CHOICE, OWNERSHIP AND RESPONSIBILITY

A Dissertation
Presented to
The Faculty of the Graduate School
At the University of Missouri

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

By
XIAOFEI LIU
Dr. Peter Vallentyne, Dissertation Supervisor

MAY 2012
The undersigned, appointed by the dean of the Graduate School,
have examined the Dissertation entitled

CHOICE, OWNERSHIP AND RESPONSIBILITY

Presented by XIAOFEI LIU

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

__________________________________________

Peter Vallentyne

__________________________________________

Peter Markie

__________________________________________

Philip Robbins

__________________________________________

Sara Rachel Chant

__________________________________________

Paul Litton
ACKNOWLEDGEMENTS

I am very grateful to Peter Markie, Paul Litton, Robert Johnson, Sara Rachel Chant, Paul Weirich, Philip Robbins, Leo Yan, Daniel Marshall, Wenwen Fan and Ashton Sperry for their helpful comments and suggestions. I am also thankful to Scott Sehon and other participants of a 2011 Pacific APA colloquium for their comments on an earlier version of a chapter of this dissertation. Finally, I want to thank my advisor, Dr. Peter Vallentyne, whose guidance, patience, and encouragement this dissertation owes its greatest debt to.
# Table of Contents

Acknowledgements......................................................................................................................... ii

Abstract................................................................................................................................................ iv

Chapter One........................................................................................................................................... 1

Chapter Two.......................................................................................................................................... 19

Chapter Three.......................................................................................................................................... 38

Chapter Four.......................................................................................................................................... 56

Chapter Five........................................................................................................................................... 81

Bibliography.......................................................................................................................................... 101

Vita.......................................................................................................................................................... 108
ABSTRACT

My dissertation is to answer these two questions: "Does moral responsibility require choice?" and "If not, what does it require?" Classic accounts of moral responsibility, such as libertarian accounts, assume a volition condition – we are not morally responsible for a behavior unless we have directly or indirectly chosen it. I call this view volitionism. Non-volitionism, on the other hand, claims that no such choice is necessary for moral responsibility. I propose and defend a non-volitionist account of moral responsibility. I first argue against volitionism by comparing a paradigm case of choice with some typical cases of non-choice mental activities and showing that no responsibility-generating power can be found uniquely in choice. Then, I examine some current non-volitionist accounts and argue that they all face a serious challenge – the Problem of Brain Manipulation. Finally, based on what I call the normative Strawsonian framework, I propose a non-volitionist account of moral responsibility that meets this challenge. According to my account, moral responsibility for a behavior requires that, roughly, the behavior be caused by a certain evaluative attitude or judgment of which the agent can claim deep ownership.
CHAPTER ONE

Introduction

“That is our choices that show what we truly are, far more than our abilities.”

– J. K. Rowling

“If you can control the process of choosing, you can take control of all aspects of your life. You can find the freedom that comes from being in charge of yourself.”

– Robert F. Bennett

On a shabby basketball court in a back street near my neighborhood, I saw a group of young men practicing some flashy basketball moves one day. One of them was making a quick drive to the basket, but his newly learned dribbling skills failed to catch up with his speed and, seeing this opportunity, his opponent tried to take advantage of his fumble. In a heated struggle, the ball accidentally got out of the hands and flew directly into the face of one of them, leaving him a bloody nose. The person who was hurt got angry. However, instead of pouring his anger towards the immediate cause of his injury—the basketball, or the one before it—the bodily conflict, he directed his anger towards the other player. Then, interestingly, after the other player came over to him and made an apology—"Sorry buddy, my bad, I didn't mean it", the anger immediately subsided and everyone moved on, putting this incident behind.

This is familiar scene for us. But, imagine a group of aliens who just landed on our planet and observed this incident. There may be something very puzzling to them:
Why would the injured player have such an emotional reaction towards other objects? Why would that reaction be directed towards the other player rather than the immediate cause of his injury, i.e. the basketball? Why would some kind of gesture or words from the other player make that reaction disappear?

One of the most important features of our human society is that we are held morally responsible for our behaviors. We blame a politician for not keeping his words. We praise a doctor for saving a life. We feel indignation when a friend betrays us. We show gratitude when a stranger goes out of his way to help us. When we show these attitudes toward each other, we are viewing them as being morally responsible for what they do. We are blameworthy or praiseworthy for our behaviors.¹

What makes a person morally responsible for his or her behavior? Consider the following scenario. A friend of mine promised to pick me up at the airport but failed to show up when I arrived. I might initially feel upset and blame him for failing to show up. But later on, I realized that I had mistakenly given him the wrong date. Upon realizing this, I no longer viewed him as blameworthy for failing to receive me at the airport. Presumably, this is because he did not know that he was supposed to pick me up on that day.

Ever since Aristotle, philosophers believe that there are two necessary conditions for moral responsibility: one is the "epistemic condition"; the other is the "control condition".² To meet the epistemic condition, it is usually believed that the agent needs to

---

¹ We can be morally responsible for outcomes, behaviors and characters (attitudes, judgments, etc.). For simplicity, my focus in this project is moral responsibility for behaviors.

² *Nicomachean Ethics*, 1999, translated by Terence Irwin (2nd edition), 1109b30-1110a2. However, it has to be pointed out that there are some people questioning whether Aristotle had the modern notion of moral responsibility. For arguments against this claim that Aristotle had a
have proper knowledge of what he is doing and the moral consequence of doing it. To meet the control condition, it is usually believed that the agent needs to have proper control over what he is doing such that the behavior reflects his own agency. In the aforementioned case, since my friend did not know that he was supposed to receive me at the airport on that particular day, he did not meet the epistemic condition for moral responsibility. This is why I should no longer view him as blameworthy upon realizing my mistake. Alternatively, if the reason why my friend failed to pick me up at the airport is that he was delayed on his way to the airport by an unexpected traffic jam, then I should also not view him as blameworthy. In this case, he had no control over the traffic jam, and thus he did not meet the control condition for moral responsibility.

For those who accept these two conditions as necessary for moral responsibility, it seems natural to believe in this *volition condition* – we are not blameworthy or praiseworthy for a behavior unless we have made a relevant choice. When we are making an informed choice, we are more likely to have the requisite knowledge of what we are going to do and its moral consequence. And, more importantly, it seems that we only have the requisite control through our choices.

Choice plays an important role in our life. The importance of choice is readily reflected in our constant demand for the opportunity to make a choice. We demand an opportunity to say "I do" (or, occasionally, "I don't") before entering a relationship with another person; we demand an opportunity to give our consent before receiving a medical treatment; we demand an opportunity to cast a vote before accepting the governance of some authority.... We care deeply about the opportunity to choose even if our choice

---
modern notion of moral responsibility see Roberts (1989), Curren (1989); and for arguments for this claim, see Irwin (1980) and Meyer (1993).
sometimes does not change how things will take place and it does not necessarily benefit our wellbeing. We care about the opportunity to choose because we believe that only through choice do we take control over our own life. Through our choices, we control our career, our relationship, our health or even our political system. So, it is no surprise that people usually believe it is our choices that put life in our own hands, and it is our choices that define who we are. The opening quotations express exactly this view.

This view – the view that we control our life through choice and our choices define the kind of person we are – has deep roots in many aspects of our life. And apparently, one of those aspects is moral responsibility. For many people, moral responsibility for a certain behavior seems to require a relevant choice. This has been the received view among not just the ordinary folks, but also philosophers who work on the issue of moral responsibility.

**A Conflict**

At a first glance, the received view seems rather intuitive. However, this truism becomes problematic when we take a closer look at our real practice. Recently, some philosophers start to call our attention to cases like the following. Suppose that my friend failed to receive me at the airport in time simply because he overslept. He did not make a choice to oversleep or to intentionally skip our meeting: there was simply no choice involved at all. But is he not blameworthy for failing to receive me at the airport? The answer is "No". In our everyday practice, finding out that one oversleeps usually provides us with a solid ground for blaming him for his failure to complete his duty. And we have good reason to

---

3 Of course, this is not to say that as long as we make a choice, we automatically meet all the conditions for moral responsibility. The current view is simply that having made such a choice is necessary for moral responsibility.
blame the person on that ground: the very fact that one oversleeps indicates that this person is not taking his duty seriously enough. If this is right, then there are occasions in which we hold people morally responsible for things that they did not choose to do.

Cases like this are not rare. Consider a few more.

*Wren*: Wren is on guard duty in a combat zone. There is real danger, but the night is quiet. Lulled by the sound of the wind in the leaves, Wren has twice caught herself dozing and shaken herself awake. The third time she does not catch herself. She falls into a deep slumber, leaving the compound unguarded.4

*Shylock*: Shylock received a message that Antonio's ships were lost at sea, he could not help but be instantly seized by a malicious rapture.

*White*: White was born in a very conservative town in a southern state during a period of time when the anti-black sentiments were at its peak. Inevitably, White was heavily influenced by these sentiments. Throughout his later years, he made careful effort to eliminate those sentiments in him. But today, when he sees a black family entering the public pool where he and his children are swimming, he cannot help but start to worry for the safety of his children.

There is obviously something blameworthy about Wren, Shylock and White for what they did or felt. Yet it is not obvious that what they did or felt is a result of their choice. White did not choose to feel threatened by the group of black people; he simply could not help it. Wren did not choose to fall asleep; on the contrary, she tried very hard to resist. If we agree with these intuitive judgments, then the following question arises: “Does moral responsibility really require choice?”

---

Classic accounts of moral responsibility, such as libertarian accounts, either explicitly assert or implicitly assume the volition condition – we are not morally responsible for a behavior unless we have, directly or indirectly, chosen it.\textsuperscript{5} I call this view \textit{volitionism}. As we have seen, there is good reason to believe in volitionism: only in choice can we find the requisite control over ourselves, and thus only through choice can we truly make a behavior our own.

\textit{Non-volitionism}, on the other hand, claims that no such choice is necessary to make a behavior our own; consequently, no such choice is required for moral responsibility. For example, some philosophers have argued that a behavior is attributable to us, and thus capable of grounding our moral responsibility, when the behavior is expressive of our attitudes, evaluative judgments, or normative commitments.\textsuperscript{6}

\textbf{The Main Questions}

(1) \textit{Is choice a precondition for moral responsibility?}

(2) \textit{If not, where can we locate the kind of control and ownership that are usually believed to be so crucial for moral responsibility?}

These are the two main questions that I intend to answer in this dissertation.

In Chapter Two, I will carefully examine the debate between volitionism and non-volitionism, and I will explain why we should reject volitionism. In Chapter Three, I will examine some current non-volitionist accounts of ownership and one enduring problem facing all these accounts – the \textit{Problem of Brain Manipulation}. In Chapter Four, I will

\textsuperscript{5} See, for example, Blum (1980), Slote (1992), Oakley (1992), Wallace (1994), Kane (1998), and Levy (2005).

propose an account of ownership that offers a solution to the Problem of Brain Manipulation. In Chapter Five, I will lay out a complete non-volitionist account of moral responsibility and argue why it is preferable to its alternatives.

Before we start, two clarifications are in order: first, the concept of moral responsibility; second, the concept of choice.

