couples a person's extent of control and moral responsibility with respect to a given action, and 6 states that these should be

sensitive to patientic considerations, i.e. uncontrolled factors. Claim 7 describes the consequences of these claims for particular moral responsibility judgments: an individual's moral responsibility for an action is a function of both controlled and uncontrolled factors. Claim 8 collects the substance of 7 into a norm about the philosophers' task in theorizing about moral responsibility. From the fact that judgments about an individual's moral responsibility should be sensitive to both agentic and patientic considerations, it follows that theories of moral responsibility should take both of these into account. I turn now to delineating different ways that this could go, and showing that many philosophers' claims about moral responsibility can be characterized as advocating for the heuristic supremacy of particular forms of the agency/patiency dynamic.

A delineation of possible dynamics between agency and patiency

The virtue of the tectonic theory of moral responsibility is not just that it makes judgments about moral responsibility sensitive to both an individual's control (agency), and relevant uncontrolled factors (patiency), but that it anchors such considerations to features of persons. Instead of focusing on agency so as to rebut eliminativist arguments, as is typical of moral responsibility proponents, or focusing on patiency so as to mount eliminativist arguments, as is typical of moral responsibility skeptics, the tectonic theory is alert to both aspects of a person and explains why this is both appropriate and valuable. The basic argument presented above is intended to show that the tectonic theory is an appropriate response to the basic features of human actions. That the tectonic theory can help to combat the homogenization problem developed in the previous chapter is intended to explain how a concern for patiency can be valuable within moral

responsibility judgments.

In addition to the metaphorical appeal to agency and patiency as tectonic plates, I indulge in importing the geological notion of *subduction* so as to characterize some of the different ways in which agency and patiency might be thought to be related to one another in particular cases of morally responsible action. Subduction is the name for the process whereby one tectonic plate sinks under another before disappearing into the earth's mantle. I think this concept adds some valuable dynamical content to the tectonic theory by conceptualizing the extremities of a spectrum of ways that agency and patiency can be related to one another in particular moral responsibility judgments. If it is judged of an action that the contribution of agentic factors was drastically more significant to the action's occurrence than the contribution of patientic factors, then this judgment classifies agency as subducting patiency. Likewise for judgments to the effect that an action is primarily a result of patientic factors; such cases involve patiency subducting agency.

The subduction motif also offers a vocabulary for discussing whether certain types of agency or patiency are particularly powerful. Forms of agency that are relatively robust against the contributions of patientic factors can be described as more subductive. Likewise, varieties of patiency that are particularly mitigating of a person's moral responsibility can be thought of as more subductive. The tectonic theory thusly allows for new ways to interrogate the various agentic and patientic factors involved in persons' actions. Although the tectonic theory is designed mainly to inject typically neglected patientic nuance into particular moral responsibility judgments, I think that it can also serve a secondary purpose of bringing to light people's

intuitions about the relative significance of particular agentic and patientic features of persons and their lives. For example, a person's upbringing is commonly invoked as both explanatory of a person's adult agency, and as an inducement toward leniency in highly unfortunate cases.²⁰⁷

The tectonic theory suggests that this is because upbringing is thought of as a highly subductive sort of patienty, i.e. a case in which uncontrollable factors such as the behaviour of one's parents can be taken to be more significant than whatever agency accumulates under their influence. I am not sceptical about the mitigating value of persons' upbringing, but I do think that the thought deserves scrutiny. The tectonic theory of moral responsibility is intended to create a philosophical discussion in which this, or any other thought about the relative importance of aspects of agency and patienty, can be scrutinized effectively. Later in this chapter, I will look at the case of Robert Harris, whose childhood was decisively unfortunate, to provide a sense of how the tectonic theory can illuminate such aspects of moral responsibility judgments. First, I turn to developing the idea that agency can be judged to subduct patienty and showing how this idea can be found in the work of prominent proponent of moral responsibility Manuel Vargas.

Agency can subduct patienty

The idea of agency subducting patienty applies to an individual who has utilized her agentic powers to neutralize patientic factors that otherwise might have disrupted her agency.

²⁰⁷ Ted Honderich sums up the way that consideration of a person's upbringing can alter one's thinking about her moral responsibility: "suppose not that I become convinced on some future day of the theory of determinism we have been considering, but just that right now I fall into a common kind of half-determinist feeling and thinking about action. This sort of thing often has to do with thoughts of the underprivileged background and history of a person. I come to feel that given the sort of person he is, and given the situation he was in, he did not have much chance of behaving differently. What happens? My retributive desire at least falters." (Honderich 2002, ch. 8)

Philosophers who argue that situationist factors can be successfully mitigated by agentic efforts can be read as picturing the subduction of patiency by agency. Manuel Vargas, for example, theorizes that so long as we give up on a traditional contra-causal conception of free will,²⁰⁸ moral responsibility practices are safe from situationist threats. In ‘Situationism and Moral Responsibility: Free Will in Fragments,’²⁰⁹ Vargas recounts several classic social psychology studies,²¹⁰ and argues that the basic claim of situationism – that features of situations exert significant influence on persons’ behaviour without the persons’ becoming aware of it – is actually valuable for moral responsibility practices. Here is Vargas’ summary of the famous Milgram experiment: “In a series of widely-duplicated experiments, Stanley Milgram showed that on the mild insistence of an authority figure, a range of very ordinary subjects were surprisingly willing to (apparently) shock others, even to apparent death, for failing to correctly answer innocuous question prompts.”²¹¹ Milgram’s results are unsettling because they show how features of situations, in this case the presence of a white coated laboratory worker authoritatively encouraging the continuation of the experiment, can cause people to behave in ways that they would not have taken themselves to be capable of.

Vargas’ view is that particular situationist claims are useful for ascertaining some of the factors that can enhance or degrade agency. He writes, “Rather than thinking of the psychological

²⁰⁸ Vargas is a revisionist about free will, claiming that the basic idea of free will being the feature of persons that legitimizes moral responsibility should be retained but in an altered form that doesn’t make a metaphysical special-case of persons. Instead, Vargas proposes to conceive of free will as constituted by a capacity to respond to moral considerations, and moral responsibility practices as legitimized on the basis that they cultivate this type of agency. The most comprehensive presentation of these aspects of Vargas’ work is in (Vargas, 2013a).

²⁰⁹ (Vargas, 2013b)

²¹⁰ Vargas looks at the *phone booth* study in which finding a dime in a payphone made subjects much more likely to help a person in distress (Isen and Levin, 1972), the *good Samaritan* study which showed that seminary students were much more likely to stop and help a stranger if they were not in a hurry (Darley and Batson, 1973), and the infamous Milgram obedience experiments which showed that many subjects were willing to administer painful shocks to others on the insistence of authority figures conducting the experiment (Milgram, 1969)

²¹¹ (Vargas, 2013b, pg. 4)

sciences as presenting threats to free will, we do well to think of them as providing us with resources for enhancing what freedom we have.”²¹² To complete the case that Vargas can be read as advocating for agentic aspects of persons as capable of subducing patientic ones (at least with respect to situationist factors), I need only to establish that he takes situationist factors to be patientic, according to my definition. Understandably given his interest in asserting the primacy of agency, Vargas does not emphasize that situationist influences on persons’ behaviour are beyond the reach of agentic control, but he does appear to appreciate that this is the case.

Considering the perspective of someone who might have hoped for a more aggrandizing take on situationism, to the effect that persons are generally capable of recognizing and overriding situationist influences, Vargas writes, “After all, we are still faced with a not-altogether-inspiring upshot that since we do not control our situations as much as we like, we are still not responsible agents as much as we might have hoped.”²¹³ Implicit in this comment are the ideas that moral responsibility can be attenuated in contexts that involve situationist factors, and that this can be traced to the fact that the situation in which persons find themselves is frequently a matter that is beyond their control, i.e. patientic.

Vargas’ description of how situationist factors are relevant to behaviour further clarifies their patientic nature. He writes, “The general lesson of situationism is that we underestimate the influence of the situation and we overestimate the influence of purportedly fixed features of the agent. Crucially, the “situational inputs” typically operate without the awareness of the agent.”²¹⁴ It is the fact that situationist factors go undetected that confirms their patientic nature. In chapter

²¹² (Vargas, 2013b, pg. 26)

²¹³ (Vargas, 2013b, pg. 28)

²¹⁴ (Vargas, 2013b, pg. 4)

two, I presented Dennett's analysis of the concept of control to use in my definition of patiency. I committed to the idea that (among other things) a person's control over something depends on knowledge about the different states that it could be in and how to put it into the state she desires. According to situationism, a person is influenced by situations in ways that she does not know about and she is thus incapable of controlling for their contribution to her behaviour.

Importantly, the reflections above apply just to occurrent situationist factors. A person can know in advance that certain situations will influence them in particular ways and thus make efforts to avoid such situations. This does not impugn the claim that situationist factors are patientic, however. In fact, the sort of agentic efforts that Vargas suggests can be put toward mitigating situationist influences supports the claim that such influences are patientic. For example, Vargas writes, "[S]uppose Samaritan-like data comes to show that for most people time pressure significantly reduces our capacity to perceive what we regard as moral considerations. Such a discovery has an important implication for our moral ecology: in those cases where we have an interest in not significantly limiting helping behavior, it behooves us to limit time-pressuring forces."²¹⁵ The suggestion that it would be worthwhile to limit 'time-pressuring forces' hinges on the idea that the influence of such forces cannot be altered through agentic efforts once it is occurrent, which is a hallmark feature of patientic factors as I have defined them. Vargas' defense of moral responsibility against situationist threats, thus, describes people as capable of agentially curating some patientic aspects of their lives, thereby mitigating their influence. This is a good example of what I mean in suggesting that agency can be thought of as subducting patiency in some cases.

²¹⁵ (ibid, pg. 26)

It is, of course, an open question whether the sorts of agentic efforts Vargas conceives as capable of mitigating the patientic contributions of situationist factors are really available for persons. In describing Vargas' take on situationism as conceiving of agency as subductive with respect to situationist patienty, I do not intend to endorse his view; the aim was just to provide an example of someone thinking that agency can subduct patienty, if only in connection to the relatively circumscribed concerns of situationism.

In addition, I have only presented Vargas' take on situationism, neglecting other aspects of his theorizing about moral responsibility. I will now show that on a much more abstract level, Vargas conceives of agency as *generally* subductive with respect to patientic factors. In this, Vargas' example is typical of compatibilist work. His concern is to neutralize apparent threats to moral responsibility rather than to conceptualize ways that the content of such "threats" might be integrated into particular moral responsibility judgments. This is the case even though Vargas' work is comparatively alert to patientic phenomena.

Unlike many philosophers, Vargas explicitly claims that agency comes in degrees, with variation being accounted for by both broadly situationist factors and the automatic/non-conscious character of many of the psychological processes from which agency emerges.²¹⁶ Both sorts of factors have already been categorized as patientic according to my definition. In response to such concerns, Vargas develops the doctrine of *circumstantialism*. This is a theory about persons' agentic powers. Vargas writes, "The operative idea in circumstantialism is this: the powers that

²¹⁶ "[C]ontemporary psychology maintains that the causal sources of our actions have less to do with features of ourselves than they do with deliberately irrelevant features of the situations in which we act. Moreover, much of human psychological activity amounts to automatic processes, often invisible and at odds with our awareness." (Vargas, 2013a, pg. 2-3)

matter for whether an agent is responsible are best characterized non-intrinsically, as functions of agents in circumstances ... Those circumstances can be structured in ways that are more and less conducive to responsible agency.”²¹⁷ Vargas’ theory thus incorporates some patientic considerations in the form of circumstantial limitations of persons’ agentic powers.

However, Vargas’ work still inherits the all-or-nothing tendencies of the moral responsibility literature, and he primarily deploys circumstantialism as a way of entertaining but ultimately dismissing eliminativist arguments. In the next-to-last chapter of *Building Better Beings*, called “History and Manipulation,” Vargas makes a case for the context dependence of agentic powers. He concedes that an individual’s capacity to detect and/or respond to moral considerations can vary across different situations. His interest, however, is to explain how continuing to hold people morally responsible is justified despite this. Vargas does not explore the criteria according to which different circumstantialist limitations on persons’ agency might be accorded different degrees of significance. Instead, he argues just that continuing to hold people responsible despite these complexities is justified on the basis that it plausibly cultivates agency.

As a sketch of the pragmatic value of moral responsibility practices, this is fine, but Vargas’ argumentation does not entail that patientic phenomena should be left out of particular moral responsibility assessments, just that they should not be considered in place of moral responsibility judgments and practices. The tectonic theory, therefore, can be taken to make some progress in comparison to traditionally compatibilist work like Vargas’. Characterizing Vargas’ particular view on situationism and his general account of moral responsibility as endorsements of the view that agency is subductive of patiency allows for evaluating his position not just as a

²¹⁷ (Vargas, 2013a, pg. 3)

good legitimization of moral responsibility judgments in general, but also as a fallible recipe for particular moral responsibility judgments.

The way that the tectonic theory makes clear that the ingredients of moral responsibility judgments should be agency and patiency allows for scrutinizing particular theories in terms of the extent to which they motivate the incorporation of detail into particular moral responsibility judgments. A construal of Vargas' work as just a general defense of moral responsibility, in contrast, can easily remain agnostic about the value of making particular moral responsibility judgments sensitive to concerns beyond the cultivation of agency. One payoff of the subduction motif, thus, is that it allows for classifying (some) existing theories about moral responsibility as well as (some) particular moral responsibility judgments as endorsements of the priority of agentic concerns over patientic ones instead of just arguments against eliminativism. This, I think, can open up interesting new ways to evaluate and refine proposals about what is important for moral responsibility judgments.

Patiency can subduct agency

It can go the other way too, of course. The concept of subduction also allows for the analysis of eliminativist theories as endorsements of the priority of patientic considerations over agentic ones, instead of just as more or less compelling condemnations of moral responsibility practices. The tectonic theory is concerned with how to improve moral responsibility judgments, rather than with their overarching legitimacy or lack thereof. As a result, I can take suggestions from eliminativists regarding what such judgments should take into account while simultaneously

rejecting their claims about how such factors undermine moral responsibility generally. Applying the subductive motif to eliminativist arguments will facilitate this. I'd like to work my way up to this, however, first recapping the basic idea of patency subducting agency and then providing some examples.

The basic idea of taking someone's patency to subduct her agency was already discussed in chapter two, when I classified manipulation cases as involving interpersonal patency. I argued for construing the behaviour of a manipulated person as brought about by factors beyond her control, factors which happen to be other people. This is an alternative to approaches like Dennett's principle of default responsibility which transfers the moral responsibility to the manipulator. Dennett's approach can be construed as making a special case of interpersonal patency, and denying that other varieties of patency have the same importance. There is no need to limit patientic considerations this way, however. Chapter two aimed to persuade that there is value in thinking of patency as a much broader phenomenon, and in chapter four, I argued that Dennett's rejection of the importance of broader patientic considerations, while a sensible response to eliminativist arguments, does not apply to the less drastic recommendations of the tectonic theory.

Here I would like to make the case that patientic considerations underlie not just the intuitive mitigation of a manipulated persons' moral responsibility for her behaviour, but also other kinds of mitigation. To make this point I need to show how different kinds of excuses for and exemptions from moral responsibility are responses to patientic factors. The clearest cases of such mitigation are ones in which patency can be seen to subduct agency, i.e. cases in which the

contributions of uncontrollable factors are taken to overwhelm the contributions of controllable ones. I will present two such cases that do not involve intentional manipulation, though one does involve interpersonal patiency: persons with Tourette's syndrome, and Robert Harris.

Considering Tourette's syndrome is a good way to illuminate judgments to the effect that patiency can subduct agency because it is basically constituted by inability to control certain behaviours. Mainly, sufferers of Tourette's syndrome exhibit motor tics, and coprolalia, the involuntary utterance of obscenities. Several researchers agree²¹⁸ that the earliest recorded account of a person with Tourette's syndrome is probably the story of a priest's son told in *The Malleus Malifecarum*,²¹⁹ a 1486 guide to identifying witches and other spiritual unfortunates such as the possessed. The priest's son took himself to have been readied for demonic possession by a woman that he failed to assist in her spiritual troubles. Unusually for a case of demonic possession, the priest's son's symptoms were intermittent. The authors of the *Malleus* write,

in all his behaviour he remained a sober priest without any eccentricity ... he showed no sign of madness or any immoderate action. But when he passed any church, and genuflected in honour of the Glorious Virgin, the devil made him thrust his tongue far out of his mouth; and when he was asked whether he could not restrain himself from doing this, he answered: "I cannot help myself at all, for so he uses all my limbs and organs, my neck, my tongue, and my lungs, whenever he pleases, causing me to speak or to cry out; and I hear the words as if they were spoken by myself, but I am altogether unable to restrain them; and when I try to engage in prayer he attacks me more violently, thrusting out my tongue."²²⁰

Even in an era of such superstition, the fact that the behavior resulting from such a possession was beyond the control of the possessed appears to have been assigned mitigating value within moral responsibility judgments. Later in the *Malleus*, it is written that, "we shall in no way despise those who are possessed; but we shall incessantly pray for them as for our own selves,

²¹⁸ (McNaught & P., 2010), (Germiniani, Miranda, Ferenczy, Munhoz, & Teive, 2012)

²¹⁹ (Kramer & Sprenger, 2007, original published in 1486)

²²⁰ (Kramer & Sprenger, 2007, part 2, question 1, chapter 10)

and pity them with our whole heart.” The basis of this view is the idea that possession is ultimately part of God’s inalterable plan, which is problematic in today’s naturalistic climate, but is still indicative of the deep roots of the intuition that uncontrollable factors mitigate a person’s responsibility for actions. Although it would be decisively insufficient solace for anyone undergoing an exorcism, the fact that the authors of the *Malleus* took a compassionate stance toward such persons because of their lack of control over themselves is telling of the deep intuitive importance of patiency.²²¹

Today, Tourette’s syndrome is thought of as a neurological disorder and the public is pervasively aware of this. For the most part, once a person is known to suffer from Tourette’s, when she suffers from her uncontrollable impulses, she will not be held morally responsible for their occurrence. That this is commonplace should not distract from its importance with respect to the mitigating value of patientic considerations. I do not think it is controversial to say that the reason sufferers of Tourette’s syndrome are excused from moral responsibility for the offensive things they say is because it is widely understood that such outbursts are not susceptible to agentic suppression in such people. In other words, the suffering of Tourette’s consists of phenomena that are clear examples of highly subductive patiency.