**The Concept of Moral Responsibility**

It is important to understand what exactly we mean by moral responsibility. When people talk about moral responsibility, they may mean two distinct things. On one conception – the commonsense one, to say that one is morally responsible for a certain behavior is to say that the agent owns that behavior in a way that makes him deserve certain treatment. Moral responsibility understood in this way is essentially a metaphysical notion – to determine whether one is morally responsible for a behavior is to find out whether that person possesses the requisite quality that establishes his ownership of that behavior. Classic libertarian view, for example, believes that the requisite quality is free will, and one is morally responsible for a behavior only if that behavior is a result of his free choice.

On the other conception, to say that one is morally responsible is to say that there is, overall, positive practical reason to treat that person in the relevant way (e.g. to blame or praise him). And moral responsibility understood in this way is essentially a practical issue – to determine whether one is morally responsible is to find out whether there is any practical justification for treating a person in a given way. An extreme example is the so-called social-regulation view, which claims that an individual is a suitable target for a
certain reaction “if and only if a reaction of this sort would likely lead to a desired change in the agent and/or her behavior.”\(^7\) In this case, the consequence of blaming or praising a person provides justification for the blame or praise.

P. F. Strawson famously argues that both conceptions of moral responsibility make the same fundamental mistake: they take what is internal to the notion of moral responsibility, attitudinal reactions involving blame or praise, which Strawson calls \emph{reactive attitudes}, “to be practical corollaries or emotional side effects of some independently comprehensible belief in responsibility.”\(^8\) According to Strawson, being morally responsible is nothing more than being an appropriate subject for certain reactive attitudes such as resentment, indignation, gratitude and approbation. What was taken to be mere practical corollaries or side effects, namely those reactive attitudes, are actually at the very core of the concept of moral responsibility. Being morally responsible for something is part of our interpersonal relationships – how we should view and treat each other. Reactive attitudes are a natural expression of an essential feature of this interpersonal nature of our way of life: they express “how much we actually mind, how much it matters to us, whether the actions of other people ... reflect attitudes towards us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other.”\(^9\) Therefore, this kind of expression “neither calls for, nor permits, an external ‘rational’ justification”,\(^10\) be it the libertarian ownership or better

---

\(^7\) Eshleman (2004).

\(^8\) Watson (1987), p.120.


social regulation; it is the kind of practice simply defined by our current social norms that govern interpersonal relationships.

Many philosophers find Strawson’s concept of moral responsibility very attractive: it recognizes a deep fact about moral responsibility – its intimate connection to interpersonal relationships. Regarding someone as morally responsible is, as Gary Watson puts it, “not just a matter of belief”, but “something in practice”. However, what most people disagree with Strawson is his claim that the appropriateness of reactive attitudes is completely determined by the current social practice. Think about slavery and homophobia. It used to be part of the social practice in some areas of the world that people had no problem with a human being selling another human being for profit; and it used to be, perhaps still is, part of the social practice that people condemn sexual behaviors between two persons of the same sex. But, clearly, just because the "current practice" condemns homosexuality and permits slavery, it does not follow that homosexuals are blameworthy and slave owners are innocent.

However, this is not to say that Strawson’s idea of taking the appropriateness of reactive attitudes (and thus ascription of moral responsibility) as an integrate part of our interpersonal relationships is a mistake. Notice that what Strawson actually emphasizes is that reactive attitudes require no “external” rational justification; but he does not say that justification \textit{internal} to interpersonal relationships are not permitted or required. So, perhaps if we look into the relevant interpersonal relationships, we may be able to find some internal rationale that enables us to illegitimate those problematic reactive attitudes we mentioned earlier. In fact, some philosophers, such as R. Jay Wallace, have proposed

this kind of amendment to Strawson’s original theory, and we shall return to a full examination of this new development in Chapter Four.

So, I understand moral responsibility in this revised Strawsonian sense: being morally responsible for a certain behavior means being an appropriate subject for certain reactive attitudes, such as resentment, gratitude, indignation and approbation, as a result of performing that behavior. And the appropriateness of the relevant reactive attitudes is subject to certain rational constraints that are internal to the relevant interpersonal relationships.

**Reactive Attitudes**

To better understand this revised Strawsonian conception, a few things need to be said about reactive attitudes. According to Strawson, reactive attitudes, such as indignation and gratitude, should be distinguished from another type of attitudinal reactions which he calls "the objective attitudes".

Objective attitudes, like reactive attitudes, are emotional reactions to external stimulus. Fear, upset, sympathy and joy are typical examples of objective attitudes. However, unlike reactive attitudes, objective attitudes need not be reactions to a moral agent like us. For example, if a rock naturally falls down and hits my car, I may feel upset; but my upset is not a reaction to any moral agent. When I am attacked by a deranged person, I will usually run away with fear; but my fear in this case is not quite the same reaction as what I would have when I was attacked by a level-headed person. According to Strawson, to adopt an objective attitude towards something is to see that thing "as an object of social policy; as a subject for what, in a wide range of sense, might be called
treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained ... "(p.52) It is important for Strawson that being an appropriate subject for objective attitudes does not entail moral responsibility for anything.

By contrast, reactive attitudes, according to Strawson, "are essentially reactions to the quality of others' wills towards us, as manifested in their behavior: to their good or ill will or indifference or lack of concern."(p.56) Reactive attitudes are not just emotional reactions; they are a particular group of attitudinal reactions, reactions to others' will or attitude – either towards us or towards things that we care – that are reflected in their behaviors. And to say one is an appropriate subject for reactive attitudes is just to say that he or she is morally responsible for something. This is how Strawson distinguishes reactive attitudes from objective attitudes.

However, Strawson's way of drawing the distinction is problematic. For example, when a psychopath attacks me, it can be true that the psychopath indeed has an ill will towards me – for instance, he may be cursing me while attacking, which indicates some ill will specifically towards me. In this situation, I may feel upset not just because I am under attack for no reason but also because the attacker has such an ill will towards me. Does this make my upset a reactive attitude rather than an objective attitude? If my upset is indeed a reactive attitude, and since it is appropriate for me to feel upset in that circumstance, it follows that the psychopath is indeed morally responsible for what he is doing. However, this seems wrong: we do not usually view a psychopath as a morally responsible agent. Therefore, my upset should be taken as an objective attitude rather than a reactive attitude, even if it is a reaction partly to a will.
Therefore, we need a better way to distinguish reactive attitudes from objective attitudes. Wallace, in his development of a Strawsonian theory, offers a better distinction. According to Wallace, what marks attitudes that are genuinely reactive from those that are merely of objective stance is that reactive attitudes are reactions to people as a result of our having certain normative expectations on them. A normative expectation is an expectation that one ought to do such-and-such or have such-and-such a belief, attitude or judgment. For example, "one ought not to torture an innocent baby for fun" is a normative expectation. We form normative expectations as a result of our acceptance of certain norms of the society.\textsuperscript{12} Once we accept a norm that one ought not to torture an innocent baby for fun, for example, we would naturally form the expectation that anyone who has the relevant capacities to understand and comply with that norm not torture a baby for fun. So, when a person whom we expect not to torture a baby for fun fails to meet our expectation, indignation, as a natural reaction, is what we feel towards that person.

By contrast, objective attitudes are not attitudes that we have by virtue of our having a certain normative expectation in the first place. I feel upset when a psychopath attacks me not because that I have a normative expectation that he ought not to do this to me. We do not have such normative expectation on psychopaths or animals because they lack the general ability to understand the norms and to comply with them, or as Wallace puts it, they lack \textit{the powers of reflective self-control}.\textsuperscript{13} It would be unfair, Wallace

\textsuperscript{12} See Wallace (1996), pp. 77-78. For Wallace, those norms are moral obligations. However, I think they should be understood more broadly as to include any norms that may concern our interpersonal relations, such as moral norms, prudential norms and even epistemic norms.

\textsuperscript{13} Wallace (1996), p. 155.
argues, to expect those who in general lack the powers of reflective self-control to comply with moral norms.\textsuperscript{14} So, since my feeling upset in this case is not a result of the psychopath's failing to meet a normative expectation that I have on him, my upset is not a reactive attitude but merely an objective attitude.

Following Wallace's treatment of reactive and objective attitudes, I shall take reactive attitudes to be reactions that result from our having a certain normative expectation on the agent. However, unlike Wallace, who seems to be mainly concerned with reactive attitudes accompanied by strong emotions, such as resentment, guilt and indignation, I believe reactive attitudes should be understood more broadly as to include reactions that are not accompanied by such strong emotions or not emotionally loaded at all. For example, sometimes we may find a wrongdoer blameworthy, but the wrongdoing is too minor to cause any such emotional reaction in us. Instead of reacting emotionally, we may simply decrease our respect for the wrongdoer.\textsuperscript{15} By respect, I mean respect in the moral sense, as opposed to respect in virtue of higher social status or greater power – it concerns how highly we think of a person in terms of their trustworthiness, moral worth and so on. We lower our respect if one fails comply with the norms, and we increase our respect if one transcends our normative expectation or successfully complies with a very demanding one.

In this dissertation, I understand reactive attitudes as including not just emotional reactions such as resentment and gratitude, but also attitudinal reactions such as readiness to increase or decrease moral respect. By highlighting the difference between emotional

\textsuperscript{14} This is what Wallace calls the \textit{Fairness Principle}, p. 161.

\textsuperscript{15} Litton (2007) and Lenman (2006), for example, make similar suggestions about reactive attitudes.
reaction and pure attitudinal reaction, I do not mean to say that readiness to increase or decrease respect is necessarily separated from moral sentiments like resentment or gratitude. Quite to the contrary, I think whenever we have those emotionally loaded reactive attitudes we also increase or decrease our respect at the same time. The point is simply that there are occasions when it is appropriate to increase or decrease moral respect but inappropriate to have stronger reactions such as resentment or gratitude.

Why Revised Strawsonian Conception

I have suggested that we shall understand moral responsibility in the revised Strawsonian way. One worry about this revised Strawsonian concept is that even if this concept correctly captures the practical nature of our commonsense notion of moral responsibility – the essential role that it plays in our interpersonal relationships, it fails to capture one equally important aspect of the commonsense moral responsibility – that is, ascription of moral responsibility entails attribution of some kind of control or ownership, an aspect that the metaphysical concept emphasizes. A concept of moral responsibility that fails to address this important part of our commonsense understanding of moral responsibility will not be adequate: it leaves out something that is also essential – the intimate connection between moral responsibility and the notion of personhood and identity.

I am sympathetic to this criticism. However, I believe the Strawsonian concept does not leave out this important aspect of our commonsense notion of moral responsibility. Here is why. We know that, according to the Strawsonian concept, reactive attitudes are reactions to other people’s will, affection or attitude. We also know
that when we resent or blame, we always resent or blame some person, not, or not just, the behavior or the will itself. As we can see, these two beliefs, putting together, give rise to the following question: “What connects the behavior or will to the person?” For example, if I am bothered by the fact that your hair grows too fast, it would be absurd for me to direct my frustration towards you. So, what does the behavior or will have to do with the agent such that the reactive attitudes aroused by the behavior or will should have the agent as its target, not the behavior or will itself? The answer, I think, must be: “The behavior or will is his own in the sense that it grounds his moral responsibility.” If this is right, then the Strawsonian concept must implicitly assume a relevant notion of ownership already, an ownership that can connect the behavior or will to the person such that what was initially aroused by the behavior or will flows eventually to the agent. So, contrary to its appearance, a certain notion of ownership is actually needed for the Strawsonian concept of moral responsibility.