²²¹ In fact, the *Malleus* contains a surprisingly nuanced discussion about moral responsibility and how it can be mitigated by patientic factors, such as demons. The authors explicitly consider the matter of whether witches deserve moral responsibility for the actions that result from their demonic possession. They write, “if it is asked whether injuries of this sort are to be ascribed rather to devils than to witches, it is answered that, when the devils cause injuries by their own direct action, then they are principally to be ascribed to them. But when they work through the agency of witches for the disparagement and offending of God and the perdition of souls ... such injuries are justly ascribed to witches secondarily, however much the devil may be the principal actor.” (Kramer & Sprenger, 2007, part 2, question 1, ch. 11) They go on to convey ambivalence regarding whether to blame witches: “when a woman dips a twig in water and sprinkles the water in the air to make it rain, although she does not herself cause the rain, *and could not be blamed* on that account, yet, because she has entered into a pact with the devil by which she can do this as a witch, although it is the devil who causes the rain, *she herself nevertheless deservedly bears the blame*, because she is an infidel and does the devil's work.” (Kramer & Sprenger, 2007, part 2, question 1, chapter 10)

Accordingly, there is a nascent debate about whether Tourette's should be used as a legal defense. The Tourette Association of America has gone on record in opposition to the idea. A statement on their website reads: "The Medical Advisory Board of the Tourette Syndrome Association is deeply concerned about media reports that attribute violent criminal behavior by individuals to the fact that they have Tourette Syndrome ... the diagnosis of Tourette Syndrome in a legal offender is no more the reason for, or an excuse for such offense than other medical diagnoses -- such as asthma or rheumatism."²²² However, one wide-ranging review paper presents a different view on the basis that Tourette's often co-occurs with other deficiencies such as generally poor impulse control and obsessive behaviour.²²³ Jankovic et al. write that Tourette's "rarely leads to criminal behavior, but patients with TS who have behavioral comorbidities are at risk of being involved with the legal system. The medical-legal community must learn to recognize the vulnerability of this patient population to potential mistreatment by the courts of justice."²²⁴

Additionally, there was a case of a 52 year old Turkish man suffering from Tourette's syndrome who swore at a referee during a sporting event. In April of 2007, the Department of Forensic Medicine in the Medical Faculty of Celal Bayar University in Turkey was asked by the local courts to issue a report as to whether the man was criminally responsible for the utterance in question. The report concluded that the man's "behavior at the football match was a[n incidence of] reflexive echolalia and coprolalia in which there was no criminal intent. Consequently, he

²²² ("Tourette Syndrome and Criminal Behaviour," 2014)

²²³ (Jankovic, Kwak, & Frankoff, 2006)

²²⁴ (ibid, pg. 86)

had no criminal responsibility for swearing at the referee during a football match.”²²⁵ Although the report based their exoneration on the fact that there was no criminal intent, the rationale can easily be reformulated in terms of patiency: the man was not morally or legally responsible for swearing at the referee because factors beyond his control (the fact that he suffered from Tourette’s syndrome) were dominant in bringing about the action. Such a case is a clear example of how patiency can subduct agency.

I now turn to the difficult case of Robert Harris, an example in which patiency’s subduction of agency is more expansive and more troubling. Watson first presented Harris’ story as a challenge to Strawsonian accounts of moral responsibility.²²⁶ In particular, Watson focuses on Strawson’s brief discussion of how consideration of an individual’s past can alter moral responsibility judgments made about her. Strawson’s over-arching aim was to illustrate the unimportance of metaphysical considerations to do with determinism for what he called ‘reactive attitudes.’ The phrase ‘reactive attitudes’ refers to the tendencies of persons to hold one another morally responsible for their behaviour, and Strawson characterizes them as inevitable within and thus crucial for human communities. Importantly for the turn away from metaphysics that Strawson advocates, he conceives of the reactive attitudes as responsive to the presence of good or ill will that persons have toward one another, judgments about which have nothing to do with determinism and free will as traditionally conceived.

²²⁵ (Gullucayir, Asirdizer, Yavuz, & Zeyfeoglu, 2009)

²²⁶ Watson states his aim thusly: “As Strawson presents it, [his theory of moral responsibility] is incomplete. I will investigate whether and how the incompleteness can be remedied in Strawsonian ways. In the end, I find that certain features of our practice of holding responsible are rather resistant to such remedies, and that the practice is less philosophically innocent than Strawson supposes” (Watson, 1987, pg. 120)

Strawson also discusses a number of ways that persons are sometimes spared a response of reactive attitudes, including consideration of a person's upbringing.²²⁷ Watson's concern in presenting the Robert Harris case is to investigate whether Strawson's denial of the importance of determinism can be squared with the idea that a person's upbringing sometimes mitigates or eliminates her moral responsibility. It might be that when a person of 'peculiarly unfortunate formative circumstances' is not held responsible because of such circumstances, the thesis of determinism needs to be invoked to explain why. If the structure of the thought that withholds blame from such a person involves the idea that her later behaviour was unavoidable due to her upbringing, this would suggest an acceptance of determinism facilitates such thoughts. Watson thus sets out to investigate whether taking Robert Harris' horrific formative circumstances to challenge his moral responsibility for later acts needs to invoke determinism.

Watson takes Strawson's theory to handle excusing conditions for moral responsibility, such as duress, well enough. With respect to exempting conditions, which Watson refers to as 'type-2 pleas,' Strawson's theory is found to be incomplete, however. Watson writes, "Evidently reactive attitudes are sensitive not only to the quality of others' wills, but depend as well upon a background of beliefs about the objects of those attitudes."²²⁸ Watson's reading of Strawson is that when a person is exempted from moral responsibility, this is because an explanation of their

²²⁷ "Seeing someone, then, as warped or deranged or compulsive in behaviour or peculiarly unfortunate in his formative circumstances – seeing someone so tends, at least to some extent, to set him apart from normal participant reactive attitudes." (P. F. Strawson, 1962, pg. 9)

²²⁸ (Watson, 1987, pg. 125-126)

behaviour is at hand.²²⁹ He notes that “Strawson’s essay does not provide an account of how this works or what kinds of explanations exempt.” The Robert Harris case is presented to make vivid the aspects of moral thinking that Strawson fails to explain. How, if not by endorsing the view that Harris’ upbringing deterministically explains his later behaviour, are his peculiarly unfortunate formative circumstances supposed to explain his behaviour to an extent that exempts him from moral responsibility?

Watson’s presentation of Robert Harris’ case is roughly tectonic. First, he presents Robert Harris as a dangerous agent, recounting his horrific crimes and awful disposition. Harris shot and killed two boys without breaking what Watson calls ‘an almost lighthearted mood,’ and once incarcerated was universally despised by the other inmates on death row. Once he has provided evidence for Harris being a prototypical example of human callousness, Watson pivots focus to Harris’ horrific childhood, thus presenting him as a patient. As a child, Harris was horrifically abused by his parents, then schoolmates, then at age 14 in a federal youth detention center, he was raped several times and attempted suicide twice. He was released at age 19 a severely troubled man and soon began gleefully torturing animals.

Watson writes, “No doubt this history gives pause to the reactive attitudes,”²³⁰ but struggles to explain why this is the case. Other Strawsonian philosophers have addressed this issue more generally. In his seminal paper, Strawson mentions that individuals who suffer ‘peculiarly

²²⁹ “It would seem that many of the exempting conditions involve *explanations* of why the individuals display qualities to which the reactive attitudes are otherwise sensitive.” (ibid, pg. 126)

²³⁰ (Watson, 1987, pg. 137)

unfortunate formative circumstances'²³¹ are candidates for suspended reactive attitudes. Michelle Ciurria's paper 'Moral Responsibility: Justifying Strawson and the Excuse of Peculiarly Unfortunate Formative Circumstances'²³² provides a good discussion of the difficulties involved in justifying the suspension of reactive attitudes towards individuals of unfortunate upbringing. Ciurria develops a justification for taking unfortunate upbringing to mitigate moral responsibility based on an externalist Strawsonianism, drawn from Andrew Sneddon's 'Moral responsibility: the difference of Strawson, and the difference it should make.'²³³ Sneddon and Ciurria construe being morally responsible as an interpersonal capacity, like playing tennis. This provides the means to explain the mitigating value of unfortunate upbringings on the basis that they reduce persons' capacities to be morally responsible, similar to the way that a novice tennis player's background explains her inability to compete with experts. I see the tectonic account of Harris' case, and cases like his, as compatible with this view, but providing further detail about the mitigation at work.

I suggest that it is important to see the factors that affected Harris' development as patientic, and not just as having reduced his capacity for morally responsible agency. If someone voluntarily reduced their capacity for membership in the moral community, say by knowingly taking a pill that would have such an effect, she could presumably be held responsible for her subsequent behaviour. These considerations and those I will develop while working through Watson's treatment of Harris will show that considering Harris as a patient instead of just as an agent lacking in capacity is the best way to explain the intuitive leniency that is stimulated by reflection on his upbringing.

²³¹ (P. F. Strawson, 1962, pg. 7)

²³² (Ciurria, 2013)

²³³ (Sneddon, 2005)

A quick Strawsonian explanation might claim that Harris' history illuminates the extent to which he is apart from any moral community, fundamentally unsuited to personal relationships. Watson notes, though, that this was known about Harris before his history was presented. There thus remains a mystery as to why Harris' upbringing gives pause to the reactive attitudes involved in holding persons morally responsible. Watson describes this as "a hard question for compatibilist views generally" because proponents of moral responsibility such as Strawson seem to be committed to the idea that whether to hold a person morally responsible is a question "settled by how individuals *are*, not by how they came to be."²³⁴

In contrast, Watson describes the intuitive importance of Harris' childhood as "grist for the incompatibilists' mill. For they will insist on an essential historical dimension to the concept of responsibility. Harris's history reveals him to be an inevitable product of his formative circumstances."²³⁵ But Watson is not happy with this sort of analysis either: "the force of the example does not depend on a belief in the *inevitability* of the upshot. Nothing in the story supports such a belief. The thought is not "It had to be!" but, again, "No wonder!"²³⁶ This is another way in which Watson's treatment of the Robert Harris case has affiliations with the tectonic theory. Watson recognizes that what I called the forensic stance in chapter one is not necessary in order to be moved by consideration of the contributions of uncontrolled factors to a

²³⁴ (Watson, 1987, pg. 137)

²³⁵ (ibid)

²³⁶ (Ibid, pg. 138)

person's behaviour. However, without the concept of patiency in place, the rest of Watson's treatment of Harris' case, by Watson's own admission, struggles toward coherence.

Watson writes, "What appears to happen is that we are unable to command an overall view of his life that permits the reactive attitudes to be sustained without ambivalence."²³⁷ Watson's response to Harris, thus, has some affinity to Smilansky's treatment of "trapped" individuals presented in chapter one, although Watson does not insist that Harris' future must be understood to have been inevitable as a result of his upbringing. Watson explains his ambivalence as an oscillation between seeing Harris as a victim of his childhood and as the agentic man he became: "in light of the "whole" story, conflicting responses are evoked. The sympathy toward the boy he was is at odds with outrage toward the man he is ... each of these responses is appropriate, but taken together they do not enable us to respond overall in a coherent way."²³⁸

The difficulties Watson confronts in trying to reconcile the different aspects of his response to Harris seem connected to the fact that he is working with the concept of 'a victim' instead of a patient. In the above paragraph, Watson writes of sympathy toward the boy and outrage toward the man. The tectonic theory can diffuse this tension. Robert Harris, the man, is the same person that underwent the terrible upbringing in question. In chapter two, I discussed how it is important to distinguish between ascriptions of victimhood and the withholding of ascriptions of agency. If being classified as a victim simply consists of being placed outside the category of agents, this

²³⁷ (Watson, 1987, pg. 138)

²³⁸ (Ibid)

leads to a neglect of the ways that the consequences of a victim's misfortune can reside within the victim's agentic features.

Watson appears to be working with a notion of victimhood that entails that victims should not be subject to moral responsibility. That Watson takes the categories of victim and victimizer to be mutually exclusive is apparent in the following passage:

Harris both satisfies and violates the criteria for victimhood. His childhood abuse was a misfortune inflicted upon him against his will. But at the same time (and this is part of his very misfortune) he unambivalently endorses suffering, death, and destruction ... With this in focus, we see him as a victimizer and respond to him accordingly. The ambivalence results from the fact that an overall view simultaneously demands and precludes regarding him as a victim.²³⁹

The first thing to mention about this passage is that the parenthetical "and this is part of his very misfortune" undermines the chronological analysis Watson presents earlier. If it is the case that Harris' present day identification with his transgressive behaviour forms part of the misfortune due to his childhood, it would seem natural enough to extend sympathy to the man that emerged from the boy's experiences since the formative experiences are still having effects. The tectonic theory can explain why this is so. The fact that Watson describes Harris' unambivalent endorsement of human horrors as part of his misfortune suggests that Watson would classify the abuse Harris incurred as a child as a subductive variety of patiency.

²³⁹ (Watson, 1987, pg. 138)

Although Harris has sufficient agency to be found morally responsible and even explicitly takes responsibility for his actions,²⁴⁰ elements of his patiency were so extreme that they seem to account for the eventual character of his agency. This is probably as good an example of characterizing patiency as subductive of agency as can be found. Watson writes, “Harris’s cruelty is a response to the shattering abuse he suffered . . . He defies the demand for human consideration because he has been denied this consideration himself.”²⁴¹ If Watson is right about the relation between Harris’ abuse and his callousness, then Harris’ moral responsibility for his actions should be minimal or non-existent, according to the tectonic theory.

Because patientic aspects of Harris subduct his agentic features, punishing him would not be legitimate. This is also Waller’s view of the Robert Harris case. Waller presents Harris again and again as clear example of someone who should be spared moral responsibility on account of the way that factors beyond his control had overwhelming influence on the agent he became.²⁴² This straightforward conclusion is resisted by Watson, but a close look at the ways that he resists it is instructive with respect to the virtues of the tectonic theory. Although he characterizes judging Harris to be both victim and victimizer as incoherent, Watson never explains what it is that’s problematic about seeing a person in both of these aspects aside from that it evokes feeling that do not usually co-occur (sympathy and antipathy). For my part, I have often felt both sympathy and antipathy towards persons, and this has never struck me as incoherent.

²⁴⁰ Watson cites a newspaper article which quotes Harris’ sister as saying that Robert understood himself to have had an opportunity to be otherwise: “He told me he has his chance, he took the road to hell and there’s nothing more to say.” (Watson, 1987. Fn. 22)

²⁴¹ (ibid, pg. 140)

²⁴² To take just one of the over 100 mentions of Harris from Waller’s book: “Robert Harris became a vicious person: a claim consistent with the conclusion that he was not morally responsible for the character shaped by those forces so far beyond his powers of control.” (Waller, 2011, pg. 181)

To be sure, moral responsibility judgments typically cast a person as either a victim or victimizer. Gray and Wegner's work on dyadic morality discussed in chapter three showed that this is a robust dichotomy, but this can be taken as a problematic aspect of moral thinking that the Harris case brings to light and challenges us to overcome. From the point of view of the tectonic theory, all agents are also patients, and there is no barrier to taking both into account. When compared to a tectonic treatment of Harris' case, Watson's difficulties in developing a coherent response result from the way that he deploys the concept of a victim instead of the concept of patiency. In the passage quoted above, Watson writes of Harris' case both demanding and precluding seeing him as a victim. I agree with the first part; Harris' childhood marks him as a victim, though unless our interest is to offload some blame for Harris' adult behaviour onto his parents, schoolmates, and administrators at his detention centre, which is not Watson's concern, the word 'patient' is better than victim. And once Harris' victimhood is construed in terms of his patiency, granting that he displayed agency need no longer preclude taking the important uncontrollable elements of his life into account.