I believe this revised Strawsonian concept has the potential to unify the practical aspect and the metaphysical aspect of our commonsense notion of moral responsibility, both of which can be an integrate part of the concept. This is why I take the revised Strawsonian concept as the concept of moral responsibility for this project.

The Concept of Choice

The concept “choice” also requires some clarification. This term is used to refer to a wide range of mental activities. Some call a deliberate decision to stop at a gas station a choice; some call one’s instinctively stepping on the brakes near a stop sign a choice. Presumably,
these are quite different mental activities. It is important, therefore, to make clear what kind of mental activities I refer to by “choice”.

Despite the aforementioned vagueness, what most volitionists have in mind when they talk about the volition condition is conscious choice, the kind of mental activity in which we are said to have control. For example, classic libertarian accounts, which take free will to be a necessary condition for moral responsibility, understand free will as such freedom that is present in a conscious choice. Descartes, in his discussion of free will, famously states “so that when a very evident reason moves us in one direction, although morally speaking we can hardly move in the contrary direction, absolutely speaking we can.” (Descartes (PWD/1984), p. 254) Peter Van Inwagen also writes “[to] say that one has free will is to say that when one decides among forks in the road of time (or, more prosaically, when one decides what to do), one is at least sometimes able to take more than one of the forks.” (Van Inwagen 2011, p. 422) The emphasis on conscious choice is also highlighted by some volitionists’ effort to link choice to direct control,\(^\text{16}\) which is a term often used to characterize the vigorous causal influence that only conscious deliberation can have. These remarks show that when libertarians talk about choice, they often have in mind, at least implicitly, conscious choice.

George Sher calls this type of view the searchlight view (Sher 2009, p. 2, p. 4): we are morally responsible only for behaviors that we consciously choose to perform, allow, or bring about. And it is conscious choice that non-volitionists like Sher reject as being necessary for moral responsibility. So, we hear Angela Smith’s contention, “most of us would probably say that choice or voluntary control is a precondition of legitimate

\(^{16}\) Wallace, for example, claims that conscious choices are subject to direct influence of reasons, and thus anyone “who lacks the capacity for deliberated choice will therefore lack the ability to control.” See Wallace (1994), p.158.
moral assessment ... yet ... we regularly do hold ourselves and others responsible for things that do not appear to reflect a conscious choice or decision.” (Smith 2005, p. 236)

Thus, it is conscious choice that is at the center of this controversy. In light of this, I distinguish between choice and non-choice mental activity by whether the mental activity involves some conscious decision-making process. Any decision-making process in which one consciously attends to one's intention, or one’s reason(s), or both is taken as a choice. For example, my decision to stop at a gas station upon realizing that my gas tank is nearly empty is a choice – in this case, I am aware of both my intention to stop and the reason why I form that intention. My decision to return the ping pong ball to my opponent’s left side is also a choice – in this case, I am aware of perhaps just the intention, not the reason behind it. By contrast, any mental process in which one does not consciously attend to either the intention or the reason is taken as a non-choice. For example, if I step on the brakes merely as a spontaneous response to an emergency situation, it is a non-choice mental process. Most of our automatic and subconscious (or unconscious) mental activities are thus non-choice by my criteria.17

Needless to say, this way of distinguishing between choice and non-choice mental activity is, by no means, adequate or accurate – it simply serves to highlight some most salient contrasts between these two categories, which are also most pertinent to the controversy at issue. So, although a more accurate and fine-grained distinction between

---

17 I leave it open whether there is a necessary connection between choice and voluntariness – this is why I prefer the term “volitionism” to “voluntarism”, which is the term used by some philosophers (e.g. Michael McKenna). I leave voluntariness aside because different people may understand it differently. If one understands being voluntary as being governed by a conscious will, then any mental process that is voluntary is a choice. By contrast, if one understands being voluntary as having the potential to be governed by a conscious will, then some mental activities that I take to be non-choice, such as subconscious exhibition of racist attitudes, turn out to be voluntary too.
choice and non-choice is definitely needed for a better understanding of these concepts, paradigm cases on the two extremes of the spectrum will suffice for our current purpose.

With these concepts clarified, we are ready to begin our investigation. The first question is whether moral responsibility requires the volition condition.
CHAPTER TWO

Responsibility and the Volition Condition

Does moral responsibility for a behavior require that one has, directly or indirectly, chosen it? Volitionists answer affirmatively. One reason for thinking so according to our commonsense view is that our choices reflect who we truly are and only through choice can we claim ownership of our behavior. But the question is non-choice mental activities such as attitudes, implicit judgments and habitual commitments can also reflect our character and personhood, sometimes in a more revealing way. They, too, seem to justify legitimate claim for ownership. So, why should we think that moral responsibility requires a relevant choice? In this chapter, I argue that volitionism is wrong. Let us start with a review of the current debate over this controversy.

Volitionism vs. Non-volitionism

Non-volitionists’ main weapon against volitionism is the type of cases that we mentioned early in the first chapter. In those cases, it appears that one can be morally responsible for behaviors that are not chosen. For example, Wren does not choose to fall asleep on duty, yet still she seems morally responsible for actually falling asleep. In fact, we frequently hold people responsible for things that are subconscious: unrecognized prejudice, involuntary negligence, bitter humor, etc. These cases show that moral responsibility does not always require a relevant choice. Thus, non-volitionists argue that we should jettison the volitionist account of moral responsibility for some alternative accounts. For
example, according to one non-volitionist account by Angela Smith – *attributivism*, “to say an agent is morally responsible for some thing ... is to say that that thing reflects her rational judgment in a way that makes it appropriate, in practice, to ask her to defend or justify it.” (Smith 2008, p. 369) So, “the condition of voluntary control should be replaced by a condition of ... judgment-dependence.”\(^{18}\) (p. 368)

In response, volitionists argue that these cases show only that moral responsibility for a behavior does not require that the behavior be *directly* chosen; but they do not show that the behavior need not be a result of the agent's choice at all. The agent’s responsibility for the current behavior may be grounded on some previous choice to which the current behavior can be causally traced.\(^{19}\) If Wren’s falling asleep, for example, is a result of her choosing to take some hypnotic drugs earlier, we may blame her for the current negligence based on that previous choice.

Thus, the key disagreement is whether, to be morally responsible for a certain behavior, an agent must make a choice at some point that causally leads to that behavior. Non-volitionists say *no*, and volitionists say *yes*. The strongest defense for volitionism for far is *the deep responsibility defense*: only choice, an exercise of the conscious will, can

\(^{18}\) It seems to me Smith moves a little too quickly here. Voluntary control may be understood in two different ways: first, one can be said to have voluntary control when she makes an actual choice; second, one can be said to have voluntary control when she possesses the capacity for choice but simply fails to exercise it. So, the kind of case we discuss here only prove that no actual choice is required for moral responsibility; they do not show that the capacity for choice is also completely irrelevant to moral responsibility. McKenna made a similar suggestion in McKenna (2008), p. 36. Furthermore, one may point out that whether it is appropriate to ask an agent to ‘defend’ or ‘justify’ a behavior – Smith’s own criterion for moral responsibility – may actually depend on whether that agent possesses the capacity for rational choice in the first place. So I think it is the exercise of choice, not the capacity for choice, that is called into question.

\(^{19}\) Wallace, for instance, argues that the moral fault in those seemingly non-choice behaviors can be traced to “an earlier episode of choice” – e.g. one's decision to take or not to take precaution to ensure something bad would not happen, or one's decision to develop or not to develop a certain attitude. See Wallace (1994), p.136. Robert Kane (1998, 2005) makes a similar move in his ultimate-responsibility account.
ground a *deep moral responsibility*, the kind of responsibility that explains the agent’s credit or fault; any account of moral responsibility that is short of this deep explanation of the agent’s credit or fault is too shallow. Non-volitionist accounts of moral responsibility, e.g. Smith’s judgment dependence account, are merely a “shallow” moral grading or description – it provides no basis for the agent’s credit or fault.\(^\text{20}\) We should, as one defender of volitionism, Neil Levy, puts it, distinguish between “having a faulty attitude” and “being at fault” (p.15): deep responsibility must account for why an individual is at fault, not just what kind of faulty attitude she has. In other words, an account of deep responsibility must explain how the agent comes to *own* the faulty attitude such that he can be properly said to be *at* fault. And to be *at* fault or to properly *own* the faulty attitude, many volitionists believe, one has to exercise his will, that is, to make a choice.

Non-volitionists can try to refute this *deep responsibility defense* in three different ways. First, they have argued that their notion of moral responsibility is indeed a deep one. For example, Smith, in her defense of her account, argues that the expressiveness of one's evaluative judgments in a behavior deeply implicates her practical agency – in particular, her deep assessment of reasons, which provides sufficient ground for a deep moral responsibility (Smith 2008, pp. 386-7). However, volitionists can respond that this kind of deep assessment, interesting as it is, still fails to provide a kind of agency as robust as what deep responsibility requires, namely, an exercise of conscious will. Insofar as volitionists take deep responsibility to essentially consist in a fault or credit that can only be found the exercise of conscious will, it seems impossible for any non-volitionist account to accommodate this particular “deepness” that volitionists demand.

\(^{20}\) For more discussions about Deep vs. Shallow Responsibility, see Wolf (1990), and Levy (2005).
Second, non-volitionists have responded by providing further cases to show that no such deep responsibility is actually required for our reactive attitudes. Consider the following case from Paul Litton.

_Daycare._ Imagine that I take all the precautions necessary to pick up my child at the end of the day. I write on my calendar that it is my turn, not my wife's, to pick up our child. I tell my boss that I have to leave a bit early to get to the daycare center before it closes. I even tie a string around my finger to remind myself. Nevertheless, as I am driving, I listen too intently to music, my mind wanders, and in accordance with my normal driving pattern, I drive home, forgetting to go to the daycare center.

(Litton 2007, p. 76)

It seems that the agent’s failure to go to the daycare center cannot be traced to any obvious choice of his; rather, all his conscious choices indicate his resolution to prevent such a failure. So, the fact that we still view the agent as blameworthy for his failure suggests that reactive attitudes do not require deep responsibility. However, volitionists can respond that there is indeed some choice to which the current failure can be traced – e.g. the agent’s decision to turn on the radio or his decision not to turn it off. It seems extremely hard to find a case in which there exists no such relevant choice at all to which a volitionist can point his finger. Thus, just by providing further cases like this, non-volitionists are not likely to have a convincing victory over volitionism.

Last, non-volitionists can argue that, in many cases, no such deep responsibility is available. According to our commonsense view of moral responsibility, anyone who is morally responsible must meet a certain epistemic condition. One popular view on the epistemic condition is that in order for an agent to be morally responsible for an outcome,
she has to be aware of what she is doing and the relevant outcome. When volitionists try to explain responsibility for the current behavior by appealing to a previous choice, a problem emerges: when the relevant choice was made, the agent might not foresee the causal influence of that choice on her future behavior. For example, one’s failure to pick up his child may be a result of his choice to turn on the radio. But when he made that choice, he might not foresee that the music would distract him completely from his plan. Indeed, if he had recognized that, he probably would not have turned on the radio. Thus, even if one made a causally relevant choice in the past, since he failed to meet the requisite epistemic condition, he is still not morally responsible for the outcome, i.e. the subsequent behavior. Consequently, appealing to previous choices does not help volitionists explain the agent’s responsibility in those cases.

This difficulty, however, does not present an insurmountable challenge to volitionism. For example, volitionists can reject the proposed epistemic condition: such a condition is too strong. Maybe responsibility for outcomes does not require the agent to actually foresee the outcome; rather, it simply requires that the agent be able to foresee the possibility of the outcome. So, even if the agent did not actually foresee that turning on the radio would distract him from his plan, he could have foreseen, at least, the possibility of the distraction – thus, the agent does meet the requisite epistemic condition and is indeed morally responsible. On the other hand, if it is the case that the agent could

21 A precise account of the epistemic condition remains controversial. But, a popular view on the epistemic condition, according to Sher, is that ‘an agent's responsibility extends only as far as his awareness of what he is doing’. See Sher (2009), p. 4.

22 Michael McKenna discusses this problem for volitionism in McKenna (2008).

23 See, for example, Ginet (2000). Ginet’s actual account is more complex than this one. But, clearly, he does not think that the actual knowledge of the outcome is always necessary.
not have even foreseen the possibility of the relevant outcome, then volitionists can insist, contrary to non-volitionists’ intuition, that the agent is indeed not morally responsible for the subsequent behavior. Since cases of the second kind are rare, denying responsibility in such cases is an acceptable price for volitionism.

Therefore, none of the three responses successfully defeats the deep responsibility defense. In what follows, I would like to propose a different argument against this deep responsibility defense, an argument that targets its basic assumption – only through choice can we come to properly own a faulty attitude, that is, to be at fault.

**What Is So Special about Choice**

What is so special about choice such that only through choice can an individual come to own a fault or credit, and thus bear deep responsibility? What unique power is there in choice? In this section, I will argue that there is no such power that belongs uniquely to choice. My strategy is to compare a paradigm case of choice to some typical cases of non-choice mental activities, and to show that no deep-deep-responsibility-generating power, whatever it is, can be found uniquely in choice. If this is successful, then we have no reason to believe that only through choice can we own a fault or credit.

Consider a paradigm choice scenario. You come across a building on fire and observe four people leaping off from the top of the building – one a few yards to your left and three a few yards to your right. You instantly notice a safety net nearby and you realize that it can be used to save some of them. You deliberate on which side you should place the net, and you decide to put it beneath the three, letting the one die.

---

24 Just to be very clear here: my argument is not that choices do not have any responsibility-generating power, but that whatever deep-responsibility-generating power it has, that power is shared by, at least, some non-choice mental activities.
(Not all choices need to involve such a fully conscious deliberation. In many cases, we attend to merely part of our reason, or even just the intention itself. But if I can show that no unique deep-responsibility-generating power can be found even in choices that involve such robust awareness, then I show no such power can be found in less robust cases.)

Now, compare that paradigm case of choice to some typical cases of non-choice: you see an old lady falls on the ground in a funny way, and you spontaneously laugh; or while you are engaging in an intense philosophical debate with your companion, your body instinctively avoids running into other pedestrians in the street. The mechanism that underlies the spontaneous laughter or the navigation system runs automatically. It does not require your consciously attending to your circumstance. And it need not be a result of any previous choice – it could be an innate mechanism that you have not been able to stifle at all.

What are the differences between that paradigm case of choice and these typical cases of non-choice? One obvious difference is that, in the choice scenario, you have robust awareness: you are aware of the motive, your options, your assessment of the options, the formation of intention and the resulting behaviors; whereas in the non-choice mental activities, usually you are aware of just the resulting behaviors. Another crucial difference is that, in the choice scenario, you seem to have control over your thoughts and behaviors; but in the non-choice scenarios, your thoughts and behaviors occur automatically.\textsuperscript{25} If choice has any unique deep-responsibility-generating power, we

\textsuperscript{25} The Dual-Process theory in the recent cognitive science literature may also help highlight the differences between choice and non-choice. Very crudely, the Dual-Process theory holds that there are two basic classes of mental processes: in one class, System 1, mental processes are relatively slow, controlled, computationally demanding, introspectively accessible, inferential and
should find it in these differences.

Let us start with awareness. Suppose that the unique deep-responsibility-generating power lies in awareness. But in what kind of awareness does that power lie? The closest connection between responsibility and our awareness is the epistemic condition – people must be aware of what they are doing and the moral consequence of their behavior in order to be morally responsible for that behavior.\textsuperscript{26} However, I could be fully aware of these things in non-choice mental activities. For example, I can be fully aware that I am laughing and that laughing at others' misfortune is inappropriate, yet I cannot help but laugh. A soldier on guard may be aware that she is falling asleep and that she should not fall asleep at that time, yet she simply cannot help it. So, if the unique power lies in these kinds of awareness, there is no reason to think that it belongs uniquely to choice.

Control seems a better place to find the alleged unique power. It seems reasonable to say that to proper own a fault or credit, we must have proper control over the fault or credit generating process. And if we have control only when we make a choice, then we can properly own a fault or credit only through a choice. But, do we have control only in choice? To answer this question, we must first make clear what exactly control means. It is hard to spell out the exact meaning of control. We may, however, borrow John M.

\textsuperscript{26} This is a rather strong epistemic condition. The idea here is that if non-choice can meet even such a strong epistemic condition, it can certainly meet those weaker ones.
Fischer's classification of two types of control here: guidance control and regulative control.²⁷ Let us start with guidance control.

**Guidance Control**

Guidance control, according to Fischer, consists mainly in a so-called *moderate reasons-responsive mechanism*. An agent has guidance control over a behavior insofar as the agent has a decision-making mechanism such that there exist some possible worlds in which there is a sufficient reason to do otherwise, and that same mechanism does otherwise because of that reason.²⁸

But if reasons-responsiveness is indeed what the alleged control amounts to, then we do have control in some non-choice mental activities, since the mechanism underlying some of those mental activities can also be reasons-responsive. When I have an intense philosophical debate with my companion, the mechanism that navigates my body through other pedestrians is usually automatic and involves no conscious choice. However, this mechanism can generate the exact results that a typical moderate reasons-responsive mechanism would generate: if a pedestrian were to approach me from a different angle and at a different speed, I would have reacted differently (without thinking about it). Likewise, in the case of spontaneous laughter, the mechanism at work can be reasons-

---


²⁸ Fischer's actual account of moderate reasons-responsiveness is more complicated than this. For example, apart from this counterfactual reasons-responsiveness, it further requires that the mechanism be regularly receptive to reasons, weakly reactive to reasons and responsive to a wide range of reasons, including moral reasons. See Fisher & Ravizza (1998), p.72.
responsive too: if the lady falls in a less funny way, I might not laugh; if she hurts herself badly, I might not even feel like laughing at all; and so on.\textsuperscript{29}

Recent developments in cognitive science give further support to the previous point. According to some theorists, our \textit{practical reasoning} mechanism consists of not only conscious and voluntary mental processes, but also subconscious and automatic processes.\textsuperscript{30} And they suggest that automatic mental processes are rather prevalent in our mental life. If automatic processes are also at work in our practical reasoning, it further supports the idea that they can be responsive to reasons – though sometimes in a defective way.\textsuperscript{31}

Thus, there is no good reason to think that reasons-responsiveeness belongs uniquely to choice. If the alleged control is indeed guidance control, we can be morally responsible for behaviors that are results of non-choice.

Some philosophers, however, insist that the alleged control should be understood as regulative control. So, let us turn to this second type of control.

**Regulative Control: Event-Causal View**

\textsuperscript{29} It is not hard to see how those further conditions for moderate reasons-responsiveness can also be met by non-choice mental activities. For example, the automatic navigation mechanism that is at work in the walk-and-talk situation can reflect regular rational patterns, and thus be regularly receptive to reasons.

\textsuperscript{30} For an example of a dual-process account of reasoning, see Jonathan St. B. T. Evans (2003), ‘In Two Minds: Dual-Process Accounts of Reasoning,’ \textit{TRENDS in Cognitive Science} 10: 454-459. In that article, Evens cites a number of interesting examples that show that some automatic processes are at work even in such high-level reasoning as deductive reasoning.

\textsuperscript{31} One example of defective automatic reasoning is the Tversky and Kahneman study (October 1983). In that study, subjects are presented with a task to assess the likelihood of some statements about Linda, who is described as deeply concerned with issues of discrimination and social justice. Most subjects judge that the statement, ‘Linda is a bank teller and is active in the feminist movement,’ is more likely to be true than the statement, ‘Linda is a bank teller.’ Psychologists ascribe this \textit{conjunction fallacy} to some automatic and heuristic reasoning mechanisms in us.
Regulative control is understood by Fischer as a dual power: “the power freely to do some act \( A \), and the power freely to do something else instead.” (Fischer and Ravizza, 1998, p. 31) Regulative control is said to be crucial for moral responsibility because the power to do otherwise is considered, by many people, a necessary condition for moral responsibility. This position is implied by the principle of alternative possibilities: we are not morally responsible for an act unless we could have done otherwise (Frankfurt 1969, p. 829). Is the power to do otherwise the unique deep-responsibility-generating power in choice? To answer this question, we need first to make clear what exactly the power to do otherwise amounts to. For some, the power to do otherwise is grounded in undetermined decision-making events.\(^{32}\) For others, the power to do otherwise is grounded in the so-called agent-casual power, the power to arbitrarily activate a causal sequence in the brain.\(^{33}\) Let us consider them in turn.

According to the event-causal view, the power to do otherwise is grounded in undetermined mental events: that is, to say that one has the power to do otherwise is to say that, at least, some of her mental events in which one of several possible courses of action becomes the actual course of action are undetermined events. For the sake of the argument, let us suppose that the world indeed contains indeterminacy. But even if indeterminacy is true, why should we think that only choices are undetermined? It seems

\(^{32}\) One such account is Robert Kane's ultimate-responsibility account. According to this account, an agent is morally responsible for an act only if that act is partly a result of a (previous) will-setting act – an undetermined event in which the agent randomly takes one of several equally appealing options, all of which genuinely reflect the agent's character and reason. See Kane (2005), pp. 127-8, pp. 135-7.

\(^{33}\) According to the agent-causation account of moral responsibility, to be morally responsibility for an act, the agent must willingly activate the causal sequence that leads to the act while such activation is not itself caused by anything external to the agent or by any other exertion on the part of the agent. For some recent discussions of agent-causation, see William Rowe (1995, 2003), and Keith De Rose (1993).
reasonable to think that at least some non-choice mental activities can be undetermined as well.

One may appeal to the apparent unpredictability of our choice to explain the asymmetry between choice and non-choice. But many non-choice mental activities also seem to be unpredictable. For example, seeing another pedestrian approaching me, my automatic navigation mechanism may make me move to the left or to the right – there is no guarantee that it will always make me move to one side. There also seems to be unpredictability in the falling-asleep case – there is no set period of time that the mechanism at work will take to put Wren to sleep, and it may not always succeed in putting her to sleep. If the apparent unpredictability of choice must be explained in terms of its being undetermined, there is simply no reason why the unpredictability of those non-choice mental activities cannot be explained in the same way. So, if the power to do otherwise is grounded in undetermined events, then we have regulative control in choice as well as non-choice.