Watson usefully specifies that his difficulty in parsing the significance of the agentic and patientic aspects of Harris and his life is not a result of oscillating between wholly different ways of looking at persons.²⁴³ Instead, Watson appears to interpret the Harris case as revealing troubling aspects about persons quite generally. One way of thinking about the case is to say that Harris was subject to bad moral luck, and this is what induces Watson to become introspective:

²⁴³ "What we have here is not exactly a clash between what Thomas Nagel has called the objective and subjective standpoints. It is not that from the more comprehensive viewpoint that reveals Harris as a victim, his responsibility is indiscernible. Rather the clash occurs within a single point of view that reveals Harris as evil (and hence calling for enmity and moral opposition) and as one who is a victim (calling for sympathy and understanding)." (Watson, 1987, pg. 138)

“the thought about moral luck turns one’s gaze inward.”²⁴⁴ He writes, “the case is troubling in a more personal way. The fact that Harris’s cruelty is an intelligible response to his circumstances gives a foothold not only for sympathy, but for the thought that if *I* had been subjected to such circumstances, I might well have become as vile.”²⁴⁵

This move has some affinity with Waller’s counterpart tactic, discussed in the previous chapter. It is different inasmuch as Watson presents himself as a contrast to Harris, but this kind of introspection seems to me to be an optional intermediary. As I argued before, comparing a target for moral responsibility judgments to others can bring to light all that moral responsibility judgments far too easily leave out of account. Where Watson’s gaze turns inward, mine and Waller’s fixes on moral responsibility judgments themselves, and charges them with the homogenization problem. At first, Watson’s introspection appears to take place instead of noticing how odd it is that he was prepared to condemn Harris as a person without having considered him as a patient at all.

Watson’s acrobatic reasoning toward a settled view of the Harris case can be read as an alternative to diagnosing the homogenization problem. Watson appears to pursue a conclusion of comfort, one that preserves a distinction between he and Harris that is not just a matter of luck.²⁴⁶ After considering the possibility that something in Harris’ genetic make-up accounts for why he

²⁴⁴ (Watson, 1987, pg. 138)

²⁴⁵ (Ibid, pg. 139)

²⁴⁶ “Horrid backgrounds do not inevitably give rise to horrid people. Some manage to survive a similar magnitude of misfortune, if not unscathed, at least as minimally decent human beings.” (Watson, 1987, pg. 140)

responded to his upbringing the way he did, Watson writes, “There is room for the thought that there is something “in me” by virtue of which I would not have become a vicious person in Harris’s circumstance.”²⁴⁷ Indeed, there is room for this thought, but I do not see its normative implications for moral responsibility practices.

As Watson concedes, “to suppose that this difference is in itself to my moral credit would be odd. To congratulate me on these grounds would be to congratulate me on being myself.”²⁴⁸

These remarks suggest that Watson understands a possible genetic difference between he and Harris to be patientic. The fact that Watson needs to go all the way to the possibility that there is something different about their genes in order to render a difference between he and Harris that legitimizes holding Harris responsible²⁴⁹ is illustrative of the way that making distinctions between agents often requires patientic considerations. Watson concedes that such a genetic distinction would be beyond either’s control, yet he manages to interpret this as safeguarding moral condemnation of Harris instead of as an indictment of the way that moral responsibility judgments typically fail to take such things into account.

Watson does eventually focus on the way that moral responsibility judgments typically ignore historical factors, but he fails to see that it is precisely because ignorance is an enduring feature of such judgments that historical considerations are significant. In other words, the homogenization problem is what makes historical considerations significant. A person’s history

²⁴⁷ (Watson, 1987, pg. 141)

²⁴⁸ (ibid)

²⁴⁹ “It might be that Harris’s genetic origins are among the determinants of his moral development.” (ibid)

can give the reactive attitudes pause because upon confronting a person's history, one realizes that the reactive attitudes were ready to be deployed (perhaps already deployed) without even having stopped to consider mitigating factors that the agent did not control. Watson writes,

Whatever the best explanation [for why historical considerations affect the reactive attitudes] may be, the remarkable fact is that we are, for the most part, quite ignorant of these considerations. Why does our ignorance not give us more pause? If, for whatever reason, reactive attitudes are sensitive to historical considerations, as Strawson acknowledges, and we are largely ignorant of these matters, then it would seem that most of our reactive attitudes are hasty ... as skeptics have long maintained.²⁵⁰

Precisely. If we keep in mind what it is to be a person, specifically that it involves patiency as much as agency, and carefully interrogate the features of moral responsibility judgments, we are struck by how little they pursue by way of patientic considerations. It is 'the remarkable fact' that moral responsibility judgments routinely operate in ignorance of persons' patiency that itself explains why patientic considerations, whether historical or not, give pause to the reactive attitudes. Nevertheless, it is important to keep this insight from being the exclusive property of moral responsibility skeptics/eliminativists, as is suggested by the final sentence in the passage above. The tectonic theory provides a way to explain why Harris' history mitigates his moral responsibility without having to be sceptical about moral responsibility. His is a relatively clear case of uncontrollable factors making an overwhelming contribution to his crimes in comparison to his personal quotient of self-control. In the vocabulary of the tectonic theory, this is described by saying that Robert Harris' patiency subducts his agency.

It is important to be clear that withholding moral responsibility from Harris does not entail that he should go free. He is far too dangerous for that. Although it is a common criticism of work

²⁵⁰ (Watson, 1987, pg. 144-145)

that is alert to patientic factors (such as Waller's) that it undermines justifications for punitive measures being taken on the part of society,²⁵¹ such criticisms are only apt if the theory in question seeks to generalize from a case like Robert Harris' (or by some other means) to the view that patiency is subductive of agency for all persons. To be fair, the majority of philosophical work that pays attention to patientic factors is of this mould.

Greene and Cohen, for example, compare ordinary persons to an imagined victim of manipulation, Mr. Puppet: "in a very real sense, we are all puppets. The combined effects of genes and environment determine all of our actions. Mr Puppet is exceptional only in that the intentions of other humans lie behind his genes and environment. But, so long as his genes and environment are intrinsically comparable to those of ordinary people, this does not really matter. We are no more free than he is."²⁵² Pereboom's famous four cases argument is also intended to challenge readers to point out what distinguishes ordinary people from clear victims of manipulation. He writes, "there are no differences between the manipulated agents and their ordinary deterministic counterparts that can justify the claim that the manipulated agents are not morally responsible while the determined agents are."²⁵³

The Strawsonian tradition is relevant to these claims. Strawson argues that moral responsibility practices are constituted by the tendencies of members of a moral community to respond to one another with the reactive attitudes. Since I am working within the Strawsonian tradition, I can

²⁵¹ (Smilansky, 2011)

²⁵² (Greene & Cohen, 2004, pg. 1780)

²⁵³ (Pereboom, 2013, pg. 58)

respond to Greene and Cohen or Pereboom by saying that people do in fact make distinctions between victims of manipulation and ordinary agents, and since these are constitutive of moral responsibility practices, a theory of moral responsibility should integrate them. Additionally, the tectonic theory has the resources to explain the distinctions in question. Manipulation cases are set up to illustrate that patiency can subduct agency, and the very fact that manipulations cases are resorted to in order to demonstrate this possibility is illustrative of the fact that most “ordinary” cases are not intuitively assessed to involve such subduction.

The tectonic theory, thus, is able to acknowledge patiency subducting agency when it occurs, whether in a manipulation case, a tragic real world case like that of Robert Harris, or in a more local way such as when someone suffers from Tourette’s syndrome, but need not insist that the features of such cases are global features of persons’ action. This allows the tectonic theory to yield different justifications for the incarceration of Robert Harris, for example, and more ordinary criminals. To elaborate, for the particular cases that the tectonic theory judges to involve patiency subducting agency, I endorse Pereboom’s justification for incarceration, which would assimilate a person like Harris to a category akin to natural disasters. Pereboom contends that a dangerous person’s incarceration can be likened to a quarantine.²⁵⁴ The difference between incarcerating someone like Harris as a quarantine-like initiative and doing so as a punitive measure consists in the way that the former lacks and the latter includes retributive motivation.

²⁵⁴ “Suppose an agent poses a danger to society by a demonstrably strong tendency to commit murder. Even if he is not in general a morally responsible agent in the basic desert sense, the state would nevertheless seem to have as much right to isolate him as it does to quarantine a carrier of a deadly communicable disease who is not responsible in this sense for being a carrier.” (Pereboom, 2013, pg. 72)

The tectonic theory, thus, enables the degree of retributivism involved in particular moral judgments to vary according to features of the case at hand. This is important because it allows for recognizing that retributivism is not appropriate when a person's patiency subduces her agency, while leaving the basic enterprise of retributive punishment sufficiently intact to justify retributive punishment if agentic factors are taken to subdue patientic ones. This is what makes the tectonic theory more agile than its competitors. Tectonic moral responsibility judgments leave people free to pursue moral responsibility judgments that are more or less retributive according to independent principles, insisting only that whatever extent of retributivism is deemed appropriate should be arrived at on the basis of a person's agency and patiency. The addition of the concept of subduction makes explicit in a way that doesn't seem to be available to other theories what it is about a person that is judged to be the case when a moral responsibility judgment pronounces the propriety or lack thereof of a retributivist response. Instead of resorting to the view that persons' lives are generally dominated by agency, as most proponents of moral responsibility suggest, or the view that they are dominated by patiency as most eliminativists claim, the tectonic theory can acknowledge that both types of cases occur.

Conclusion

This chapter had two aims. First, I presented the basic argument for the tectonic theory, which was intended to show that the tectonic theory can be justified on the basis of uncontroversial features of persons' actions. This move was supposed to show that the homogenization problem is not necessary to motivating the theory. Instead, the homogenization problem can be taken as indicative of what it is about moral responsibility practices that the tectonic theory can

ameliorate. The basic argument establishes what it is about persons' actions to which the tectonic theory attunes so as to enable amelioration of the homogenization problem.

The second goal of the chapter was to present the concept of subduction, drawn from geological work on tectonic plates, as a way to categorize different verdicts that the tectonic theory might give to different sorts of cases. By analyzing Vargas' work, I showed that the tectonic theory allows for recognizing that apparent threats to morally responsible action, such as situationism, can be forestalled on the basis of judging agency as capable of subducting the relevant patientic factors. Additionally, I showed that Vargas' general view about moral responsibility should be classified as taking agency to be generally subductive of patiency, which enriches the ways that such a theory can be evaluated.

Finally, I discussed two cases of human behaviour, to which intuition responds leniently in such a way that can be explained by recognition that patiency is subductive of agency in such cases. Just as with Vargas' work on situationist threats in particular, and moral responsibility in general, it was shown that patiency can be taken as subductive of agency in connection to a circumscribed set of factors, as with sufferers of Tourette's syndrome, or much more generally, as in the eliminativist arguments of Greene and Cohen and Pereboom.

Watson's probative treatment of the case of Robert Harris was given special attention because Harris' story provides a relatively clear case of patiency subducting agency. In contrast to Watson's analysis, the tectonic theory easily explains why Harris' upbringing makes trouble for the prospect of holding him morally responsible. Harris' upbringing was so horrific as to invite the judgment that these patientic factors in his life outweigh the contributions of the agentic powers he developed. The tectonic theory also handles the question of what to do with Harris comparatively easily. For such cases, Pereboom's quarantine-based justification for incarceration is natural. The tectonic theory suggests this on the basis of facts about Harris in particular, though, and therefore avoids culminating in general eliminativist conclusions. The tectonic theory, thus, integrates the major concerns of moral responsibility proponents and eliminativists, enabling the deploy of such concerns to be motivated by features of the persons subject to moral responsibility judgments, instead of motivated by the comparatively abstract and contested matters of what the universe is like and whether moral responsibility practices do it justice.

Chapter 6- The Moral Significance of Neuroscience

In this chapter, I will present what I take to be the implications of the tectonic theory of moral responsibility for the moral/legal significance of neuroscience. I focus on a debate instigated by Joshua Greene and Jonathan Cohen's provocative paper "For the law neuroscience changes nothing and everything".²⁵⁵ Greene and Cohen argue that neuroscience will clearly and accessibly demonstrate human behaviour to be mechanistic, leading to an abandonment of intuitive beliefs about free will that will undermine familiar retributive forms of moral responsibility. Once neuroscience has thusly transformed people's moral intuitions, Greene and Cohen think this will in turn re-orient our criminal punishment practices away from the aim of inflicting suffering on deserving transgressors and toward the goal of dealing with criminals humanely. Many have been skeptical. I will discuss responses from Dennett, Stephen J. Morse, and Pardo and Patterson.²⁵⁶ I will argue that the tectonic theory provides a way to integrate the concerns of both Greene and Cohen and their critics. Greene and Cohen suggest that neuroscience can be put into service making justice systems less punitive. I agree with this, but explain it in terms of the tectonic theory instead of in reference to the alleged pervasiveness of libertarianism. Greene and Cohen's critics all see Greene and Cohen's prediction about how the public will respond to neuroscience as implicitly committed to all sorts of regrettable social changes in addition to the alleged legal reform. A tectonic account of the moral significance of neuroscience can retain Greene and Cohen's basic idea about justice systems becoming less punitive while avoiding the pitfalls seized on by their critics.

²⁵⁵ (Greene & Cohen, 2004)

²⁵⁶ (D. Dennett, 2011), (Morse, 2011b), (Pardo & Patterson, 2013)

Greene and Cohen's Argument

In this section, I will show that Greene and Cohen's view of how neuroscience will bring about legal reform changes subtly as their paper goes on. At first, they appear to be committed to very strong claims about how neuroscience will change people's moral thinking, which in turn is supposed to spur the envisioned legal reform. Later, however, they concede that their early claims about how neuroscience will change people's moral thinking are likely too strong, but continue to insist that in a legal context, neuroscience can facilitate the adoption of a non-retributive outlook towards criminals on the part of judges and jurors. This results in a gap in Greene and Cohen's argument. Without the support of their early claims about neuroscience changing people's general moral outlooks, it is not clear how the envisioned legal reform is supposed to come about. Additionally, even if they were right about how neuroscience will affect people's general moral outlook, Greene and Cohen's critics point out many ways that this could turn out to be an undesirable result.

Nevertheless, I suggest that Greene and Cohen were on the right track insofar as they saw a need to reform the way people think about moral responsibility in order to make justice systems sufficiently responsive to neuroscientific facts to create an opportunity for a downturn in punitiveness. Greene and Cohen just misdiagnose the nature of the change in moral outlook that is required. Instead of the widespread dissolution of beliefs in free will, which would face the pitfalls mentioned by Greene and Cohen's critics, the tectonic theory's concern for patency can do the job. There are good reasons to believe this. Kurt Gray and his team have shown that emphasis on a person's patency is more likely to mitigate her responsibility than considerations

of her agency,²⁵⁷ and that patiency is more associated with people's bodies than is agency.²⁵⁸ So long as it is accepted that neuroscience is also more associated with the body than more traditional psychological ways of probing a case of moral responsibility, these findings explain how neuroscience could, under the purview of the tectonic theory, make for less punitive justice systems.

The claim that neuroscience will be more associated with the body than the mind relies on the idea that most people are intuitive dualists, a phenomena amply documented by Paul Bloom in *Descartes' Baby*.²⁵⁹ In other words, due to the fact that most people implicitly distinguish between the mind and the body (of which the brain is a part), and it is not yet commonplace to deduce a person's precise mental states from information about her brain, the presentation of neuroscientific information in courtrooms can be expected to be interpreted as pertaining to a person's body. Any impact of neuroscientific information on judges and jurors' reasoning about defendants' mental states will plausibly result from their own idiosyncratic views as to how a person's mind can and cannot influence her body, and how a person's mind can and cannot evade the influence of her body.

Greene and Cohen present the view that neuroscience's legal significance will follow from its intuitive moral significance as a way to meet the challenge created by Morse's trendsetting work on the subject of neuroscience and the law. Morse dubbed the tendency to morally absolve

²⁵⁷ (K. Gray & Wegner, 2011)

²⁵⁸ (K. Gray et al., 2011)

²⁵⁹ (Bloom, 2004)

criminals in response to neurological information about the causation of criminal behaviour the *fundamental psycholegal error*. The fundamental psycholegal error is “to believe that causation, especially abnormal causation, is per se an excusing condition.”²⁶⁰ This concept has two important elements. First, Morse argues that legal doctrine is concerned with an agent's rationality, and not whether she has free will. Second, despite legal doctrine being this way some judges, jurors and commentators mistakenly hold the view that neuroscientific knowledge about the causation of criminal behaviour should mitigate punishment for that behaviour.²⁶¹

Greene and Cohen's strategy is to elaborate upon the second element, while conceding the first. They argue that what Morse thinks of as an error on the part of some judges, jurors, and commentators is more like a moral vanguard, to which the rest of society will catch up as neuroscience becomes more sophisticated and accessible. Thus, Greene and Cohen argue that it may be presently legitimate for Morse to describe the fundamental psycho-legal error as a mistake, but that this is only because extant legal doctrine has yet to confront a public whose morality has been influenced by neuroscience.²⁶² This is why Greene and Cohen set out to predict that neuroscience will influence the law indirectly, through its influence on intuitive morality.

²⁶⁰ (Morse, 2004, pg. 180)

²⁶¹ “If you think that neuroscientific information about the causes of human action, or some particular human's action, can, by itself, make for a legitimate legal excuse, you just do not understand the law. Every action is caused by brain events, and describing those events and affirming their causal efficacy is of no legal interest in and of itself.” (Greene and Cohen, 2004, pg. 1778)

²⁶² “In our opinion, Morse and like-minded theorists are absolutely correct about the relationship between current legal doctrine and any forthcoming neuroscientific results. For the law, as written, neuroscience changes nothing ... But, we maintain, ... the legitimacy of the law itself depends on its adequately reflecting the moral intuitions and commitments of society. If neuroscience can change those intuitions, then neuroscience can change the law.” (Greene and Cohen, 2004, pg. 1778)

Greene and Cohen's account of how neuroscience will change people's moral outlooks consists of three elements. First, Greene and Cohen speak to what it is that neuroscience conveys about human behaviour so as to have its purported effect on intuitive morality. Mainly, Greene and Cohen use the language of mechanism for this, claiming that in the future neuroscientific demonstrations will be sufficiently sophisticated and accessible to show that human cognition is a fully mechanical process. The second element is Greene and Cohen's view regarding what it is that people believe now that neuroscience is supposed to change by illuminating the mechanical nature of human cognition. They claim that most people have libertarian intuitions about free will, such that human agents are pervasively understood to have the ability to do otherwise, which view Greene and Cohen take to be incompatible with the knowledge that human cognition is thoroughly mechanistic. Finally, Greene and Cohen claim that the retributive component of moral responsibility depends on the libertarian beliefs just mentioned. With all of this in place, Greene and Cohen argue that neuroscience will show more and more people that human cognition is mechanistic, which will cause them to abandon their intuitions about human free will, thus suppressing retributivist ways of thinking. Greene and Cohen envision the subsequent legal reform coming about because the law will need to become non-retributive in order to stay credible to such a public.

There are many claims made by Greene and Cohen that illustrate their initial aim of establishing that neuroscience has general implications for morality. In their abstract, Greene and Cohen describe a "change in moral outlook" that will result from neuroscience "transforming people's

moral intuitions about free will and responsibility.”²⁶³ In their introduction, they claim that “neuroscience will challenge and ultimately reshape our intuitive sense(s) of justice.”²⁶⁴ Neither of these remarks mentions criminal punishment, which makes it clear that Greene and Cohen start out with the idea that neuroscience has general implications for morality, and that the specific implications of neuroscience for criminal punishment flow from these. They write that,

Cognitive neuroscience, by identifying the specific mechanisms responsible for behaviour, will vividly illustrate what until now could only be appreciated through esoteric theorizing: there is something fishy about our ordinary conceptions of human action and responsibility, and that, *as a result*, the legal principles we have devised to reflect these conceptions may be flawed.²⁶⁵

Greene and Cohen's use of the phrase 'as a result' expresses that the legal implications of neuroscience are supposed to be enabled by the widespread recognition of the flaws in our ordinary conceptions of human action and responsibility.

Thus, Greene and Cohen put the bulk of their early efforts into explaining how neuroscience will change the way people think about human action so as to eventually delegitimize and deter retributivist thinking. They write, “New neuroscience will undermine people’s common sense, libertarian conception of free will and the retributivist thinking that depends on it.”²⁶⁶ The two claims here, that neuroscience will undermine libertarianism and that retributivism depends on widespread libertarianism, both suffer from a lack of resolution. This enables Greene and Cohen to encourage stronger and then weaker readings of them as their argument progresses, thus obscuring the abandonment of their original argumentative goals. Before making the case that this is how Greene and Cohen proceed, a brief word about their appeal to libertarianism is required.

263 (Greene and Cohen, 2004, abstract)

264 (Greene and Cohen, 2004, pg. 1775)

265 (Ibid, italics added)

266 (Greene and Cohen, 2004, pg. 1776)

Throughout this thesis, I have rejected the idea that libertarianism is the only doctrine that can legitimize moral responsibility. In chapter one, I explained that the tendency to think that libertarianism is needed results from a failure to realize that even if it is the case that libertarianism would satisfyingly legitimize moral responsibility practices, this does not provide any support for the claim that it is the only way to legitimize moral responsibility practices. This would be among Greene and Cohen's problems if it weren't the case that they ultimately retreat from their initial pronouncements. Any serious attempt to follow through on their project as initially stated would require that Greene and Cohen show that their prediction can also apply to a public whose acceptance of retributivist moral responsibility practices is based on non-libertarian beliefs about agency, or make a case that such a public could not exist. That Greene and Cohen fail to do this is a stroke against them, to be sure, but it is beside the point given the way that they eventually settle on a plea for neuroscience's importance in justice systems that is *not* a result of people abandoning intuitions about moral responsibility en masse.

My plan is to rehabilitate Greene and Cohen's work in terms of the tectonic theory, which is designed to operate in what might be called a post-libertarian world-view. Therefore, I will treat Greene and Cohen's argument as though it intended to apply to a public that has no special allegiance to libertarianism, essentially swapping their concept of free will for the basic concept of agency used throughout this thesis. This will allow me to work toward a more robust account of neuroscience's moral significance, and results in minimal and ultimately inconsequential distortion of Greene and Cohen's argument. For Greene and Cohen, the only account of agency

that can support retributivist morality is libertarian; taking their argument to apply to agency in general instead of just libertarian forms simply corrects this oversight on their part. What this means practically is just that I will be substituting different formulations of the phrase ‘morally responsible agency’ for ‘free will’ when citing or discussing Greene and Cohen.

I should note that this is not quite exegetically fair to Greene and Cohen. They do embrace a form of moral responsibility that is entirely consequentialist: “The consequentialist approach to responsibility generates a derivative notion of free will that we can embrace ... In the name of producing better consequences, we will want to make several distinctions among various actions and agents ... If we like, then, we can say that the actions of rational people operating free from duress, etc. are free actions, and that such people are exercising their free will.”²⁶⁷ Greene and Cohen are aware that this is nearly a compatibilist view of the sort they wish to avoid: “At this point, compatibilists such as Daniel Dennett may claim victory: ‘what more could one want from free will?’. In a word: retributivism.”²⁶⁸ The idea is that a consequentialist form of moral responsibility would be just like what we have now except that it would not invoke the concept of desert. There are two major problems with this idea. First, it is dubious whether holding someone responsible for reasons other than her desert qualifies as moral responsibility. My view is that the concept of desert is essential for moral responsibility.

Second, Greene and Cohen’s eventual concession that retributivist thinking is not likely to be wholly suppressed by neuroscience means that some other way of fostering the wholly

²⁶⁷ (Greene and Cohen, 2004, pg. 1783)

²⁶⁸ (Greene and Cohen, 2004, pg. 1783)

consequentialist view is required. Greene and Cohen do not fill this gap, and even if they did somehow, their reluctant acceptance of the persistence of retributivist thinking suggests that however it is that the wholly consequentialist approach is motivated, it will still be subject to retributivist inflection, potentially resulting in officially non-retributive but nevertheless cruel forms of punishment. Dennett makes this point, writing that,

What is the alternative [to compatibilist forms of morally responsible agency] proposed by Green and Cohen? Consequentialist medicalization, in effect, which replaces punishment in all its forms with . . . treatment and education designed to encourage a revision of one's goals. Greene and Cohen think that this sort of consequentialism is humane, and it is clearly more humane than the horrific vindictiveness of Kant's retributivism, but otherwise I think it is quite horrible in its own right. It is, or generally leads to, totalitarianism. Think of all the people in the Soviet Union who, it was decided by the state, needed "curing" of the mental disorders that caused them to commit anti-state acts.²⁶⁹

Given these severe difficulties facing Greene and Cohen's positive claims about wholly consequentialist moral responsibility practices, I will ignore this part of their argument in favour of focusing on their negative claims about neuroscience making trouble for desert-based forms of moral responsibility.

Returning to the lack of resolution in Greene and Cohen's claims, the idea that neuroscience will undermine intuitions about persons having morally responsible agency can express either that neuroscience will destroy or just weaken these intuitions. Likewise, the claim that retributivism depends on libertarianism can express either a strict or weak dependency.²⁷⁰ Greene and Cohen start out encouraging the strong readings of both of these claims, but eventually settle on the weaker readings. I can ignore the second of these because of the way I am considering Greene

²⁶⁹ (D. Dennett, 2011)

²⁷⁰ A strict dependency between the two would be expressed as a bi-conditional: retributivism occurs if and only if libertarianism is in bloom, and so whenever libertarianism is absent, so too is retributivism. A weak reading would cast libertarianism as just one among retributivism-enabling rubrics, making no claim about retributivism in the absence of libertarianism. See Chapter one for a fuller discussion of this distinction.

and Cohen's argument as though it targeted any morally responsible agency without insisting that such a thing must be as libertarians claim. After swapping out Greene and Cohen's libertarian free will for a non-metaphysical account of morally responsible agency, the dependency of retributivist moral responsibility upon the viability of morally responsible agency is assured. I will therefore focus on Greene and Cohen's drift from claiming that neuroscience will abolish people's beliefs in morally responsible agency to claiming that neuroscience will just facilitate episodic adoption of a non-retributive stance toward criminals on the part of judges and jurors.

Greene and Cohen make several remarks in the early sections of the paper that suggest they envision neuroscience as eradicating rather than just weakening intuitions about morally responsible agency. In their introduction, they write, "the net effect of this influx of scientific information will be a *rejection* of [morally responsible agency] as it is ordinarily conceived, with important ramifications for the law."²⁷¹ Greene and Cohen's use of the word 'rejection' seems to indicate that they envision a time wherein persons are decidedly without ordinary conceptions of morally responsible agency. A little later they describe recognition that morally responsible agency is an illusion as an impetus for the rejection of retributivist legal principles.²⁷² This also suggests that Greene and Cohen meant their claim about neuroscience undermining people's beliefs in morally responsible agency in the strong sense.

271 (Greene and Cohen, 2004, pg. 1776, italics added)

272 "We can, however, recognize that [morally responsible agency], as conceptualized by the folk psychology system, is an illusion and structure our society accordingly by rejecting retributivist legal principles that derive their intuitive force from this illusion." (Greene and Cohen, 2004, 1776)

The main reason for taking this to be Greene and Cohen's position, however, is that their overall argument (as initially laid out) requires the undermining claim to be meant in the strong sense. The basic idea of neuroscience's legal significance flowing from the way it changes people's moral intuitions would require that people stop believing that they and others have morally responsible agency. Greene and Cohen write, "We maintain ... that our operative legal principles exist because they more or less adequately capture an intuitive sense of justice. In our view, neuroscience will challenge and ultimately reshape our intuitive sense(s) of justice."²⁷³ It is clear that Greene and Cohen locate the site of neuroscience's influence in intuitive morality, which in turn is supposed to change justice systems.²⁷⁴

Greene and Cohen's vision of a neuroscientifically-enlightened justice system is comprehensively consequentialist. They suggest that the concept of desert will stop playing a role in future justice systems, and that this will be prompted by the public having already stopped using it.²⁷⁵ If Greene and Cohen did not intend to claim that neuroscience will undermine intuitions about morally responsible agency in the strong sense, it is difficult to see how enough pressure could be put upon justice systems to result in the drastic reform they envision. If the public is still at all prone to interpreting one another's behaviour as subject to deserved moral responsibility, this prediction falters severely.

²⁷³ (Greene and Cohen, 2004, pg.1775)

²⁷⁴ "The legitimacy of the law itself depends on its adequately reflecting the moral intuitions and commitments of society. If neuroscience can change those intuitions, then neuroscience can change the law." (Greene and Cohen, 2004, pg. 1778)

²⁷⁵ "With the rejection of common-sense conceptions of [morally responsible agency] comes the rejection of retributivism and an *ensuing* shift towards a consequentialist approach to punishment, i.e. one aimed at promoting future welfare rather than meting out just deserts." (Greene and Cohen, 2004, pg. 1778, italics added)

Greene and Cohen's weakening of the undermining claim is subtle at first, but by the end of the paper, they openly concede that neuroscience might *not* change people's everyday moral thinking as initially claimed. Greene and Cohen concede that "our intuitive sense of free will runs quite deep, and it is possible that we will never be able to fully talk ourselves out of it."²⁷⁶ The implication is that we may be able to *partially* talk ourselves out of our free will intuitions, which suggests that neuroscience will undermine intuitions about morally responsible agency in the weak sense of reducing their power. This is reinforced by Greene and Cohen in a footnote where the claim that the relevant intuitions themselves will be undermined by neuroscience gets repackaged as the claim that the force of these intuitions will be undermined.²⁷⁷ Greene and Cohen also go on to suggest that the tension between intuitions about persons being morally responsible agents and the knowledge that persons are natural entities may never dissipate. They write that "The problem of free will and determinism will never find an intuitively satisfying solution because it arises out of a conflict between two distinct cognitive subsystems."²⁷⁸ The lack of a satisfying solution for the problem of disparate cognitive subsystems plausibly entails the persistence of intuitions about morally responsible agency throughout neuroscience's development. If neuroscience really did destroy such intuitions, then this should resolve the conflict in question because it would mean that we are no longer construing one another as anything but mechanical systems. That Greene and Cohen predict the conflict will endure suggests that intuitions about moral responsibility will endure too, which suggests that neuroscience's undermining of them must be weaker than required by their argument.

276 (Greene and Cohen, 2004, pg. 1781)

277 "Our suggestion is simply that neuroscience will eventually advance to the point at which the mechanistic nature of human decision-making is sufficiently apparent to undermine the force of dualist/libertarian intuitions." (Greene and Cohen, 2004, f.n. 6)

278 (Greene and Cohen, 2004, pg. 1783)

Greene and Cohen drift farther from their stated aims when it comes time to defend their argument against objections. When confronting the near orthodox view that an institution of desert-based moral responsibility is necessary to keep civilization civilized, Greene and Cohen suggest that this concern is overblown and move on to the allegedly more moderate view of Daniel Wegner. Greene and Cohen write, “In a less hysterical vein, Daniel Wegner argues that free will, while illusory, is a necessary fiction for the maintenance of our social structure.”²⁷⁹ In response to the necessary fiction claim, Greene and Cohen write, “We disagree. There are perfectly good, forward-looking justifications for punishing criminals that do not depend on metaphysical fictions. (Wegner’s observations may apply best to the personal sphere.)”²⁸⁰ Greene and Cohen’s concession that Wegner might be right about the personal sphere (even though it's tucked away in parentheses) is fatal to their objective as stated at the outset. Once they remove the personal sphere from neuroscience's influence, it is difficult to see how neuroscience will effect enough of a change in people's general moral outlooks to prompt the drastic legal reform that Greene and Cohen envision.

Having retreated from the strong reading of the undermining claim that was needed for their argument, Greene and Cohen start to chart a new course. Rather than explaining how people will stop thinking in retributive ways altogether, they now claim that neuroscience will perform a far subtler social surgery, replacing retributivism with a more compassionate outlook just when criminal punishment is at issue. The problem with this is that Greene and Cohen do not provide any means to explain why neuroscientific revelations should matter inside but not outside of

279 (Greene and Cohen, 2004, pg. 1783) Greene and Cohen refer to (Wegner, 2002, ch. 9)

280 (Greene and Cohen, 2004, pg. 1783)

courtrooms. Instead of defending their earlier claims that commitment to morally responsible agency will gradually disappear resulting in a public that insists on extending to criminals the non-retributive moral outlook that will by then be widely enjoyed outside of courtrooms, Greene and Cohen acknowledge that there are many good reasons to accept the view that retributivism is an inevitable aspect of human life.²⁸¹

At this eleventh hour, Greene and Cohen start looking for something special about criminal punishment that explains why neuroscience might have a practical role here that it lacks in day-to-day life. In doing so, Greene and Cohen draw an analogy between the sorts of counter-intuitive truths about reality that physics has delivered, and the neuroscientific wisdom they recommend be heeded by judges and jurors of the future. They write that, “modern physics tells us that space is curved. Nevertheless, it may be impossible for us to see the world as anything other than flatly Euclidean in our day-to-day lives,” and then ask if it follows “that we are forever bound by our innate Euclidean psychology?”²⁸² Greene and Cohen don't think so, and claim that to remain guided by our innate psychological dispositions is a choice that can be made according to the situation at hand.

281 Greene and Cohen mention three streams of evidence for the view that retributivism is an inevitable part of human life: philosophical, evolutionary, and neuroscientific. “Regarding responsibility and punishment, one might wonder if it is humanly possible to deny our retributive impulses ... This challenge is bolstered by recent work in the behavioural sciences suggesting that an intuitive sense of fairness runs deep in our primate lineage ... and that an adaptive tendency towards retributive punishment may have been a crucial development in the biological and cultural evolution of human sociality ... Recent neuroscientific findings have added further support to this view, suggesting that the impulse to exact punishment may be driven by phylogenetically old mechanisms in the brain ... These mechanisms may be an efficient and perhaps essential, device for maintaining social stability. If retributivism runs that deep, one might wonder whether we have any serious hope of, or reason for, getting rid of it. Have we any real choice but to see one another as free agents who deserve to be rewarded and punished for our past behaviours?” (Greene and Cohen, 2004, pg. 1784) See the previous chapter for mention of their sources.