It should be noted that, for many libertarians, there is more to regulative control than the mere undetermined-ness. For example, Robert Kane famously advocates the ultimate responsibility account of moral responsibility. For an individual to be ultimately responsible for a behavior, not only must the behavior be a result of some will-setting act that is itself an undetermined events, but also the will-setting act must meet what Kane calls the “plurality conditions” (2005, pp. 128-30). To meet these conditions, roughly the decision must be made in accordance with one’s character, with the necessary knowledge, and as a result of one’s reason. However, there is no reason to think that non-choice mental activities cannot satisfy all these conditions for ultimate responsibility. We have
already seen how non-choice mental activities can be undetermined. It is not hard to see how they can also meet the “plurality conditions”. For example, people like Sher and Smith have made a strong case that behaviors that are not results of choice can be expressive of one’s characters, judgments and reason. Indeed, this is exactly why they argue that non-choice mental activities can be sufficient ground for moral responsibility. Thus, we can be ultimately responsible for, at least, some non-choice mental activities.

**Regulative Control: Agent-Causal View**

Alternatively, we may understand the power to do otherwise in the agent-causal sense. According to the agent-causal view, the power to do otherwise is grounded in an arbitrary power of the conscious will, which flows free of the compelling force of reason or impulse and can activate a causal sequence arbitrarily. There are serious doubts about the plausibility of this agent-causal power. But even if we grant that this agent-causal power exists, there is still no reason to think that we have the power to do otherwise only when the behavior is, directly or indirectly, chosen.

To start, we may view a non-choice mental process as a process from which one's conscious will currently stays away. If the conscious will, as the agent-causal view claims, really has the power to arbitrarily decide its own course, it must then have the power to step into, at least, some mental processes from which it currently stays away. Consider the walk-and-talk case again. Even though the navigation mechanism is currently not governed by my conscious will, but if I indeed have the power to arbitrarily steer my conscious will, then my conscious will should be able to stop my current conversation,

---

34 Many philosophers find agent-causation highly implausible, for it seems radically at odds with our scientific knowledge. For a good discussion of the problems for the agent-causation view, see Pereboom (2001), Chapter 3.
take over the automatic navigation mechanism, and decide consciously to, say, halt for a few seconds. So, if we indeed have such arbitrary agent-causal power, we should be able to do otherwise even in a non-choice mental activity.

Some may argue that if the conscious will intervenes, then the mental activity ceases to be non-choice – this simply proves that we have the requisite control only in choice. But this argument misses the point. It is true that the agent-causal power exists in the exercise of the conscious will, but it does not follow that we could have done otherwise only in conscious choices. Remember, when we have a non-choice mental process going on in our brain, our conscious will does not necessarily cease to work. When my automatic navigation mechanism guides my body through a crowd, my conscious will is constantly at work at the same time – to keep my philosophical conversation going. So, we exercise the agent-causal power even when some automatic mental processes are going on in the brain. If the agent-causal power is indeed of such arbitrary nature as the agent-causal view claims, then it should enable my will to arbitrarily step into, at least, some non-choice mental processes whenever I exercise that power. This is to say, I could have done otherwise and thus have the requisite control in, at least, some non-choice mental activities too.

Some may further argue that moral responsibility for a behavior does not just require the possession or exercise of the agent-causal power; it also requires that the behavior in question enter our consciousness as an option when we are exercising our agent-causal power to make a decision. For example, they may argue that one can be responsible for acting or not acting only when one consciously contemplates that option and decides arbitrarily to take it or not to take it. And since non-choice does not enter our
consciousness, we cannot be responsible for a non-choice behavior. But this cannot be right. Consider the following case: I see a child drowning in a pool; I contemplate two options – jumping into the pool to save him or leaving him to die; and I decide to leave him to die. Suppose that there is yet a third option – calling for help, but that option simply fails to register in my mind during my contemplation. Am I, therefore, not morally responsible for not calling for help? I think the reasonable answer is ‘No’. We can be morally responsible for failing to consider an option that we could have considered, just as we can be morally responsible for outcomes that we failed to foresee but could have foreseen. If this is right, then we should be morally responsible for non-choice behaviors too. This is because non-choice mental processes are nothing but mental processes from which our conscious will currently stays away – in other words, they are, just like the option of calling for help, options that fail to enter our consciousness. So, if we can be morally responsible for failing to call for help, why cannot we be responsible for those non-choice mental activities? Thus, if we indeed have the arbitrary agent-causal power, we should be responsible for, at least, some non-choice behaviors.

Masked Abilities

There is yet another way to understand the power to do otherwise. According to some compatibilists, the power to do otherwise should be understood as a general ability or disposition.\(^{35}\) For example, according to Michael Fara, to say that an agent had the ability to do otherwise is to say that “she was disposed to act otherwise if she tried to act otherwise”. (Fara 2008, p. 853) This general disposition or ability, however, can be \textit{masked} in particular circumstances. For instance, a good ping pong player has the ability

to perform a beautiful forehand loop, yet he may fail to manifest that ability if he is playing with an unfamiliar paddle: the use of an unfamiliar paddle masks that ability (without undermining it). Now, according to this view, one can still be said to have the ability to do otherwise even if she actually could not have done other than what she did in a particular circumstance – her ability to act otherwise was masked in that circumstance. One’s moral responsibility, then, is grounded upon the possession of the general abilities or dispositions to rationally form beliefs and desires that would lead her to do otherwise (Fara 2008, p. 856). If these general abilities or dispositions to rationally form beliefs and desires are the alleged deep-responsibility-generating power, it is clear that moral responsibility does not require choice: the absence of an actual choice does not mean the absence of those general abilities or dispositions in the agent. So, again, we have no reason for thinking that we have the requisite control only in choice.

**Self Control**

Last, let us consider a quite different kind of control. Neuroscientists and psychologists talk about self control, which is usually understood in terms of internal neurobiological constraints.36 In other words, self control is understood as the degree to which the subject is truly psychologically capable of avoiding a certain mental activity. For example, a person who suffers from compulsive disorder does not have self control over certain behaviors because the degree to which she is psychologically capable of avoiding them is very low. Now, one may argue that the degree to which we are psychologically capable of avoiding automatic behaviors is usually very low, and this is why we lack the relevant

---

36 See, for example, Churchland (2002), p. 211-2. Churchland believes that control understood this way comes in degrees, and therefore no sharp distinction can be drawn between what we typically view as voluntary acts and what we typically view as involuntary acts.
control in non-choice behaviors. However, self control cannot be the kind of control that volitionists are looking for, because we do not have that kind of control in many choice situations either. In a lot of choice situations, the degree to which the agent is actually psychologically capable of avoiding a certain decision can be as low as that in many automatic mental processes – the more reasonable (moral ...) you are, the less capable of avoiding a rational (moral ...) decision. Thus, self control cannot be used by volitionists to justify the choice/non-choice dichotomy in moral responsibility, for it would render them unable to accommodate a large number of cases involving decisions hard to resist.

**Conclusion**

After a careful examination of several possible candidates for the deep-responsibility-generating power, we found that none of them belongs uniquely to choice. Thus, if there is indeed deep responsibility, it does not belong uniquely to choice. Unless further accounts of the unique deep-responsibility-generating power present themselves, we have no good reason to believe that only when we have directly or indirectly chosen a behavior can we come to own a fault or credit.

People usually believe that there is a *fundamental* difference between our conscious will and our automatic mental processes. This belief has been seriously challenged by psychologists and neuroscientists. Daniel M. Wegner, for example, argues that there is no fundamental difference in the causal path between seemingly intentional actions and unintentional actions: both of them are caused by the actual and unconscious causal events in the brain (Wegner 2002, pp. 67-8). However, even without such scientific evidence, a simple reflection on our own experience well discloses to us that
the long-held centrality of conscious will in our cognitive life is much exaggerated. Consider just how a motive usually emerges in our mind (e.g. how I suddenly desire to help someone), how our computation is usually done (e.g. how I come to the conclusion that three lives are weightier than one life), and how we usually find some information in the environment to be relevant to the task at hand (e.g. how I find that the safety net nearby can be used to save lives) – most of these mental processes are automatic. Automaticity is everywhere. And there are definitely good reasons why this is the case – to the least, automatic processes tend to be cognitively more efficient. All these show that, to the least, the contrast between choice and non-choice has been over-emphasized. Awareness, agency, control and automaticity all come in degrees. So, it is highly questionable that we can draw a clear line between choice and non-choice such that one is so essential to moral responsibility while the other is completely irrelevant.

That being said, it is undeniable that there are still some important cognitive differences between paradigm cases of choice and typical cases of non-choice. Choice has some obvious cognitive advantages over non-choice – to the least, it enables more sophisticated computations and development of new strategies to cope with ever-changing environmental challenges. By contrast, this creativity is usually absent in non-choice mental activities, which are mostly established patterns. But the absence of exercise of this creativity, by no means, implies the lack of the general creative capacity in the agent. And, not to forget, our choices are not always creative and more sophisticated.

Thus, I conclude that the volition condition is simply too strong – it excludes many non-choice mental activities through which we can be morally responsible for our
behaviors. But if choice is not the answer for the kind of ownership that we are looking for, where can we find the answer? It is now time to turn to this second question.
CHAPTER THREE

Ownership and the Problem of Brain Manipulation

Being morally responsible for a certain behavior requires some of kind of ownership. The received view is that the essential mark of this requisite ownership is *having made a relevant choice*. We have seen why this received view is ungrounded: after examining a variety of potential responsibility-generating powers, we found that none of them belongs uniquely to choice. Thus, choice is not the right place to locate this requisite ownership.

Non-volitionists have proposed their own ownership-based accounts of moral responsibility. Instead of looking for an ownership that builds on some kind of robust control, such as the power to do otherwise, almost all non-volitionists believe that a weaker kind of ownership will suffice – a kind of ownership that better fits into the narrative of contemporary natural science. One thing that all theorists on moral responsibility agree upon is that some kind of practical reasoning capacities are crucial for being a morally responsible agent. Thus, for many non-volitionists, a natural place to locate the desired ownership is *the exercise of the relevant rational capacities*, which includes both conscious and non-conscious exercise.

However, all these non-volitionist proposals face one serious challenge – the Problem of Brain Manipulation. To see why the Problem of Brain Manipulation presents such a serious challenge, let us start with some current non-volitionist accounts of moral responsibility and the different notions of ownership implied in these accounts.
**Endorsement Account**

Some philosophers try to capture the idea of *being one's own* in terms of *being endorsed by one's deep self*. David Hume is, perhaps, among the first to suggest this. According to Hume, an individual is responsible for an action only if the action reflects his or her "durable or constant" quality of mind, such as character.\(^{37}\) The action is one's own because it is "endorsed" by those durable or constant qualities, because unlike those "temporary and perishing" qualities, these durable and constant qualities have "a close and lasting relation to the agent".