282 (Greene and Cohen, 2004, pg 1784)

Greene and Cohen's limited support for this claim comes from the way that rocket scientists on the job manage to conceive of reality according to “relativistic physical principles that are less intuitive but more accurate.”²⁸³ They reason that if “a Euclidean perspective is not necessary for all practical purposes,” then “the same may be true for our implicit commitment to [morally responsible agency] and retributivism.”²⁸⁴ Once Greene and Cohen start to extract the moral from their analogy between relativity and neuroscience, there is no longer any trace of the strong undermining claim made earlier. They write,

For most day-to-day purposes it may be pointless or impossible to view ourselves or others in this detached sort of way. But—and *this is the crucial point*—it may not be pointless or impossible to adopt this perspective when one is deciding what the criminal law should be or whether a given defendant should be put to death for his crimes.²⁸⁵

Prior to this section, Greene and Cohen had been arguing that viewing ourselves and others in the “detached sort of way” will be made irresistible by neuroscientific advances and that this is all that will be required to suppress retributivist thinking. Now that they have conceded that neuroscience will not eradicate intuitions about morally responsible agency, Greene and Cohen's positive argument for legal reform is just an allusion to undiscovered reasons that it would be appropriate to adopt a detached neuroscientifically enlightened perspective when criminal punishment is at issue.

Even though it is not what Greene and Cohen set out to argue for, the suggestion that there are reasons to foster sensitivity to neuroscientific facts in criminal punishment situations that are not

283 “The answer depends on the domain of life in question. In navigating the aisles of the grocery store, an intuitive, Euclidean representation of space is not only adequate, but probably inevitable. However, when we are, for example, planning the launch of a spacecraft, we can and should make use of relativistic physical principles that are less intuitive but more accurate.” (ibid)

284 (ibid)

285 (ibid, italics added)

present in daily life is plausible. However, an enhanced sensitivity to neuroscience in the courtroom can still benefit from alliance with theorizing about moral responsibility. If there were a way to reform theories of moral responsibility such that neuroscientific facts worked with them instead of against them, this would mean that judges and jurors do not need to depart from their everyday moral outlook in order to confront the possibility that neuroscientific facts about a criminal should mitigate her moral/legal culpability. Such a schema is both simpler and more psychologically realistic than Greene and Cohen's vision of institutionalized moral detachment towards criminals that is akin to rocket scientists' adoption of a non-euclidean framework. Additionally, it would preserve the idea that the law's legitimacy involves congruence with the moral thinking of the citizenry, which was an important and intuitive part of Greene and Cohen's initial argument.

I now turn to how the tectonic theory can rehabilitate Greene and Cohen's abandoned path along the lines just described. The tectonic theory's concern for patiency, although not motivated by neuroscientific facts, can explain what it is about such facts that holds the promise of fostering a more compassionate and less punitive justice system. As Greene and Cohen eventually concede, it is almost certainly not the case that neuroscience will eradicate moral responsibility practices, but if such practices were to independently become tectonic, neuroscientific contributions would fall into place illuminating predominantly patientic facts about a person. After developing these ideas, I will show how they evade the problems delineated by Greene and Cohen's critics.

A tectonic account of neuroscience's moral and legal significance.

The tectonic theory's integration of the concept of patiency provides a promising way to explain the moral significance of neuroscience. Instead of taking neuroscience to undermine moral responsibility by demonstrating that persons are not the right sort of thing to be held morally responsible, a tectonic account can enlist neuroscience to combat the homogenization problem discussed in chapter four. In other words, instead of showing that persons do not have enough control over their behaviour to be held morally responsible (which following Strawson, I take to be a non-starter), neuroscience can enhance the fairness of moral/legal culpability judgments by recovering information about what a given person does not control.

The first reason to expect a large portion of neuroscientific facts to convey patientic information is that neuroscientific facts are first and foremost about a person's body, and these are more patientic than agentic. To be fair, the recent bloom of philosophical interest in neuroscience is motivated by the idea that neuroscientific facts, though biological, are revelatory with respect to features of persons' mental lives. However, this conception of neuroscience as straddling the line between the physical and the mental is highly theory-laden. Some of the philosophers most familiar with neuroscience would characterize this line as temporary, likely to vanish as conceptions of mental features are gradually assimilated to a neuroscientific account of cognition.²⁸⁶ Others insist that neuroscientific claims must be subject to formulation in terms of mental features to have any moral significance.²⁸⁷

²⁸⁶ Paul and Patricia Churchland are the most well-known eliminativists about mental states. See (Paul Churchland, 1981), and (Patricia Churchland, 1987)

²⁸⁷ See (Morse, 2011b), (Sifferd, 2006)

My view intends to occupy a middle ground between these. For broadly Strawsonian reasons, I do not think that neuroscience holds the promise of altering the way that persons interpret one another's behaviour in terms of mental states. I agree with the Strawsonian claim that consideration of a persons' good or ill will is an inevitable fixture of life in a community of moral responsibility. However, I disagree with the claim that neuroscience must be translated into such terms in order to have moral significance. The tectonic theory allows for a third way. A person's agentic actions can still be parsed according to familiar mental state concepts like good or ill will, and the contributions of neuroscientific facts can be determined by investigating whether they qualify as patientic, and whether they are relevant to the occurrence of the action in question. At this time, there is not much a person can do to modify her neurology,²⁸⁸ and so most neuroscientific facts about a person will illuminate patientic things about her that she cannot change.

Additionally, as noted by both Greene and Cohen and Stephen J. Morse, neuroscience is simply not needed to judge whether a person has a degree of self-control that qualifies her as an agent that can be held responsible.²⁸⁹ Old fashioned, folk-psychology-based interpersonal observation

²⁸⁸ This might change as enhancement drugs/treatment come into the picture. Nicole A. Vincent wrote a paper investigating whether in the future, enhanced individuals should be subject to greater moral responsibility; see (Vincent, 2013). To say anything useful about the prospect of post-human enhancements for the tectonic theory would require a focused discussion. I plan to undertake this task in future work.

²⁸⁹ Greene and Cohen (2004) and Morse (2011b) both make this point in relation to a neuroscientific study of adolescent brains that swayed judged and jurors to advocate against use of the death penalty for adolescents on account of the relatively underdeveloped executive control systems in their brains. Both authors point out that the neuroscience is not necessary to make the point that adolescents suffer from undeveloped executive control, which point can made on the basis of behavioural observations of adolescents. Morse suggests that the relevant judges and jurors exemplified the fundamental psycho-legal error. Greene and Cohen characterize them as having implicitly gleaned that persons are deterministic systems from the neuroscience, which in turn prompted the leniency. The

is typically sufficient to sort agents from non-agents. This does not entail that neuroscience cannot be used to make more fine grained distinctions between persons' agency than are presently on offer. The claim is just that neuroscience is not needed to categorize persons as agents. To illuminate patientic aspects of persons, in contrast, neuroscience can be more useful because neurological features of persons may not be as clearly expressed in persons' behaviour as is their basic level of agency.²⁹⁰

Also, there is empirical work from Gray and his team that shows patiency to countenance bodily features of a person in a way that agency does not. In a paper called "More Than a Body: mind perception and the nature of objectification,"²⁹¹ Gray et al. perform several studies that explore the nature of objectification. They show that traditional conceptions of objectification, according to which it involves a de-emphasis or denial of psychological traits (dementalization), are flawed. Instead the authors find that when a person's body is emphasized, subjects attribute more patiency and less agency to that person. Importantly, this involved subjects finding more "objectified" people to be less morally responsible and more susceptible to harm.

The studies performed all work with the two-dimensional model of mind attribution discussed in chapter three: attributed mental states can be either agentic or experiential, which correspond respectively with agency and patiency. In support of the hypothesis that patiency and the body

tectonic theory's concept of patiency allows me to split the difference between these perspectives, which this chapter is intended to demonstrate.

²⁹⁰ The discussion of Mr. Oft in chapter two is a good example of how neuroscience can bring to light neurological features of persons that lie outside their control, i.e. are patientic, and cannot be easily ascertained through the use of folk-psychology-based behavioural analysis.

²⁹¹ (K. Gray et al., 2011)

are linked, they note that flesh seems to be required for attributions of patiency. Experimental subjects are typically not disposed to attribute patientic characteristics to machines or corporations, while they do attribute agentic features to such things.²⁹² The authors go as far as claiming that naive dualism may be less a distinction between mind and body and more a distinction between the agentic and experiential/patientic aspects of an individual, both of which involve mindedness but the latter of which also implicates the body.²⁹³

One method used in the studies operationalizes emphasis on bodies by showing subjects pictures of people with more or less or no clothing and querying them as to which characteristics the pictured would be likely to have. Another method simulates a Milgram-like set-up where subjects determine how many mild shocks would be appropriate to administer to a man (with or without a shirt on). Additionally, Gray et al. presented subjects with vignettes involving two people one of whom is emphasized in his agency while the other is described with bodily details then forcing subjects to choose which would be likely to act/suffer in a crisis situation. All studies bore out the authors' hypothesis about objectification, with emphasis on the body bolstering patientic attributions and reducing agentic ones.

²⁹² (Knobe and Prinz, 2008)

²⁹³ "rather than a conflict between a physical object and an immaterial soul, dualism may be a conflict between rational agency ("mind") and the seething passions of experience ("body")." (K. Gray et al., 2011, pg. 2) Note that the use of 'seething passions' reinforces the idea that a person's experiential aspect is a patientic phenomenon, i.e. not subject to her control.

These results suggest that emphasis upon a person's body induces a heightened sensitivity toward her experiential aspect, which partially comprises her patiency. The authors conclude that in addition to complicating the picture of objectification, their findings

also have implications for the moral domain. Ascriptions of moral responsibility are linked to perceptions of agency, so a body focus should decrease perceptions of moral responsibility. Alternatively, ascriptions of moral rights are linked to perceptions of experience ... so if a body focus increases perceived experience, it should also lead to a greater perceived entitlement to avoid harm.²⁹⁴

This suggests that the focus on persons' agency typical of non-tectonic moral responsibility judgments shifts focus to a person's mind and away from her body. The agentic bias and the homogenization problem that results from it, therefore, can thus be characterized in part as neglectful of bodily facts about a person. Neuroscience thus can plausibly contribute to enhancing the fairness of moral responsibility judgments by injecting content about persons' bodies to the patientic considerations that the tectonic theory puts in place to countervail the agentic bias.

Working with Wegner, Gray also showed that patiency is a more effective mitigator of a persons' moral responsibility than agency. In a paper called, *To Escape Blame, don't be a hero – be a victim*, Gray and Wegner investigate which way of escaping blame is more effective: presenting oneself as a hero (calling attention to past good deeds) or presenting oneself as a victim (calling attention to sufferings over the course of a life). Their moral typecasting thesis suggests that the latter should be more effective, and the former should be ineffective or even harmful with respect to blame-mitigation. Gray and Wegner set out to test this while guarding

²⁹⁴ (K. Gray et al., 2011, pg. 3)

against the possible confound that victimhood may be a more effective mitigator of blame than heroism on the grounds that victimhood stimulates a more emotional/sympathetic response.

In study 1, subjects were presented vignettes describing a man named George who either regularly gives some pay to charity (the hero condition), or else has the money stolen by a coercive supervisor (the victim condition), or else spends normally (the neutral condition), and then asked about the blame-worthiness of George when he fails to return some money to a stranger who has dropped it, and whether they sympathize with George. Assignments of blame were much lower in the victim condition and slightly higher in the hero condition in comparison with the neutral condition. In study 2, subjects were given vignettes involving a pair of employees that were given contrasting histories (victims had been hit by a drunk driver in the distant past, heroes had started a charity, and neutrals worked in a hardware store) and then described as having neglected a customer's request for a peanut-free salad, almost killing her. Subjects were asked which of the employees should be held more responsible and which should be fired. Results support the moral typecasting-derived hypothesis: heroes were more, neutrals were less, and victims much less responsible.

These results suggest that when patientic features of a person's life are emphasized, this decreases assignments of moral responsibility. Although these results provide support for the basic claim made on behalf of the tectonic theory – that patientic considerations can result in less punitive moral responsibility judgments – they also suggest that this effect can be irrational. For example, it would be difficult to philosophically defend the idea that having been hit by a drunk

driver long ago should reduce one's degree of moral responsibility for neglectfully serving peanuts to someone who is allergic. This would be problematic for the tectonic theory if it didn't have measures in place to guard against such effects. In the first place, the tectonic theory's concern for patiency is related to the actions for which persons might be held morally responsible. In chapter two, I presented a definition of patiency as it figures in tectonic moral responsibility judgments that insists on relevance to the performance of particular actions, and defended this structure against worries about irrelevant patientic factors.

More broadly, the two-step structure of the tectonic theory is designed to prevent agentic and patientic considerations from distorting or obscuring one another. The subjects in Gray and Wegner's studies were not aware of whether they were encountering a description of a more agentic/more patientic person, and were thus not provided with an opportunity to balance these considerations. Gray and Wegner's findings illustrate the different ways that agentic and patientic considerations contribute to intuitive moral responsibility thinking, as discussed in detail in chapter three, but do not have implications as to how such considerations relate to one another in conscious deliberative reflection. The two step structure of the tectonic theory insists on separate consideration of a person's agency and patiency, leaving it to individuals to decide for themselves which, if either, was more important in connection to the relevant action.

This leaves room for disagreement about particular cases, but will render the disagreement in terms of the tectonic theory (agency, patiency and subduction). This enables the grounds of particular moral responsibility judgments to be brought to light and compared to other

possibilities in relatively nuanced ways in comparison to agency-centric theories of moral responsibility. This should minimize the contributions of the typecasting tendencies discovered by Gray and Wegner, while simultaneously exploiting the distinctness of agentic and patientic considerations so as to enhance moral responsibility judgments' sensitivity to the nuances of particular cases. I now turn to showing how the tectonic theory can account for the moral significance of neuroscience in a way that evades the problems that critics saw in Greene and Cohen's proposal.

Defending the tectonic account of neuroscience's moral significance

The main problem that critics have with Greene and Cohen's proposal concerns what I have called their initial argument. Greene and Cohen set out to argue that neuroscience will do away with dualistic and libertarian intuitions, which will in turn prompt consequentialist legal reform. Their critics unanimously see such a prospect as gravely problematic. Morse directly describes Greene and Cohen's vision of the future as dystopian,²⁹⁵ Pardo and Patterson accuse Greene and Cohen of 'denying our shared humanity,'²⁹⁶ and Dennett invokes *A Clockwork Orange* to provide a sense of what he thinks is wrong with Greene and Cohen's vision.²⁹⁷

²⁹⁵ "Although I predict that we will see far more numerous attempts to introduce neuroevidence in the future, the dystopia that Greene and Cohen predict is not likely to come to pass." (Morse, 2011, pg. 546)

²⁹⁶ "We share what we believe to be the sentiment of Greene and Cohen that criminal punishment ought to be more humane, but we do not believe that the way to get there is by denying our shared humanity." (Pardo & Patterson, 2013, pg. 152)

²⁹⁷ "To a first approximation, the reason [that Greene and Cohen's consequentialist vision is problematic] has been well known for many years. One of the most vivid portrayals of it is Stanley Kubrick's movie, *A Clockwork Orange*, and the novel it came from by Anthony Burgess." (D. Dennett, 2011)

These criticisms do not apply to the tectonic account of neuroscience's moral significance because the tectonic theory doesn't purport to reform or eliminate conceptions of agency that legitimize moral responsibility. All of the critics insist that retributivism/desert is an important part of justice proceedings,²⁹⁸ and the tectonic theory, although designed to create opportunities for lessening the contributions of retributivist thinking to moral/legal responsibility judgments, does not aim to do away with retributivism. Still, Greene and Cohen's critics voice other concerns about taking neuroscientific facts into account in moral and legal responsibility judgments that can be instructively addressed by the tectonic theory.

The tectonic theory leaves it to individuals to decide for themselves how significant agentic and patientic factors are in particular cases. Such judgments will be subject to idiosyncrasies of those doing the judging, which could result in problems. The fate of persons on trial would be in part determined by judges and jurors' intuitions about subduction or lack thereof between the relevant agentic and patientic factors. Pardo and Patterson voice a similar concern about Greene and Cohen's argument. They write, "the first problem with [Greene and Cohen's] argument is the assumption that the intuitions of most people necessarily answer the normative questions of whether criminal punishment is justified or how it ought to be distributed ... [but] it is possible

²⁹⁸ "[the] criminal justice system as we know it ... includes robust notions of personhood and desert," (Morse, 2011, pg. 546)

"even if [judges and jurors] ... abandon retributivism as a basis for justifying punishment; Greene and Cohen are wrong to suppose that we would be "lucky" and punishment would necessarily be more "humane." ... a brief history of actual criminal-sentencing practices in the United States over the last thirty years suggests the converse to be true ... abandoning retributivist rationales for punishment in favor of deterrence, incapacitation, and general crime control has paved the way for harsh sentences for drug crimes, prosecuting juveniles as adults, strict liability crimes, proposals to abolish the insanity defense, and the felony-murder rule. The absence of retributivist constraints also allows for the indefinite "civil commitment" of criminals after the completion of their prison sentences." (Pardo and Patterson, 2013, pg. 151)

"We do not want to get rid of retributivism root and branch, we just want to tame it and reform it." (Dennett, 2011)

for widely shared intuitions about what is just punishment to be mistaken.”²⁹⁹ The tectonic theory of moral responsibility is subject to this worry but in a very different way than is the case with Greene and Cohen’s proposal.