Harry Frankfurt further develops this endorsement idea. According to Frankfurt, there are two different types of desires: first-order desires are desires to act or behave in a certain way – e.g. desires to watch movie or eat ice-cream; second-order desires are desires about the first-order desires – e.g. desires to rid oneself of the desires to eat more ice-cream.\(^{38}\) First-order desires can be in conflict with second-order desires. For example, one may have a first-order desire to eat more ice-cream, but at the same time have also a second-order desire to control one's appetite for ice-cream. For Frankfurt, having second-order desires is a distinctive feature of *personhood*. Only when one acts in accordance with her second-order desires, the act reveals her deeper self: it reflects her will not merely at the surface level but also at the level of more rational and thoughtful reflection. Accordingly, Frankfurt proposes that moral responsibility for action requires that the action be endorsed by the agent in a special way – that is, the agent has the will to do it and she has that will because she has the second-order desire to have that will.


\(^{38}\) See Frankfurt (1971).
These accounts of moral responsibility employ an endorsement notion of ownership: for an action or behavior to be one's own, that action or behavior must be endorsed by her deep self – her "durable or constant" characters or her second-order desires.

**Attributivist Account**

Following T. M. Scanlon (1998), philosophers like Angela Smith hold a *rational-relations* view of moral responsibility. According to this view, moral criticism – a negative form of holding somebody morally responsible, addresses, by its very nature, a particular demand to its target: "it calls upon the agent to explain or justify her rational activity in some area, and to acknowledge fault if such a justification cannot be provided." Since only when an activity reflects one's rational judgment or commitment, can it be appropriate to demand a justification of it, being morally responsible for some activity means that the activity reflects an attitude or judgment in a way that makes it appropriate to ask the agent to defend or justify it.

---

39 Scanlon's view on blame undergoes some significant change in Scanlon (2008). Scanlon (2008) distinguishes blameworthy from blame. To judge someone to be blameworthy is to judge that his or her conduct shows something about his attitudes or judgments that impair the relations one has with them. By contrast, to blame someone requires no such attitude-based judgment: it is simply to form a judgment that the other impairs a certain interpersonal relationship with you, and "to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate." Whether an interpersonal relationship is impaired or not is completely defined by that relationship at issue. This view of blame is definitely not a volitionist view. But it is not completely clear that it is an attributive view either.


On the basis of this rational-relations view of moral responsibility, Smith proposes the following condition for morally responsible. What determines moral responsibility for a behavior (or the appropriateness to ask for a justification for a behavior) is whether that behavior is indicative and expressive of one's evaluative judgments.\textsuperscript{42} Evaluative judgments, according to Smith, need not be states of mind that are consciously recognized, rather they are "tendencies to regard certain things as having evaluative significance."\textsuperscript{43} My forgetting to attend my friend's wedding can indicate my evaluative judgment that I do not take our friendship quite seriously, but I may, at the same time, fail to explicitly recognize that judgment. Thus, to say something is indicative or expressive of a certain evaluative judgment is not to say that one must be consciously motivated by that evaluative judgment; rather it means just that there is a "normative connection" between that thing and that particular evaluative judgment. A mental state or behavior is normatively connected to a particular evaluative judgment just in case "if one sincerely holds [that] evaluative judgment, then the mental state [or behavior] in question should (or should not) occur."\textsuperscript{44} One example Smith gives is this: "[If] I sincerely judge that there is nothing dangerous or threatening about spiders, I should not be fearful of them."\textsuperscript{45} What Smith means by "normative connection" seems to be that the evaluative judgment in question provides good reason for believing that the agent will perform or not perform the behavior in question. So, staring at the sight of a spider is expressive of

\textsuperscript{42} Smith (2005), p. 251; Smith (2008a), p. 368.

\textsuperscript{43} Smith (2005), pp. 251-251.

\textsuperscript{44} \textit{Ibid.}, p. 253.

\textsuperscript{45} \textit{Ibid.}, p. 253.
an evaluative judgment – "spiders are dangerous", because that the evaluative judgment provides good reason for believing that I will startle at the sight of a spider.

Smith calls this account of moral responsibility *attributivism*. What underlies this attributivist account is the idea that the kind of ownership required for moral responsibility consists in a normative connection between the behavior in question and some evaluative judgments in the agent: the behavior *is expressive of* those evaluative judgments. As long as this normative connection is held, the behavior is attributable to the agent.

**Causal Account**

Some philosophers endorse a causal notion of ownership in their account of moral responsibility. Hume, at some point of his discussion, claims that a necessary condition for moral responsibility is that the action be caused by one's “desires and willings”, which seems to suggest that, alongside his endorsement idea, causation is also a necessary condition for ownership. Other philosophers take a more straightforward causal approach to ownership, one of which is George Sher.

Sher is also mainly concerned with the negative form of moral responsibility, namely, moral blame. Sher understands the concept of blame in terms of a two-tiered structure. To blame \( X \) is to have certain attitudinal and behavioral dispositions, and these dispositions are derived from a belief-desire pair of those who hold \( X \) for blame. Such a belief-desire pair consists of a belief that \( X \) has acted badly or has a bad character, and a desire that \( X \) not have acted badly or not have that bad character. To blame someone is,

---

47 Sher (2006b), p. 112.
first and foremost, to hold a belief-desire pair of this kind, which forms the first tier of blame. Also constitutive of blame is the set of blame-related attitudinal and behavioral dispositions, such as indignation, resentment and disdain, which form the second tier of blame.

Much like Wallace's idea of normative expectation, Sher explains the origin of the belief-desire pair in relation to our normative commitment. When we come to accept certain moral or prudential norms of the society, we form a relevant normative desire, which, according to Sher, is conceptually entailed by one's acceptance of the norm. To say one accepts a certain norm is to say that one desires people to behave or think in the relevant way. Or, as Sher puts it, the full acceptance of a moral principle is "conceptually bound up with" desiring that the principle not be violated. The second-tier blame-related attitudes and behaviors arise as a natural reaction to the frustration of that normative desire.

Like the Strawsonian conception of moral responsibility, Sher's two-tiered conception of blame also faces the following question: if our blame-related attitudes and behaviors are simply reactions to the frustration of our normative desires, which is caused by people's performance, then why should the target of those attitudes or behaviors be the agent, not just their performance? The answer, again, must be that the performance is the agent's own. Thus, one key determining factor of whether one is a suitable subject for blame in Sher's view is whether the performance can be properly said the agent's own.


49 *Ibid.*, p. 112, and p. 115. As we shall see, this concept is, in many aspects, similar to the normative Strawsonian framework that I will discuss in Chapter Four.
Like other non-volitionists, Sher does not believe that it requires a relevant choice to make the performance one's own. But, *contra* non-volitionists like Smith, Sher also does not believe that a mere normative connection is robust enough to ground the requisite ownership. I think there are good reasons for Sher to reject the mere normative-connection idea of ownership. Suppose that, when Wren falls asleep on duty, she has, at the same time, an evaluative judgment that her duty is not important and she really needs a good sleep. But suppose also that her falling asleep is not actually caused by that evaluative judgment – it is simply a result of food poisoning. In this case, surely we cannot attribute Wren's falling asleep to herself – her agency simply has nothing to do with her falling asleep, even if there is a "normative connection" between her actual performance and the evaluative judgment that she has in mind. Therefore, a mere "normative connection" is not enough to establish the kind of ownership needed here. There must be a more robust relation between the behavior and the relevant evaluative judgment or attitude that one has in mind.\(^50\)

According to Sher, the kind of relation needed here is a causal relation: the questionable behavior can be causally traced back to the "interplay" of a collection of "desires, beliefs, and fine-grained dispositions" within the agent, and these psychological elements "combine to make him the person he is, while through their interaction, they give rise to his failure to act as he should."\(^51\)

\(^{50}\) It has to be pointed out that it is not entirely clear whether, on Smith's view (2008), the causal condition is completely irrelevant. At some places, Smith criticizes Sher for putting too much emphasis on the causal connection between a behavior and the relevant evaluative judgment or attitude, which seems to suggest that she does not completely reject a causal component.

In summary, we have looked at three different proposals for the kind of ownership that is believed to be necessary for moral responsibility. Some suggest that a behavior is one’s own if it is endorsed by the agent's deep self. Some suggest that a behavior is one’s own if it is indicative or expressive of one's evaluative judgments. Some take a straightforward causal approach and suggest that a behavior is one’s own if it can be causally traced back to some conscious or non-conscious “interplay” of the agent’s rational capacities, such as attitudes, judgments and other brain activities.

Admittedly not all these accounts of moral responsibility can be properly called a non-volitionist account. Hume and Frankfurt, for instance, seem to put some emphasis on volition – though I doubt that this emphasis on volition is essential to their accounts. Nevertheless, these accounts can all be called a compatibilist account, an account that takes moral responsibility to be compatible with the truth of causal determinism. And it is no surprise that the kind of ownership adopted in these accounts is much weaker than those incompatibilist accounts of ownership, which build upon more robust control such as the agent-causal power.

One challenge the compatibilist accounts face is to explain why this weaker kind of ownership, rather than a stronger one, is enough for a reasonably intuitive concept of moral responsibility and a reasonable adequate account of it. However, there is a more pressing challenge: it is argued that these compatibilist accounts of ownership have serious internal tension, which makes them theoretically incoherent. This challenge is best illustrated by the Problem of Brain Manipulation (or Mind Control)\textsuperscript{52}. In what follows, I am going to discuss why this problem presents such a serious challenge. In the

\textsuperscript{52} There is extensive discussion of this problem in Pereboom (2001). Also see Shabo (2012).
next chapter, I will explain how this challenge can be met by non-volitionists. I will come back to the less pressing challenge in the final chapter.

**The Problem of Brain Manipulation**

Consider the following two cases from Derk Pereboom.53

*Direct Manipulation:* Plum is directly controlled by a neuroscientist with advanced technology. The technology allows the neuroscientist to directly produce every mental state in Plum's brain from moment to moment. The neuroscientist likes to manipulate Plum in a way that his brain would work exactly like how an ordinary human brain does: the brain goes through a reasoning process before every behavior, the reasoning process will result in different choices when the reasons are different, the brain has first-order desires as well as second-order desires, which first-order desires always conform to ... The neuroscientist manipulates Plum to undergo a process of reasoning by which his egoistic desire to kill Bill is brought about and carried out.

*Indirect Manipulation:* the neuroscientist has no direct control over Plum this time; however he has managed to program him in such a way that Plum is often but not exclusively rationally egoistic. The egoistic reasons do not always determine Plum’s choice – he has the general ability to regulate his behavior by moral reasons, but these egoistic reasons can become very powerful in some circumstances. Plum is now in such a circumstance in which he finds himself forced by the program to undergo a reasons-responsive process that leads to his killing Bill.

---

53 See Pereboom (2001), pp. 112-5. Some details are changed.
Most people would agree that Plum is not morally responsible for the killing in these two cases: whatever mechanism that is there making him do the killing is not his own. However, Plum seems to meet all the non-volitionist conditions for ownership that we mentioned earlier.

First, consider the endorsement account, which says that one owns a behavior just in case that behavior is endorsed by one's deep self. A sophisticated manipulator can manipulate Plum’s brain in a way that Plum indeed has some deep, endurable egoistic desires, and these egoistic desires are why he forms the intention to kill Bill. What about the attributivist account? Again, a sophisticated manipulator can manipulate Plum in a way that he forms evaluative judgments or attitudes, and his act of killing is expressive of those judgments or attitudes. Likewise, the neuroscientist can also manipulate Plum's brain in a way that Plum is able to form certain judgments, attitudes or dispositions, better yet different judgments, attitudes or dispositions in different circumstances, and these judgments, attitudes or dispositions in fact cause Plum to kill Bill. As we can see, Plum can meet all the conditions in these different accounts of ownership, yet, intuitively, we do not think Plum should be morally responsible in these cases. Thus, an account of moral responsibility based on any of these accounts of ownership fails to explain our intuition about Direct Manipulation and Indirect Manipulation.

However, the problem goes beyond mere inadequacy: these cases show an internal tension in all compatibilist accounts of moral responsibility. On the one hand, compatibilists insist that one can be morally responsible for behaviors that are predetermined by ordinary causes, such as environmental influence and personal disposition. On the other hand, compatibilists deny that one can be morally responsible
for behaviors that are predetermined by brain manipulation. But what distinguishes ordinary determination from brain manipulation such that moral responsibility is established in one but undermined in the other? After all, the behavior is equally causally determined by external factors in both cases. Without an adequate explanation of this dichotomy, a compatibilist account risks inconsistency.

In light of this internal tension in compatibilist accounts, incompatibilists conclude that we need to go back to an account of ownership that builds on a more robust kind of control in order to solve the Problem of Brain Manipulation.\textsuperscript{54} For example, \textit{source incompatibilism}\textsuperscript{55} holds that moral responsibility requires that the agent be the \textit{ultimate source} of the behavior – that is, among other things, the agent's decision to perform the behavior is an event that is not causally determined. Another incompatibilist view, the agent-causal view, holds that in order to be morally responsible for a behavior, that behavior must be agent-caused by the agent. In the two cases mentioned above, Plum is not the ultimate source or the agent-cause of his behavior, and this is why he is not morally responsible for his killing. When it comes to ordinary determination, incompatibilists, unlike compatibilists, have no problem saying that we are not morally responsible if our behavior is actually causally determined – but they deny that our behaviors are, in general, causally determined. Thus, incompatibilism like these do not suffer the challenge from the Problem of Brain Manipulation.

\textsuperscript{54} See, for example, Pereboom (2001), Chapter 4.

It seems that this challenge from brain manipulation leaves us with only two options. First, we give up the ownership-based concept of moral responsibility. That is, we admit that we never deserve any reactive attitudes in the sense that they are appropriate because we own the behavior in question – our behavior does not "calls for" blame or praise on us. Then, we find other grounds on the basis of which cases involving manipulation and cases involving ordinary determination can be distinguished, such as a consequentialist ground (e.g. how efficient is blame or praise as a regulatory measure in each case). Second, if we want to keep the ownership-based moral responsibility, we have to go back to a stronger notion of ownership, one that builds on a more robust control such as the power to do otherwise.

Neither option seems attractive to a compatibilist who would like to preserve a commonsensical account of moral responsibility. I believe there is a third option for such a compatibilist: the option of constructing an account of ownership that is robust enough to distinguish manipulation from ordinary determination, yet moderate enough to not invoke any incompatibilist power to do otherwise. I will explore this option in the next chapter.

But before I explore this third option, I would like to consider another potential move that compatibilists can make: that is, to take the first option. Could compatibilists simply jettison the ownership-based concept of moral responsibility, yet have a reasonable account of moral responsible that avoids the problem of brain manipulation? In the rest of this chapter, I am going to discuss one such proposal and explain why such a proposal must ultimately fail.
An Unsuccessful Solution

Seth Shabo (2012) proposes a compatibilist solution to the Problem of Brain Manipulation by appealing to a principle of fairness. First, Shabo acknowledges that the Problem of Brain Manipulation presents a challenge precisely because compatibilists presuppose a desert-based account of blame. He understands a desert-based account of blame as an account which claims that we are morally responsible for our actions because we are "morally at fault for what we do in such a way that we deserve to be blamed for it" and our act "calls for" or "demands" such reactive attitudes. And I take this desert-based account of blame to be something very close to the ownership-based notion of moral responsibility that we have been talking about. Shabo suggests that if we can build an account of blame that is not desert-based, then we may have a way out of the Problem of Brain Manipulation.

However, Shabo also points out that this non-desert-based account of blame should not be anything like a consequentialist account of blame, an account in which blame is taken to be merely a method to promote good outcome in the future. Rather, this non-desert-based account should still be an account of “unattenuated blame”. By "unattenuated blame", Shabo means reactive attitudes such as resentment and indignation, which have "characteristically retributive sentiments" and express "hostility" or a

56 Ibid., pp. 159-60.

57 At one place, Shabo says that "If an action 'calls for' blame in this sense, it is pro tanto inappropriate (because unfair) that the agent be spared condemnation, where this is a primitive fact about her 'ownership' of that action", which suggests that he takes a desert-based notion and an ownership-based notion to be on a par with each other. That being said, the main point of my discussion in this section is to show that any theory of moral responsibility that intends to adequately explain unattenuated blame needs some notion of ownership. Thus, even if Shabo may not reject the ownership-based concept of moral responsibility completely, this discussion can still serve as an illustration of why we cannot bypass a relevant notion of ownership.
"withdrawal of goodwill towards their target". Unattenuated blame should be distinguished from attenuated blame such as objective attitudes (e.g. upset and joy) or mere "character assessment" (e.g. "This is an evil trait" and "That is a wrongdoing"), which are the kind of blame that a consequentialist account is mainly concerned with.

Shabo suggests that compatibilists can isolate justification for unattenuated blame from considerations of desert, by resorting to a fairness-based justification – people can be morally responsible for their behaviors insofar as "it isn't unfair to blame people for what they are deterministically caused to do". In particular, Shabo appeals to a principle of fairness concerning cynical appeals, which he believes can help us answer the challenge from brain manipulation.

To start with, Shabo makes the observation that sometimes a moral demand can be denied when we have reason to believe that it is cynically motivated. Suppose that a wife finds out that her husband is involved in an extramarital affair with another woman, and she asks for a divorce. If the husband then demands the wife to honor her oath of marriage by appealing to the principle of promise-keeping, the wife has legitimate reason to reject his appeal because it is "cynically motivated". The wife is justified in rejecting such a moral demand because the husband's appeal to that moral demand is "manifestly insincere", and it seems nothing more than an attempt to exploit her moral commitment (Shabo, 173). The very fact that the husband fails to honor his own commitment to promise-keeping, yet at the same time demands others to honor their same commitments just to his own benefit is, according to Shabo, a clear indication that he does not have "a

---

59 Ibid., p. 161.
genuine commitment to the relevant moral value". Consequently, it is not unfair to reject such a moral demand in that particular circumstance.

Shabo then extends this principle to moral responsibility. It is quite common that we can ask to be exempted from blame if we have genuine misunderstanding or we are temporally delusional when acting. However, a demand for exemption is not always accepted. Suppose that John made a very malicious and false remark about Smith, and when he was blamed for making it, John quickly responded, "I didn't mean it. I just happened to have a difficult time. Shall we move on?" According to Shabo, the fact that John is too quick to ask to be let off the hook suggests that he does not take his transgression seriously enough.60 In other words, John's demand for exemption compounds "a presumption of disregard by exhibiting [his] evident cynicism towards our moral standards" – John merely exploits our moral commitment to let himself off the hook. Consequently, just as it is fair for us to reject the husband's cynical appeal to promise-keeping, it is fair to reject John's request for exemption.

It is incumbent, Shabo continues, on the person who makes a demand for exemption to prove that the demand does not compound a presumption of disregard. When no such proof can be offered, the agent has no legitimate complaint against others if they blame her for what she did.

The consideration from fair blame, according to Shabo, can help us answer the Problem of Brain Manipulation. When one is manipulated into murdering another person,

---
60 In his response to an objection that this example presupposes a desert-based notion of blame – what John's reason for exemption does is simply to show that he does not deserve the blame, Shabo claims that there could be desert-independent grounds for believing John should be blamed. For example, we might think he failed to give due consideration to how Smith might feel. However, an immediate worry is that this kind of blame sounds much like an attenuated blame – a mere moral assessment of John's faulty attitude. So, there is a serious question as to whether the focus here is indeed "unattenuated blame".
it is hard to see that "her conduct and attitude on this particular occasion are symptomatic of the abiding deficiencies in her moral commitment". (Shabo, 181) Thus, were she to make a demand for exemption, her demand would not compound a presumption of disregard. By contrast, suppose that one murders another person simply because she has the egocentric character, which is a result of years of bad environmental influence and problematic character building. In this ordinary-causation case, "her conduct and attitude are straightforward manifestation of her considered refusal to govern her behavior by moral standards" (Shabo, 181). Consequently, were she to make a demand for exemption, it would not be unfair for us to reject this cynically motivated appeal. Shabo concludes that there is no need to turn to any desert-based notion of moral responsibility: the idea of fair blame gives us good reason to treat these two cases differently.

An immediate worry is this. Why should an incompatibilist agree that one's conduct and attitude are not "symptomatic of" a disregard to moral commitments when she is under manipulation whereas they are "symptomatic of" such a disregard when she is caused in the ordinary way? After all, a sophisticated manipulator can make his subject's mental states be exactly the same as those of an ordinary agent – they can both have a disregard of their moral commitment, that disregarding attitude causes them to form the intention to murder, and so on. It is really hard to see how one’s behavior manifests a disregard in one case but not in the other when an actual disregard can be present in both cases, unless Shabo has already presupposed some kind of "ownership" – that is, when one is under manipulation, her behavior does not manifest a disregard that can be properly said her own; whereas in the ordinary determination case, her behavior manifests a disregard that can be properly said her own. But then, Shabo clearly owes us
an account of ownership. He has to explain what role this notion of ownership plays in his compatibilist account of blame and why it does not turn his account of blame into a desert-based account.

Not only does Shabo's solution to the Problem of Brain Manipulation require an account of ownership, any account that aims to explain unattenuated blame must have some kind of notion of ownership. As Shabo admits, unattenuated blames are reactive attitudes loaded with retributive sentiments, such as resentment and indignation. These reactive attitudes have as their target the agent, not just the behavior or attitude – we resent people, not just their behavior or attitude. By contrast, attenuated blames are blames that are either mere objective attitudes or mere moral assessment – e.g. "This is a really bad attitude" or "This act is clearly undesirable". If Shabo is right that the key for the fairness of blame is whether the conduct or attitude manifests some disregard to moral commitments, then the immediate target of the blame should be the *conduct or attitude* – after all, the disregard is manifested in the conduct or attitude. So, why should the agent become the target of the blame? The answer must be that the conduct or attitude is *the agent's own*. So, without a proper account of ownership that connects the behavior or attitude to the agent, we can, at best, have an account of blame that has the conduct or attitude as its target – in other words, an account of attenuated blame.