Greene and Cohen describe neuroscience as an alternative to argumentation as a way of aligning citizen intuitions with facts about persons. They write,

not everyone cares for philosophical arguments. Arguments are nice, but physical demonstrations are far more compelling. What neuroscience does, and will continue to do at an accelerated pace, is elucidate the ‘when’, ‘where’ and ‘how’ of the mechanical processes that cause behaviour. It is one thing to deny that human decision-making is purely mechanical when your opponent offers only a general, philosophical argument. It is quite another to hold your ground when your opponent can make detailed predictions about how these mechanical processes work, complete with images of the brain structures involved and equations that describe their function.³⁰⁰

With respect to beliefs in libertarian free will, Greene and Cohen might be right that sophisticated neuroscientific demonstrations of the future will erode such commitments in a way that does not require complex argumentation. This would be a welcome development because libertarian commitments could work against the consideration of patiency prescribed by the tectonic theory, as discussed in chapter one.

However, the prospect of libertarian commitments evaporating from the public does not have the wide-ranging moral implications that Greene and Cohen allege. Such a development would help the tectonic theory get off the ground, but might not be necessary. The literature from experimental philosophers on the abstract/concrete paradox discussed in chapter one suggests that even people who profess libertarian commitments are also prone to more Strawsonian moral

²⁹⁹ (Pardo and Patterson, 2013, pg. 144)

³⁰⁰ (Greene and Cohen, 2004, pg. 1781)

responsibility judgments that are primarily reactions to persons' deeds, not their metaphysical natures.

In contrast to Greene and Cohen's view that neuroscience will succeed where naturalistic philosophical arguments sometimes fail,³⁰¹ the tectonic theory provides a philosophically motivated rethinking of the nature of the person in terms of both agency and patiency.

Neuroscientific facts, as argued above, can be expected to do a very different sort of work than that supposed by Greene and Cohen. Instead of undermining intuitions about persons having morally responsible agency, neuroscience can serve as an illuminator of a person's patiency.

Although people's intuitions will play a role in determining whether a particular factor qualifies as agentic/patientic or subductive or not, the tectonic theory provides a vocabulary with which to interrogate and refine such intuitions, as discussed above.

Pardo and Patterson's worry in pressing Greene and Cohen's appeal to people's intuitions is that the envisioned shift is supposed to bypass philosophical arguments. They write, "the key issue is whether neuroscience will contribute evidence that provides epistemic support for arguments concerning determinism, compatibilism, moral blameworthiness, and just punishment."³⁰² In contrast to Greene and Cohen, the tectonic theory provides philosophical motivation for considering persons' patiency that does not rely on neuroscience, but which yields a context in which neuroscientific facts can be relatively easily integrated into judgments about particular cases. Even though individuals' possibly problematic intuitions about agency and patiency play a

³⁰¹ "we argue that neuroscience can help people appreciate the mechanical nature of human action in a way that bypasses complicated arguments." (Greene and Cohen, 2004, pg. 1780)

³⁰² (Pardo and Patterson, 2013, pg. 144)

role in the tectonic theory, they are subject to philosophical pressures that can expose and refine them.

Pardo and Patterson also charge Greene and Cohen with implying an untenable view of folk psychology.³⁰³ Greene and Cohen describe folk-psychology as ‘the gateway to moral evaluation’ and claim that folk psychological interpretations of a person’s behaviour depend on attributing mind to that person.³⁰⁴ When Greene and Cohen claim that neuroscience will become decisively persuasive about the mechanical nature of human cognition, they intend this prediction to describe an alternative to the habit of interpersonal mindedness attributions. Pardo and Patterson take issue with this on the ground that folk psychology is a crucial part of both everyday life and legal doctrine.³⁰⁵

Additionally, Pardo and Patterson challenge Greene and Cohen’s conception of mindedness as involving uncaused causation.³⁰⁶ Pardo and Patterson take this claim as entailing a commitment on Greene and Cohen’s part to the view that folk psychological judgments don’t track any genuine phenomena.³⁰⁷ I do not think that this is fair to Greene and Cohen,³⁰⁸ but the importance

³⁰³ “Neither determinism in general nor neuroscience in particular undermines folk psychology in the ways [Greene and Cohen] presuppose.” (Pardo and Patterson, 2013, pg. 149)

³⁰⁴ “folk psychology is the gateway to moral evaluation. To see something as morally blameworthy or praiseworthy (even if it is just a moving square), one has to first see it as ‘some- one’, that is, as having a mind.” (Greene and Cohen, 2004, pg. 1782)

³⁰⁵ “reflect for a moment on what it would mean for folk psychology generally to be an illusion (i.e., that there are no such things as beliefs, desires, wants, fears, knowledge, intentions, or plans). One obvious implication is that there is no difference between us ... and someone engaged in bodily movements caused by a seizure or conditions beyond their “rational” control. A second implication is that psychological explanations would be false, and there would be nothing real (or nothing that affects behavior) for psychology to explain. A third implication is that human action would indeed cease to be “human” or “action,” at least as we currently conceive of these notions. Thus, if the argument by Greene and Cohen depends implicitly on these more radical claims about folk psychology, then this provides a *reductio ad absurdum* for their argument.” (Pardo and Patterson, 2013, pg. 151)

³⁰⁶ “intuitively, a mind is, among other things, an uncaused causer.” (Greene and Cohen, 2004, pg. 1782)

³⁰⁷ Pardo and Patterson reconstruct Greene and Cohen’s thinking thusly: “(1) Retributivist punishment depends on moral evaluation; (2) moral evaluation requires that people punished had some control over their actions; (3) to have

of folk psychology to moral/legal responsibility judgments is a common theme in discussion about neuroscience and the law/morality. It plays a central role in Morse's response to Greene and Cohen, and in others' work on the subject.³⁰⁹ I should explain, therefore, how the tectonic theory characterizes folk psychological judgments.

Morse's critique of Greene and Cohen is predicated on the idea that Greene and Cohen present neuroscientific construals of persons as alternatives to folk psychological appraisal. Like Pardo and Patterson, Morse reads Greene and Cohen as denying that folk psychological judgment is a viable way of explaining people's behaviour, and advocating for replacing folk psychological judgments with mechanistic explanations based on neuroscience. Contemplating the prospect of Greene and Cohen's prediction coming true, Morse writes, "Suppose neuroscience convinces us that agency and folk psychology are an illusion, that intentional bodily movements and reflexes are morally indistinguishable because both are simply the outcomes of mechanistic biophysical processes."³¹⁰ Morse's concern is that the contributions Greene and Cohen foresee neuroscience making to people's moral thinking threaten to dissolve extant conceptions of personhood. He writes, "Some people ... think that new discoveries about the causation of behaviour are leading inexorably to a purely mechanistic view of the link between the brain and behaviour, and thus to

such control they must have been able to act or refrain from acting; (4) the ability to act or refrain from acting requires that their mental states played a causal role in regulating their behaviour; (5) but mental states do not exist or they do no causal work; (6) thus people have no control over their actions; (7) thus retributive punishment is unjustified." (Pardo and Patterson, 2013, pg. 149-150)

³⁰⁸ Pardo and Patterson's reconstruction from the previous note is problematic. A more charitable treatment of Greene and Cohen would suggest that the reason for 7 (retributivism not being justified) is not 6 (that people completely lack control over their actions), but rather that the control people do have over their actions is a product of deterministic forces. In addition, it should be clear that the reason that G+C take people to have a degree of control over their behaviour that is *not* sufficient for retributive punishment is not 5 (that mental states lack a causal existence), but that the causal reach of mental states is significantly constrained by biology and environment.

³⁰⁹ See (Sifferd, 2006)

³¹⁰ (Morse, 2013, pg. 545-546)

a purely mechanistic view of human behaviour. Will the agentic person disappear and be replaced by the biological victim of neuronal circumstances?”³¹¹

The tectonic theory should alleviate Morse’s concern by substituting the concept of patiency for victimhood. The phrase ‘victims of neuronal circumstances’ is Greene and Cohen’s,³¹² and it is unfortunate. To insist that all persons can be seen as a victim of neuronal circumstances would erase important distinctions between persons. Above, Morse summarizes the worry as conflating reflexes and intentional bodily movements, which are distinguished by the way that folk psychological explanation makes sense of the latter, but is not recruited in response to the former. Checking to see if a person’s behaviour can be explained by folk psychological attributions is therefore a means of sorting agentic actions from non-agentic actions. In contrast to Greene and Cohen, the tectonic theory preserves this process in the first part of its two step structure. Investigating the extent to which a person’s behaviour can be explained folk psychologically can still serve to distinguish agents from non-agents. However, the tectonic theory is designed to go beyond this by considering the patiency of anyone deemed an agent, a corollary of which is that folk psychology is not sufficient for fair moral responsibility judgments. I should explain, therefore, how a purely folk psychological method of determining moral responsibility is problematic.

Folk theoretic knowledge is a function of evolutionary pressures applied to humans’ cognitive

³¹¹ (Morse, 2013, pg. 545-546)

³¹² “The law will continue to punish misdeeds, as it must for practical reasons, but the idea of distinguishing the truly, deeply guilty from those who are merely victims of neuronal circumstances will, we submit, seem pointless.” (Greene and Cohen, 2004, pg. 1781)

architecture and should not be expected to pass all tests of philosophical oversight. Even if we can naturalize the control component of moral responsibility in terms of agency rather than libertarian magic, so long as the presence of agency is discerned in folk psychological terms, it is plausible to expect epistemological trade-offs in the system. After all, folk psychology evolved because it facilitates the survival and proliferation of humans, not to provide philosophically satisfying theories about the nature of behaviour.

It's plausible that folk psychology evolved to provide a socially viable explanation of the behavioural heterogeneity of humans. The highly variegated behaviour of persons, if not implicitly taken to be underlaid by something unique to humans such as agency, would require that individuals develop hugely complex and idiosyncratic beliefs about why people behave the way they do.³¹³ Folk psychology, instead, provides a collective touchstone according to which the relative natural improvisatoriness of persons can be not only understood but expected. It is *prima facie* far easier for a society to develop and enforce behavioural norms if everyone agrees that the causes of persons' behaviour are beliefs and desires.

Conceiving of persons as purely folk psychological, however, contributes to the homogenization problem. Instead of taking a wide variety of facts about an individual into account when endeavouring to explain her behaviour, persons are typically satisfied that folk psychological explanations are sufficient. Indeed they are from a point of view that seeks just to certify the individual as a candidate for moral judgment. However, in the interest of fine-tuning such

³¹³ Wilfrid Sellars makes a similar point in a paper called 'Fatalism and Determinism.' He writes, "Whereas the inductively unexplainable behaviour of a merely material thing would be said to be unintelligible, a matter of chance, an inductively unexplainable action of a person – i.e. one which could not be explained in terms of character – can be intelligible in terms of being capable of explanation with reference to the practical reasoning and, ultimately, the volition of which it is the expression." (Sellars, 1966, pg. 148)

judgments so as to enhance their fairness, the limitations of folk psychological explanations need to be noted. Joshua Knobe has done compelling work on this issue.

Knobe launched his career with studies showing that folk psychological judgments are sensitive to moral considerations.³¹⁴ These were counter-intuitive findings because the usual view is that folk psychological appraisals supply information about what sort of an action a person has performed, which is only subsequently subject to moral judgment. Knobe showed that whether or not subjects thought an action was right or wrong influenced their pertinent folk psychological judgments. This phenomenon, sometimes called the Knobe-effect, is a challenge to the view that folk-psychology is an innocuous intermediary between persons' behaviour and moral judgments about it.

In a 2007 paper called, 'Folk Psychology: science and morals,'³¹⁵ Knobe investigates whether there are compelling reasons to think of folk psychology as akin to a scientific theory. He reasons that many of his colleagues' aversion to the claim that folk-psychological judgments are sensitive to moral considerations (as suggested by Knobe's earlier work) derives from associating folk psychology with scientific theories. Knobe argues that folk psychology is indeed a theory in that it is revisable according to empirical evidence (people's behavior), and deals with unobservable phenomena (beliefs and desires), but that it is not a scientific theory because it does not exclude considerations in the same ways that scientific pursuits do. Scientific claims are not subject to distortion by pre-existing moral views in the same way that folk psychological judgments are.

³¹⁴ (Knobe, 2010) provides the best overview of this work.

³¹⁵ (Knobe, 2007)

Knobe suggests that folk psychology's relative corruptibility by moral views is a result of the way that folk psychology evolved to serve two distinct purposes for humankind. He contends that folk psychology serves to facilitate prediction and explanation of persons' behavior, and also to facilitate the making of moral judgments about the behaviour. Knobe concludes that folk psychology evolved to adequately (but not expertly) facilitate these two things. He writes, "This single system of concepts [i.e. the beliefs and desires of folk psychology] might not do a perfect job either at generating predictions or at making moral judgments, but it could do at least an adequate job of both."³¹⁶

Knobe's work provides a way to stress that moral judgments based exclusively on folk psychology, although adequate, may still leave much to be desired. In terms of the tectonic theory, it could be said that folk psychological appraisals of a persons' behavior, while sufficient for determining whether or not she is an agent, are agnostic about (if not turned away from) patientic factors. Knobe compares thinking about persons to thinking about the weather: "In thinking about the weather, there is a need to make predictions about what conditions will arise in the future, and there is also a need to make evaluations of whether these conditions are good or bad for certain purposes."³¹⁷ For meteorologists, these tasks come apart easily. One system of concepts is used for predicting the weather, and a second collection of concepts awaits the results of the predictive schema before doing the work of evaluating them.

³¹⁶ (Knobe, 2007, pg. 168-169)

³¹⁷ (Knobe, 2007, pg. 169)

Knobe is pessimistic about there being an analogous arrangement for thinking about the actions of persons: “One approach would be to have a system of concepts that was specifically suited to the task of making predictions and then another, entirely separate system of concepts that was specifically suited to the task of evaluation. But such an approach might leave us with a large and unwieldy array of distinct ways of carving up the same class of phenomena.”³¹⁸ With some adjustments, however, I think the tectonic theory approximates such an arrangement without being unwieldy. The first step of the tectonic theory, determining whether a person qualifies as an agent, contains an element of prediction insofar as it implicitly suggests that a person so classified can be expected to be responsive to phenomena in ways that are broadly similar to the ways that other agents are. In other words, attributing agency to a person serves to indicate that hers is not an anomalous case of human action, but rather that what is known about agents generally will plausibly apply to her as well.

The second step of the tectonic theory, consideration of a person’s patiency, is a more purely evaluative and particularized enterprise. The tectonic theory prescribes considering patiency as a possible mitigator of a person’s moral responsibility, and although patientic considerations might provide some predictive content, their role in the tectonic theory is just to temper the punitiveness of moral responsibility judgments where appropriate. Thus, the tectonic theory goes some distance toward addressing the impoverishment of purely folk psychological moral responsibility judgments by performing the explanatory and evaluative components of such judgments in sequence. The tectonic theory deals first in descriptive explanatory facts concerning whether a person is sufficiently agentic to be held responsible for an action, but waits

³¹⁸ (Knobe, 2007, pg. 169)

until patientic considerations have been addressed to render an evaluation as to how much moral responsibility is appropriate. Such a schema is more complex than its alternatives, but is not unwieldy, and is philosophically well motivated.

Thus as Morse insists, Greene and Cohen's envisioned replacement of folk psychological methods of parsing the features of persons' actions with purely neuroscientific tools is rightly taken to make trouble for conceptions of personhood that involve agency. Greene and Cohen's proposal would undermine the way that classifying someone as an agent makes general knowledge about agentic persons applicable to her, which would plausibly make moral responsibility judgments unwieldy. However, as I've argued throughout this thesis, persons are patients as much as they are agents, and folk psychology deals primarily with persons' agency. Thus, even though Morse is right that folk psychology is essential for moral responsibility judgments, moral responsibility judgments that only involve folk psychological judgments will fail to address the homogenization problem by continuing to leave patientic considerations out of account.

To complete the defense of the tectonic theory's account of neuroscience's moral significance as pertaining to patientcy, I will show how operationalizing the tectonic theory in the justice system would work similarly to a proposal of Morse's. When he is not aiming to euthanize Greene and Cohen's argument, Morse appears to be amenable to the concerns I collected under the label of the homogenization problem. In a paper called 'Diminished Rationality, Diminished

Responsibility,³¹⁹ Morse advocates for “a generic doctrinal mitigating excuse of partial responsibility that would apply to all crimes.”³²⁰ The basic idea is to formalize a process whereby criminals can be given reduced sentences when factors beyond their control make significant contributions to the relevant action.

Morse comes close to describing the homogenization problem as it figures in the justice system. He writes, “Present law is unfair because it does not sufficiently permit mitigating claims. Criminal defendants, like people generally, display an enormously wide range of rational and control capacities.”³²¹ Morse’s concern is to sophisticate the justice system’s capacity to respond to details of particular cases, which he takes to be needed because existing tools for doing so are too blunt.³²² To combat this, he proposes that jurisdictions institute a new verdict that would be applicable to all crimes: guilty but partially responsible.³²³ Though he does not use the word ‘patency,’ Morse takes care to note that any diminution of responsibility must be traced to conditions that were beyond the actor’s control, and he specifically mentions that neuroscience could be useful for determining when this sort of mitigation will apply.

³¹⁹ (Morse, 2003)

³²⁰ (Morse, 2003, pg. 289)

³²¹ (ibid, pg. 296)

³²² “with precious few exceptions, present criminal law contains doctrinal all-or-nothing, bright line tests: the defendant was or was not legally insane; the defendant did or did not act under duress. Moreover, the criteria for such tests tend to be quite narrow, allowing few defendants to succeed with an excuse. Lesser rationality or control problems, or less grievous threats, may be considered only as a matter of discretion at sentencing.” (Ibid, pg. 295-296)

³²³ “I propose the adoption of a fourth verdict, “Guilty But Partially Responsible,” (GPR), that would be available in appropriate cases in addition to guilty, not guilty, and not guilty by reason of insanity.” (Ibid, pg. 299)

Morse writes, “The formula I would use would be this: “The jury may find the defendant GPR if, at the time of the crime, the defendant suffered from substantially diminished rationality *for which the defendant was not responsible* and which substantially affected the defendant’s criminal conduct.”³²⁴ This formulation clearly invokes patiency in specifying that the agent’s rationality diminishment must not be something for which she is responsible. It also insists that the diminishment ‘substantially affected’ the agent’s criminal conduct, which is akin to what I termed subductive patiency in the previous chapter. The main difference in Morse’s proposal is that it explains mitigation in reference to the agent and her rationality. Rather than acknowledging that factors beyond the agent’s control contribute to bringing about an action, Morse proposes integrating patientic considerations into culpability judgments by investigating whether they made an agent less capable of performing rational reflections about her own behavior. This is understandable given the normative weight that justice systems place on the concept of rationality, and though I think Morse’s schema would catch most of the considerations I call patientic, the tectonic theory provides a more satisfying explanation of the envisioned mitigation.

Morse’s conception of rationality is intentionally imprecise. In a recent paper, Morse makes this clear,

the law’s view is that people are capable of acting for reasons and are capable of minimal rationality according to the predominantly conventional, socially constructed standards. The type of rationality the law requires is the ordinary person’s common sense view of rationality, not the technical notion that might be acceptable within the disciplines of economics, philosophy, psychology, computer science, and the like.³²⁵

³²⁴ (Morse, 2003, pg. 300, italics added)

³²⁵ (Morse, 2011a, pg. 840)

Although Morse declines defining rationality, that which he takes a person's rationality to contribute to her behavior is fairly clear. My reading of Morse has him casting rationality as a means of avoiding criminal/immoral behavior. Describing how a mitigating rationality diminution affects a person, he writes, "Diminished rationality ... must also have substantially impaired the defendant's ability to access and to consider reflectively the good reasons not to commit the crime on this occasion."³²⁶

If 'rationality' is operationally defined as a capacity to avoid crimes through the application of reason, then when any factors that diminish this capacity result in criminal/immoral action, describing the occurrence of the action either as a result of the patientic factors at play or as a result of how they reduced the actor's capacity to avoid the action would seem to be equivalent. I suggest, however, that the concept of patientcy is actually doing most of the work. My support for this claim derives from the way that Morse insists that any mitigation due to rationality diminishment requires that the diminishment was non-culpable, i.e. not a result of agentic actions. Morse writes, "the proposal also requires that the rationality diminution must be non-culpable, which means that the defendant's impaired rationality must be justifiable or excusable. This reflects a socially based evaluative judgment that some rationality diminutions, and the defendant's response to them, are not the defendant's fault."³²⁷ This remark connects the envisioned mitigation to a social norm whose content is that a person is not responsible for that which they do not control, which claim figured in the basic argument for the tectonic theory from the previous chapter.

³²⁶ (Morse, 2003, pg. 301)

³²⁷ (Morse, 2003, pg. 300)

The importance of patiency is clearer still when Morse provides examples of culpable and non-culpable rationality diminishments. He writes, “Mental disorder and grief are examples of conditions that are non-culpable; voluntary ingestion of mind-altering substances, including ethanol (alcohol), is culpable and would not furnish the basis for mitigation.”³²⁸ This distinction is not well captured purely in terms of rationality diminishment. The difference between mental disorder and grief and drunkenness does not lie in the ways that they diminish rationality, but in the fact that becoming drunk is almost always something an agent can choose to do or not while grief and mental disorder arrive in a person’s life comparatively unbidden. Thus, when compared to Morse’s focus on agents’ rationality, the tectonic theory’s concept of patiency seems to do a better job of capturing the reasons for which conditions in persons’ lives sometimes mitigate her moral responsibility.

Returning to the focus of the chapter, the moral significance of neuroscience, Morse asserts that neuroscience could be relevant for the generic mitigating doctrine we’ve been considering. He writes,

Anglo-American criminal law does not have a generic mitigating doctrine that is applicable at trial. Unless a defendant meets the criteria for a complete excuse such as legal insanity or duress, the law expects the defendant to control himself despite temptations and provocations. Mitigation is then left to the discretion of the sentencing judge. I have argued, in contrast, that criminal law should provide defendants with a partial mitigating doctrine applicable at trial that would in appropriate cases reflect a defendant's substantially diminished rational or control capacities. Neuroscience studies

³²⁸ (Morse, 2003, pg. 300-301)

might help confirm that some defendants have much greater difficulty obeying the law ... than criminal law and common sense presuppose.³²⁹

As argued above, revelations to the effect that an agent faced difficulties obeying the law are better rendered in terms of patency than diminishment of the agent's capacities. Neuroscientific facts are no different, except insofar as they are perhaps particularly suited to being construed in terms of patency because they are facts about persons' bodies, as explained in the previous section.

Morse's worry about the use of neuroscientific facts in the courtroom is that they could induce judges and jurors to conflate causation with excuse, a circuit of reasoning that involves appeal to determinism.³³⁰ This is what motivates Morse's mainly deflationary work on neuroscience and the law. He rightly sees proposals like Greene and Cohen's as a threat to agentic conceptions of personhood. This leads him to make what I take to be a problematic distinction between internal and external critiques of moral/legal responsibility. Considering the possible uses of neuroscience for the law, he writes,

We must begin with a distinction between internal relevance and external relevance. An internal contribution or critique accepts the general coherence and legitimacy of a set of legal doctrines, practices or institutions and attempts to explain or alter them. For example, an internal contribution of criminal responsibility may suggest the need for doctrinal reform, of, say, the insanity defense, but it would not suggest that the notion of criminal responsibility is itself incoherent or illegitimate. By contrast, an externally relevant critique suggests the doctrines, practices or institutions are incoherent, illegitimate, or unjustified.³³¹

³²⁹ (Morse, 2011b, pg 542)

³³⁰ Morse's reticence toward such inferences is clear when he writes of his generic mitigating doctrine that, "the proposal does not take an *objectionably mechanistic* approach to the effect of untoward mental states on criminal behavior." (Morse, 2003, pg. 300)

³³¹ (Morse, 2011, pg. 536)

This distinction underlies Morse's insistence that neuroscientific facts be translated into folk psychological terms before they acquire moral/legal significance. The idea that appeals to neuroscience must either culminate in 'external' criticisms that aim to undermine moral/legal responsibility altogether, or accept moral/legal responsibility as it is in a broad stroke before endorsing minor 'internal' alterations results in Morse's insistence that neuroscience either work with or undermine the folk psychological tools at the heart of the system. However, the tectonic theory provides another option.

The tectonic theory is motivated first and foremost by Reader's insights about patiency being constitutive of personhood. My claim that existing accounts of moral responsibility neglect patientic considerations, resulting in the homogenization problem, could be interpreted as either an internal or external critique of moral responsibility, according to Morse's criteria. It is 'external' insofar as it advocates for fundamentally changing the way moral/legal responsibility judgments are carried out, but it is 'internal' insofar as it retains existing conceptions of morally responsible agency. The tectonic theory preserves folk psychological methods of making moral responsibility judgments, but sophisticates the context in which they contribute by supplementing them with patientic considerations.

The tectonic theory thus provides a way to make moral/legal culpability judgments alert to the significant differences between agents, which Morse agrees is important, but without having to formulate the differences in agentic or folk psychological terms. This preserves the heuristic and social value of agentic considerations but can significantly enhance the fairness of moral

responsibility judgments. This should alleviate Morse's worry about neuroscience making the agentic person disappear because the tectonic theory is built from the ground up to avoid the determinism-invoking incompatibilism of the sort Greene and Cohen predicted neuroscience facilitating. If we were faced with a choice between incorporating neuroscientific facts into courtrooms and keeping folk psychology intact, the former would be preferable, but the tectonic theory provides another way forward. Neuroscience may yet make justice systems more humane if the tectonic theory were adopted. It would do so because it provides patientic facts about persons and these are important, not because it undermines any of the existing folk psychological tools we use to understand one another.

Conclusion

In this chapter, I presented Greene and Cohen's (in)famous argument from 'For the law neuroscience changes nothing and everything.' Greene and Cohen argue that neuroscience will have a transformative effect on intuitive moral thinking, which will eventually result in a non-retributive justice system. Greene and Cohen's critics point out a number of reasons to be suspicious of such a prospect, but I showed that Greene and Cohen ultimately undermine their own case. This is important because it is illustrative of the way that the tendency of people to consider one another as morally responsible agents is robust enough not to be threatened by neuroscientific facts about the mechanical nature of human cognition.

However, I also explained that the basic idea behind Greene and Cohen's proposal was sound. Intuitive morality does need to change if justice systems are to benefit significantly from

neuroscientific developments. The change required, though, is the concern for patiency prescribed by the tectonic theory. Appealing to work from Gray, Wegner, and others, I showed that neuroscientific facts are well suited to being interpreted through the lens of patiency because they are about persons' bodies, which are more associated with patiency than agency.

Finally, I defended the tectonic account of neuroscience's moral significance against the concerns voiced by Greene and Cohen's critics. I explained how the tectonic theory provides the means to illuminate and hopefully refine people's intuitions about which factors mitigate a persons' responsibility and which do not. I also showed that the tectonic theory can take mitigating patientic factors into account without threatening the agentic element in conceptions of personhood, by preserving but also supplementing folk psychological considerations. The tectonic theory of moral responsibility enables culpability judgments to take neuroscientific facts into account in a principled way that can plausibly go some distance towards Greene and Cohen's vision of more humane/less punitive justice systems while guarding against the losses that worried their critics.

Conclusion

Dennett describes the state of philosophical debate about free will and moral responsibility as “a rather unfortunate stew of misapprehensions.”³³² After getting to know the literature, I came to think this is apt, but also that the situation is rapidly improving. I interpret the compatibilist mainstream of the moral responsibility literature to be engaged in clearing away some of the particularly problematic bits of the stew that trace back to the traditional accounts of free will that Waller described as *miracle-requisite*. Most of this work focuses on clarifying the notion of agency on which moral responsibility lies so that no more is demanded of morally responsible agents than can be reasonably expected within a naturalistic framework. I think this can be likened to brushing one’s teeth after the stew, a crucial first step toward cleaning up the enterprise of moral responsibility. But as burdensome and unrequired as it may seem, just as with oral hygiene, one ought to do more than brush one’s teeth after eating. I conceive of the tectonic theory’s concern for patiency as akin to flossing, a crucial second step that can clear away smaller and better hidden remnants of the stew that remain even after the notion of agency has been cleaned up.

My goal with this thesis was to develop a conceptual apparatus that would make moral responsibility judgments more sensitive to the details of particular cases. Usually, the philosophers who call for such a thing are skeptics about the legitimacy of moral responsibility practices, such as Greene and Cohen, Derk Pereboom, or Bruce Waller. These philosophers argue that once sufficient attention is paid to the nuanced antecedents of particular actions, it

³³² (Dennett, 2011)

ought to be impossible to maintain that people have sufficient control over themselves to be rightly held responsible for their behaviour. I agree that looking more closely at the conditions from which an action originates can attenuate the severity of moral responsibility judgments, but I do not think there is reason to adopt a skeptical position regarding moral responsibility. In the first place, I don't think that the elimination of moral responsibility practices is feasible. Such practices play too large a role in human societies to be excised. In the second place, and I take this to be the more important point, I do not think that one needs to adopt a skeptical position towards moral responsibility judgments to enhance their sensitivity to details of particular cases.

Instead, the tectonic theory of moral responsibility that I develop reconceptualizes what moral responsibility judgments ought to be judgments of. Extant theories of moral responsibility tend to conceive of moral responsibility judgments as in the business of deciding whether a person qualifies as an agent or not. This is an understandable approach given that proponents of moral responsibility are generally working out responses to moral responsibility skeptics, whose arguments tend to target the very idea of categorizing anyone as an agent. I contend, however, that there is both room, and reason to sophisticate what it is about a person that ought to factor into moral responsibility judgments. The tectonic theory suggests that moral responsibility judgment should take into account not just persons' agency, but also their patiency.

A person's patiency refers to those aspects of her self and life that are not under her control. These sorts of factors are usually just presented as symptoms of the deterministic nature of persons by skeptics intending to undermine the legitimacy of moral responsibility. The tectonic

theory, on the other hand, sensitizes moral responsibility judgments to uncontrolled factors that are relevant to persons' actions, but without the usual trappings of incompatibilist skepticism about what persons do control. By exchanging incompatibilist appeals to determinism for a concern for persons' patiency, the tectonic theory allows for considering the uncontrolled aspects of an action within moral responsibility judgments rather than presenting such factors as reasons not to make such judgments.

To conclude the thesis, I would like to address how the tectonic theory of moral responsibility touches on some of the same issues that are present in the philosophical literatures on moral luck, and autonomy. Some philosophers working on moral luck have wrestled with phenomena that I have called patientic. The literature that has bloomed around the initial work of Bernard Williams and Thomas Nagel's papers (both called 'Moral Luck')³³³ has occasionally addressed the question of how factors outside of a person's control can mitigate her moral responsibility. Different types of moral luck have been delineated, including resultant, constitutive, causal, and circumstantial luck. To explain how the tectonic theory relates to discussions of moral luck, the important distinction lies between resultant luck, and the other varieties. Resultant luck refers to ways in which the consequences of a person's actions often lie outside of her control. For example, two drunk drivers, one of whom makes it home safe while the other happens to come across a child all of a sudden darting in front of the car and incurring injury, were subject to different resultant luck.

³³³ (Nagel, 1979), (Williams, 1981)

Resultant luck, thus, considers the interval of time subsequent to an action. The tectonic theory, in contrast, has been presented as purely backward looking, recommending that both agentic and patientic factors relevant to an action's occurrence be taken into account. It would be interesting to survey ways of extending the tectonic theory to cover the consequences of actions, perhaps developing ways of distinguishing between agentic and patientic consequences, but I have not attempted to do so in this thesis. I think it is far more pressing to install a tectonic form of assessment of the interval of time leading up to an action's occurrence. The antecedents of actions provide plenty of considerations that can be profitably parsed by the tectonic model, and I hope to have shown that such a parsing can enhance the fairness of moral responsibility judgments.

The other sorts of luck (causal, circumstantial, and constitutive) do refer to uncontrolled antecedents of actions, and authors have discussed whether and how they should alter moral responsibility appraisals. However, there is a significant difference between discussions of moral luck and the patientic considerations encouraged by the tectonic theory. The literature on moral luck is nearly entirely driven by the perceived importance of limiting the content of moral responsibility appraisals to agentic considerations. This usually takes the form of checking to see whether a control principle should govern moral responsibility assessments. Such a principle states that a person should only be morally assessed in regards to matters under her control.

To my judgment, this principle exemplifies what Reader called the agentic bias, a tendency to emphasize what a person does control at the expense of also considering what she does not

control. Even use of the word ‘luck’ to characterize patientic factors is problematic insofar as it assigns such factors a secondary status and places them outside of a person. As Reader emphasized, it is important to refrain from thinking that a person is less of a person to the extent that she is subject to factors beyond her control. All agents are also patients at all times regardless of any forms of luck to which they are subject. The tectonic theory I have developed in this thesis can be characterized as an alternative to discussions of moral luck. Instead of relegating factors beyond a person’s control to the realm of luck, the tectonic theory installs patiency as an inevitable and significant feature of persons and their actions, thus deserving of attention within moral responsibility judgments.

As with the literature on moral luck, philosophical discussions about autonomy also share some of the concerns to which the tectonic theory is a response. ‘Autonomy’ is a feature of entities of whom it is legitimate to say that they are self-governing or self-ruled. It is thus similar to agency. However, the conception of agency that I have been working with is tethered directly to moral responsibility judgments. Autonomy is a more wide-ranging notion. The questions of whether a person should be judged as self-ruled, and whether a person has enough self-control to be a candidate for moral responsibility are distinct. The former has aspirational value that the latter lacks. Autonomy can be likened to an ideal, a state in which a person’s actions and life are actively endorsed by their beholder. Agency, on the other hand, is intended in this thesis just as a feature that separates persons who should be subject to moral responsibility assessments, from those that should not (such as the very young, or the mentally disabled).

The autonomy literature also harbors some ideas that resemble patiency. Heteronomy, for example, is a concept tracing back to Rousseau and Kant, which has some kinship with patiency. Heteronomy is a term for actions that are influenced by forces outside the actor. Autonomy and heteronomy might be taken to be related in a way that compares to the relation between agency and patiency. However, as with autonomy and agency, the difference between agency/patiency and autonomy/heteronomy can be explained on the basis of the latter relation having a wider scope. To describe a person as either autonomous or heteronomous is to render a verdict about that person as opposed to just an assessment of her actions.

The concepts of autonomy/heteronomy can conceivably be deployed in such a way that they incentivize the sorts of wide-ranging considerations of actions recommended by the tectonic theory, but their native habitat, so to speak, is in the wider world of ethical theory. For Kant, autonomy was a feature of persons important for justifying deontological ethical theory, and heteronomy was used to landmark situations in which persons could be prevented from developing the deliberative prowess required for the kinds ethical character Kant thought to be valuable and/or obligatory. In contrast to these wide ranging concerns, the tectonic theory's dyad of agency and patiency is intended to range just over the antecedents of actions for which a persons' moral responsibility is at issue. While a more full-fledged inventory of how agency and patiency might serve to shape general ethical theory would be an interesting project, the purpose of this thesis was to show how inclusion of patientic considerations in moral responsibility judgments (a relatively minor adjustment), could have significant and wide ranging consequences whether or not this, that, or even any general ethical theory is subscribed to.

References

- Arico, A. J. (2012). Breaking Out of Moral Typecasting. *Review of Philosophy and Psychology*, 3(3), 425–438. doi:10.1007/s13164-012-0108-1
- Bean, H. (2001). *The Believer*.
- Baron-Cohen, S. (1997): ‘Are Children with Autism Superior at Folk Physics?’, in H. Wellman and K. Inagaki (eds.), *The Emergence of Core Domains of Thought: Children’s Reasoning about Physical, Psychological, and Biological Phenomena* (pp. 45--54), San Francisco: Jossey-Bass.
- Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). The Neurocognitive Profile of Williams Syndrome: A Complex Pattern of Strengths and Weaknesses. *Journal of Cognitive Neuroscience*, 12(Supplement), 7–29.
- Blair, R. J. R. (1995). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition*, 57, 1–29.
- Blair, R. J. R. (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26(5), 571–579. doi:10.1007/BF02172277
- Block, N. (1996). On A Confusion About a Function of Consciousness. *Behavioral and Brain Sciences*, 6, 227–287. doi:10.1017/S0140525X00038188
- Bloom, Paul. (2004). *Descartes' Baby: How the Science of Child Development Explains What Makes us Human.*, 271 pp, Basic Books
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65, 17–28. doi:10.1016/j.tpb.2003.07.001
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3531–3535. doi:10.1073/pnas.0630443100
- Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought* (p. 376). Basic Books.
- Brink, D. O. (2013). Situationism, Responsibility, and Fair Opportunity. *Social Philosophy & Policy*, 30, 121–149.
- Brosnan, S. F., & De Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299. doi:10.1038/nature01963

- Churchland, P. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, 78(2), 67–90.
- Churchland, P. (1987). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT Press. doi:10.1016/0306-4522(87)92990-3
- Ciurria, M. (2013). Moral Responsibility: Justifying Strawson and the Excuse of Peculiarly Unfortunate Formative Circumstances. *Ethical Theory and Moral Practice*, 17(3), 545–557. doi:10.1007/s10677-013-9457-9
- Copp, D. (1997). Defending the Principle of Alternate Possibilities : Blameworthiness and Moral Responsibility. *Nous*, 31(4), 441–456.
- Darley, John, and Daniel Batson. 1973. ‘From Jerusalem to Jericho’: A Study of Situational and Dispositional Variables in Helping Behavior. *Journal of Personality and Social Psychology* 27
- Davies, P. S. (2013). Skepticism Concerning Human Agency: Sciences of the Self Versus “Voluntariness” in the Law. In *Neuroscience and Legal Responsibility*, ed Nicole A. Vincent (pp. 113–130). doi:10.1093/acprof
- Dennett, D. (1984a). *Elbow Room: The Varieties of Free Will Worth Wanting* (p. 210). Oxford University Press.
- Dennett, D. (1984b). I Could not have Done Otherwise--So What? *The Journal of Philosophy*, 8(10), 553–565.
- Dennett, D. (2003). *Freedom Evolves*. Viking.
- Dennett, D. (2011). “My Brain Made Me Do It” (when neuroscientists think they can do philosophy). In *Max Weber Lecture Series* (p. 21).
- Dennett, D. (2012). Review of Bruce N. Waller, *Against Moral Responsibility*. Retrieved January 13, 2015, from <http://www.naturalism.org/DCDWallerreview.htm>
- Dennett, D. C. (1989). *The Intentional Stance* (Seventh Pr.). The MIT Press. doi:10.1017/S0140525X00058611
- DeScioli, P., Gilbert, S. S., & Kurzban, R. (2012). Indelible Victims and Persistent Punishers in Moral Cognition. *Psychological Inquiry*, 23(2), 143–149. doi:10.1080/1047840X.2012.666199
- Doris, J. M. (1998). *Lack of Character: Personality and Moral Behaviour*. Cambridge University Press.
- Doris, J. M., & Murphy, D. (2007). From My Lai to Abu Ghraib : The Moral Psychology of Atrocity. *Midwest Studies in Philosophy*, 31, 25–55.

- Dretske, F. (1991). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.
- Eagleman, D. M. (2012). *Incognito: The secret lives of the brain*. Vintage.
doi:10.1080/13546805.2012.715448
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.
doi:10.1038/415137a
- Fischer, J. M. (2012). *Deep Control: Essays on Free Will and Vlaue*. Oxford University Press.
- Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829–839.
- Frankfurt, H. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5–20.
- Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state–trait distinction. *Journal of Personality and Social Psychology*, *50*(1), 169–174. doi:10.1037/0022-3514.50.1.169
- Frith, U. (2003). *Autism: Explaining the Enigma*, 2nd edition. Oxford: Blackwell.
- Germiniani, F. M. B., Miranda, A. P. P., Ferenczy, P., Munhoz, R. P., & Teive, H. a. G. (2012). Tourette's syndrome: from demonic possession and psychoanalysis to the discovery of gene. *Arquivos de Neuro-Psiquiatria*, *70*(February), 547–549. doi:10.1590/S0004-282X2012000700014
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science (New York, N.Y.)*, *315*(5812), 619. doi:10.1126/science.1134475
- Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–20. doi:10.1037/a0025883
- Gray, K., & Wegner, D. M. (2009). Moral typecasting: divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–20. doi:10.1037/a0013748
- Gray, K., & Wegner, D. M. (2011). To escape blame, don't be a hero—Be a victim. *Journal of Experimental Social Psychology*, *47*(2), 516–519. doi:10.1016/j.jesp.2010.12.012
- Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry*, *23*(2), 101–124. doi:10.1080/1047840X.2012.651387

- Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1451), 1775–85. doi:10.1098/rstb.2004.1546
- Gullucayir, S., Asirdizer, M., Yavuz, M. S., & Zeyfeoglu, Y. (2009). Criminal and Legal Responsibilities in Tourette ' s Syndrome. *Psychiatry: Interpersonal and Biological Processes*, 46(3), 221–225.
- Haji, I. (2009). *Incompatibilism's Allure: principal arguments for incompatibilism*. Broadview Press
- Harman, G. (2003). No character or personality. *Business Ethics Quarterly*, 13, 87.
- Honderich, T. (2002). Dismay and Intransigence; Compatibilism and Incompatibilism. In *How Free Are You? The Determinism Problem*. Oxford University Press.
- Isen, Alice, and Paula Levin. 1972. Effect of Feeling Good on Helping. *Journal of Personality and Social Psychology* 21 384-88
- Jack, A. I. (2014). A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis. *Oxford Studies in Experimental Philosophy (Vol. 1)*, 1–32.
- Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccio, A. H., & Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *NeuroImage*, 66, 385–401. doi:10.1016/j.neuroimage.2012.10.061
- Jack, A. I., & Robbins, P. (2012). The Phenomenal Stance Revisited. *Review of Philosophy and Psychology*, 3(3), 383–403. doi:10.1007/s13164-012-0104-5
- Jack, A. I., Robbins, P., Friedman, J. P., & Myers, C. D. (2014). More Than a Feeling: Counterintuitive Effects of Compassion on Moral Judgment. In J. Sytsma (Ed.), *Advances in Experimental Philosophy of Mind* (pp. 125–179). Bloomsbury Academic.
- Jankovic, J., Kwak, C., & Frankoff, R. (2006). Tourette's syndrome and the law. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 18, 86–95. doi:10.1176/appi.neuropsych.18.1.86
- Kane, R. (1998). *The Significance of Free Will*. Oxford University Press.
- Karlsson, M. M. (2002). Agency and Patiency: Back to Nature ? *Philosophical Explorations*, V(May 1999), 59–81.
- Knobe, J. (2007). Folk Psychology: Science and Morals. In *Folk Psychology Re-Assessed* (pp. 157–173). Springer.

- Knobe, J. (2010). Person as scientist, person as moralist. *The Behavioral and Brain Sciences*, 33, 315–329; discussion 329–365. doi:10.1017/S0140525X10000907
- Knobe, J. (2014). Free Will and the Scientific Vision. In E. Machery & E. O'Neill (Eds.), *Current Controversies in Experimental Philosophy* (pp. 1–15). Routledge.
- Kramer, H., & Sprenger, J. (2007). *The Malleus Maleficarum*, 292.
- Levine, J. (2001). *Purple Haze: the Puzzle of Consciousness* (p. 216). Oxford University Press.
- Mandelbaum, E., & Ripley, D. (2012). Explaining the Abstract/Concrete Paradoxes in Moral Psychology: The NBAR Hypothesis. *Review of Philosophy and Psychology*, 3(3), 351–368. doi:10.1007/s13164-012-0106-3
- McNaught, & P., K. St. (2010). 125 Years of Tourette Syndrome: The Discovery, Early History and Future of the Disorder. *Inside TSA*, 38(3), 1–7.
- Mele, A. R. (1995). *Autonomous Agents*. Oxford University Press. doi:10.5772/236
- Milgram, Stanley. 1969. *Obedience to Authority*. New York: Harper and Row
- Monterosso, J., Royzman, E. B., & Schwartz, B. (2005). Explaining Away Responsibility: Effects of Scientific Explanation on Perceived Culpability. *Ethics & Behavior*, 15(2), 139–158. doi:10.1207/s15327019eb1502_4
- Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology. General*, 138(4), 535–545. doi:10.1037/a0016796
- Morse, S. J. (2003). Diminished rationality, diminished responsibility. *Ohio St. J. Crim. L.*, 1(1998), 289.
- Morse, S. J. (2004). New neuroscience, Old problems. In B. Garland (Ed.), *Neuroscience and the Law: Brain, Mind, and the Scales of Justice* (pp. 157–198). New York: Dana Press.
- Morse, S. J. (2011a). Avoiding Irrational NeuroLaw Exuberance: A Plea for Neuromodesty. *Law, Innovation and Technology*, 3, 209–228. doi:10.5235/175799611798204932
- Morse, S. J. (2011b). Lost in Translation? An Essay on Law and Neuroscience. *Law and Neuroscience : Current Legal Issues Volume 13*, 13(January), 1–42. doi:10.1093/acprof
- Nagel, Thomas. (1979). *Mortal Questions*. Cambridge University Press.
- Nichols, S., & Knobe, J. (2007). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. *Nous*, 41(4), 663–685. doi:10.1111/j.1468-0068.2007.00666.x

- Pardo, M. S., & Patterson, D. (2013). Neuroscience, Normativity, and Retributivism. In *The Future of Punishment*, ed. Thomas Nadelhoffer. Oxford University Press.
- Pereboom, D. (2013). Free Will Skepticism and Criminal Punishment. In T. Nadelhoffer (Ed.), *The Future of Punishment* (pp. 49–78). Oxford University Press.
- Pettit, P. 2002 The capacity to have done otherwise. Rules, reasons, and norms: selected essays. Oxford University Press.
- Reader, S. (2007a). *Needs and Moral Necessity*. Routledge.
- Reader, S. (2007b). The Other Side of Agency. *Philosophy*, 82(322), 579–604.
- Reader, S. (2010). Agency, Patiency, and Personhood. In T. O'Connor & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 200–208). Blackwell Publishing Ltd.
- Robbins, P., & Jack, A. I. (2006). The Phenomenal Stance. *Philosophical Studies*, 127(1), 59–85. doi:10.1007/s11098-005-1730-x
- Roskies, A., & Nichols, S. (2008). Bringing moral responsibility down to earth. *The Journal of Philosophy*, (1999), 1–16.
- Sanfey, A. G., Rilling, J. K., Aronson, J. a, Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(June), 1755–1758. doi:10.1126/science.1082976
- Schlosser, M. E. (2013). Conscious will , reason-responsiveness , and moral responsibility. *Journal of Ethics*, 17(section 2), 205–232.
- Sellars, W. (1966). Fatalism and Determinism. In K. Lehrer (Ed.), *Freedom and Determinism* (pp. 141–174). New York: Random House.
- Sifferd, K. L. (2006). In Defense of the Use of Commonsense Psychology in the Criminal Law. *Law and Philosophy*, 25(6), 571–612. doi:10.1007/s10982-005-3802-7
- Sinnott-Armstrong, W. (2008). Abstract + Concrete = Paradox. In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy*. Oxford University Press.
- Smilansky, S. (2011). Hard Determinism and Punishment: A Practical Reductio. *Law and Philosophy*, 30, 353–367. doi:10.1007/s10982-011-9099-9
- Smilansky, S. (2012). Free Will and Moral Responsibility: The Trap, the Appreciation of Agency, and the Bubble. *The Journal of Ethics*, 16(2), 211–239. doi:10.1007/s10892-012-9126-6

- Sneddon, A. (2005). Moral Responsibility: The Difference of Strawson, and the Difference it Should Make. *Ethical Theory and Moral Practice*, 8, 239–264. doi:10.1007/s10677-005-2484-4
- Sneddon, A. (2006). *Actions and Responsibility*. Springer.
- Strawson, G. (1986). *Freedom and Belief*. Oxford University Press. doi:10.1093/acprof:oso/9780199247493.001.0001
- Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25.
- Strickland, B., Fisher, M., & Knobe, J. (2012). Moral Structure Falls Out of General Event Structure. *Psychological Inquiry*, 23(2), 198–205. doi:10.1080/1047840X.2012.668272
- Sytsma, J., & Machery, E. (2012). The Two Sources of Moral Standing. *Review of Philosophy and Psychology*, 3(3), 303–324. doi:10.1007/s13164-012-0102-7
- Taylor, J. S. (2009). *Practical Autonomy and Bioethics*. Routledge.
- Taylor, R. (1982). Agent & Patient: Is There a Difference? *Erkenntnis*, 18, 223–232.
- Theriault, J., & Young, L. (2014). Taking an “Intentional Stance” on Moral Psychology. In J. Sytsma (Ed.), *Advances in Experimental Philosophy of Mind* (pp. 101–124). Bloomsbury Academic.
- Tourette Syndrome and Criminal Behaviour. (2014). doi:10.1037/e502522010-006
- Van Inwagen, P. (1983). *An Essay on Free Will* (p. 248). Oxford: Clarendon Press.
- Vargas, M. (2010). The Revisionist Turn: a brief history of recent work on free will. In J. Aguilar, A. Buckareff, & F. Keith (Eds.), *New Waves in Philosophy of Action*.
- Vargas, M. (2013a). *Building Better Beings*. Oxford University Press. doi:10.1093/acprof
- Vargas, M. (2013b). Situationism and Moral Responsibility: Free Will in Fragments. In *Decomposing the Will*. New York: Oxford University Press.
- Vincent, N. A. (2013). Enhancing Responsibility. In *Neuroscience and Legal Responsibility*, ed Nicole A. Vincent (pp. 305–331). doi:10.1093/acprof
- Waller, B. N. (2011). *Against moral responsibility* (p. 352). The MIT Press.
- Waller, B. N. (2012). Reply from Waller. Retrieved January 23, 2015, from <http://www.naturalism.org/Wallerexchange.htm>

Watson, G. (1987). 'Responsibility and the Limits of Evil: Variations on a Strawsonian Theme. In F. Schoeman (Ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 256–286). Cambridge: Cambridge University Press.

Wegner, D. (2002). *The Illusion of Conscious Will* (p. 416). MIT Press.

Williams, Bernard. (1981). *Moral Luck: Philosophical Papers, 1973-1980*. Cambridge University Press

Williams, B. (2006). *Ethics and the Limits of Philosophy*. Fontana Press.
doi:10.1017/S003181910400004X

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: studies in the attribution of moral responsibility. *Cognition*, *100*(2), 283–301. doi:10.1016/j.cognition.2005.05.002