Therefore, in order to account for unattenuated blame, we need some kind of ownership that is able to connect a behavior or attitude to the agent. And since any moderately ambitious account of moral responsibility aims to account for unattenuated blame or praise, it is fair to conclude that an ownership-based concept of moral responsibility is necessary. This is why we cannot take the first option.
What is left for us is the third option: to construct an account of ownership that is strong enough to enable a relevant distinction between manipulation and ordinary determination yet, at the same time, moderate enough to not invoke some incompatibilist power to do otherwise.
CHAPTER FOUR

A Compatibilist Account of Ownership

In the last chapter, we have seen why we need a proper notion of ownership in order to account for "unattenuated" blame or praise, that is, reactive attitudes such as resentment or gratitude. We have also looked at some compatibilist accounts of ownership: the endorsement account, the attributivist account and the causal account, and see why they all fail to meet the challenge from the Problem of Brain Manipulation.

What is noteworthy about their failures is that they all fail because the notion of ownership that they focus on is only the ownership of behaviors – we own a behavior when that behavior is endorsed by our second-order desires, expressive of some evaluative judgments, attitudes or commitments, or caused by them. These accounts fail to explain further how we come to own those mental qualities. Those mental qualities – second-order desires, evaluative judgments or attitudes – can all be imposed upon us from the outside, which is exactly what the Problem of Brain Manipulation shows. Therefore, a proper notion of ownership requires not just an account of how a certain behavior can be called our own, but also an account of how the mental qualities by virtue of which we own the behavior can also be called our own. To answer this challenge, we need a deeper notion of ownership.

Can we have a coherent and adequate compatibilist account of ownership of desires, attitudes or judgments, an account that does not appeal to any incompatibilist power? I believe the answer is yes. In what follows, I would like to propose, based on
what I call the *normative Strawsonian framework*, an account of ownership (more accurately, a condition for ownership) that is able to answer the Problem of Brain Manipulation. Let me start with the normative Strawsonian framework.

**The Normative Strawsonian Framework**

According to the Strawsonian concept of moral responsibility, being morally responsible is just being an appropriate target of certain reactive attitudes, such as resentment, gratitude, indignation and appraisal. Reactive attitudes are an integrative part of our interpersonal relationships, and thus, for Strawson, they require and permit no *external* rational justification. The second part of his theory, as we have seen, has received much criticism. The social practice of condemning homosexual behaviors in some societies, for example, does not make homosexuals truly blameworthy. A solution to this problem, I suggested, is to look for some rational constraints *internal* to the relevant interpersonal relationship, constraints that will illegitimate outrageous reactive attitudes like homophobia.

But how can we do this? According to philosophers such as R. Jay Wallace, George Sher and James Lenman, the answer is to be found in the relation between reactive attitudes and moral fairness.\(^6^1\) In particular, the solution lies in the recognition of the unique normative nature of the kind of interpersonal relationships that reactive attitudes like resentment and gratitude are connected to.

Recall Wallace's normative expectation theory of reactive attitudes. According to Wallace, reactive attitudes like resentment and indignation are natural expressions of our frustration when an agent fails to meet a certain normative expectation that we justifiably

---

\(^6^1\) See Wallace (1996), Sher (2006b), and Lenman (2006).
have on him, and reactive attitudes like appraisal and gratitude are natural expression of our appreciation when the agent fulfills or transcends a certain normative expectation that we have on him. For example, my friend’s resentment towards me as a result of my forgetting to attend his wedding is a natural expression of his frustration – he has good reason, based on norms of friendship, to expect me to attend his wedding. Normative expectations are expectations that one ought to do such-and-such, and they are expectations that we have on members of the moral community as a result of our accepting the moral or other social norms of that community. For instance, once we accept the norm that breaking a promise is wrong, it follows that we would have the expectation that people not break their promise.

Thus, being in this norm-complying interpersonal relationship means that we would naturally feel, for instance, resentment or indignation when someone fails to meet his moral obligation, which we are justified to expect him to fulfill. Once we see this relation between reactive attitudes and normative expectations, we can see that the appropriateness of reactive attitudes is subject to, at least, two internal rational constraints. First, the normative expectation must be based on a justifiable norm. Thus, homophobia is not appropriate because no just moral norm would prohibit homosexuals from having a loving relationship. Second, the individual that is subject to that normative expectation must be a qualified member of the moral community. For Wallace, this

---

62 See Wallace (1996), pp. 77-78. Wallace restricts his discussion of reactive attitudes to a number of negative moral sentiments: resentment, indignation and guilt. However, I think his general theory can easily be extended to positive reactive attitudes as well.

63 Wallace focuses on moral obligations only. However, as I have suggested in Chapter One, I think the relevant norms should be understood more broadly as to include any norm that concerns interpersonal relationships, such as prudential norms or even epistemic norms. This broader interpretation seems to be the suggestion by T. M. Scanlon in his recent book, *Moral Dimensions.*
qualification requires that the individual has the general rational capacities to understand and comply with moral reasons, which he calls *the powers of reflective self-control*. It would be unfair to expect, say, animals or psychopaths, who in general lack the powers of reflective self-control, to comply with norms that they do not understand. Ordinary human beings, on the other hand, have these general rational capacities and are thus subject to normative expectations.

There are three things that I want to emphasize about Wallace’s normative Strawsonian project. First, for Wallace, reflective self-control should be understood as a set of general capacities, capacities to understand and comply with normative reasons in general, not the local capacity to understand and comply with a particular norm on a particular occasion. The difference between a general capacity and a local capacity can be illustrated by this simple example: Reggie Miller has the general capacity to shoot a three pointer, however he may not have the local capacity to do this on a particular occasion, if, say, he is distracted by the noise in the audience. For Wallace, only general capacities to understand and comply with normative reasons are required for being a morally responsible agent. Some may ask "In cases of coercion or genuine misunderstanding, we are usually exempted from moral responsibility – does this not show that the local capacity to understand and comply with a certain norm is also a condition for moral responsibility?" Of course, if one admits that the local capacity is also necessary for moral responsibility, then he will soon be led to admit that we are not morally responsible when we lack the local capacity, in other words, we are not morally responsible unless we could have done otherwise on that particular occasion – a phrase sounds all too

---

familiar to an incompatibilist's ear. Thus, this is a route that compatibilists like Wallace must resist.

But how does Wallace explain cases like coercion and genuine misunderstanding? The answer, according to Wallace, is this: we are exempted from moral responsibility in cases of coercion or genuine misunderstanding not because the local capacity to comply with moral reasons is a precondition for moral responsibility, but because we did not violate our moral obligations in those cases. When you turn on a switch without knowing that it will set off a bomb, your behavior simply does not manifest any ill will or attitude that is in violation of a moral norm. So, it would be unfair to blame you if you have done nothing wrong. Likewise, when a bank robber threatens to kill you if you do not drive him away from the scene, you will not be blameworthy for following his order. This is because you are under no moral obligation to risk your life to stop the robber. We do not need to appeal to any local capacity condition to explain why you are exempted from moral responsibility in these cases.

This type of answer brings up the second thing I want to highlight in Wallace’s theory. The reason why we can say that the agents do not violate any normative expectation in these two cases is that the attitude or judgment behind their behaviors does not show disrespect to their moral obligations: either they simply do not understand what they are doing or there are more important moral values for them to preserve. Thus, whether an agent violates a normative expectation or not is to be determined by the attitude or judgment behind the behavior. A behavior is a sign of failure to meet a
normative expectation only in the sense that it manifests a morally flawed attitude or judgment.\textsuperscript{65}

Last, in addition to the general capacities to understand and comply with normative reasons, I think one other condition is also needed for a person to qualify as a morally responsible agent: the proper opportunity to become aware of the norms. Even if one has the requisite general capacities, it would still be unfair to expect him to comply with a norm which he has no opportunity to be acquainted with. Young children, for instance, may indeed have the requisite rational capacities to understand and comply with a norm like "Don't lie". But if they had never been taught about this norm, it would be unreasonable to expect them to comply with it (unless we want to say that knowledge of this norm is innate). Of course, a proper opportunity to become aware of a norm does not entail a genuine acceptance of it: all it requires is that the agent has a proper exposure to the relevant norm to the extent that an ordinary person in that circumstance would come to know it.

To sum up, upon accepting the norms in our community, we will naturally expect every ordinary member of the community, who possesses the general capacities to understand and comply with these norms and has a proper opportunity to become aware of them, to comply with these norms. Reactive attitudes are natural expressions of our frustration or appreciation when people fail to meet, successfully comply with, or transcend our normative expectations on them: reactive attitudes are an integrate part of the norm-complying interpersonal relationship. Whether an agent violates or obeys a certain norm is to be determined by the attitude or judgment that causes his behavior.

\textsuperscript{65} Wallace's emphasis on the agent's choice further proves that he takes the failure to lie in the will of the agent. But, of course, we have seen why such failure need not require a conscious choice.
Thus, reactive attitudes are reactions to, ultimately, the attitudes or judgments manifested in the behaviors. To say that one is morally responsible for a certain behavior is to say that the behavior manifests an attitude or judgment which is subject to normative expectation, and that the agent's failure to meet or excellence in complying with the relevant expectation makes certain reactive attitudes appropriate toward that agent. This is what I call the *normative Strawsonian framework*.

I believe this framework provides the resources to construct an account of ownership that is strong enough to enable a relevant distinction between manipulation and ordinary determination yet, at the same time, moderate enough to not invoke any incompatibilist power to do otherwise. Let us now turn to the account of ownership.

**A Condition for Ownership**

When we say someone is morally responsible, he is always morally responsible *for* some thing – in our current context, some behavior. But when we blame or resent, we always blame or resent *people*, not or not just the behavior. What connects the behavior with the person such that what was aroused by the behavior flows eventually towards the person? We have seen different proposals. Some say that it is the endorsement by our deep, rational self; some say that it is the causal connection between the behavior and our attitudes, desires or judgments. However, our deep self, attitudes, desires and judgments are all vulnerable to manipulation. So, an account of ownership that can properly ground responsibility must go deeper: what we need is an account that explains how those mental qualities – those attitudes, desires and judgments, can be properly called one’s own.
However, instead of starting with a complete account of ownership, I think a better approach is to start with the question, “Why should we think that manipulation, but not ordinary determination, usually undermines the kind of ownership necessary for moral responsibility?” There are two reasons for doing this. First, a precise account of ownership would be very hard to have, considering the difficulty around a precise analysis of such basic concepts as “knowledge” and “personal identity”. Second, we already have some crude and intuitive notion of ownership to start with. For example, the fact that an attitude is formed in one’s head gives us, at least, a prima facie reason to think that the attitude is his own. What is of greater interest to us is what exactly manipulation does to undermine this initial judgment – this is the more relevant issue.

Obviously, for compatibilists, the answer to this question cannot be that the attitude is externally determined in cases of manipulation. Our attitude and judgment is equally externally determined in both manipulation and ordinary determination. Moreover, if external determination is what undermines the relevant ownership and thus responsibility, then one may be led to suspect that ownership requires undetermined-ness or even the power to do otherwise.

Many compatibilists answer this question by claiming that manipulation undermines responsibility because the agent lacks a particular type of rational capacity in cases of manipulation. For example, in the case of Indirect Manipulation, it might be claimed that Plum lacks the ability to rationally modify the programs imposed upon him by the neuroscientist, and this explains why he is not qualified as a responsible agent. However, there are two reasons why this answer does not work. First, to specify which exact part of our mental states needs to be rationally modifiable in order to qualify as a
morally responsible agent would be an impossible task for compatibilists. It certainly cannot be the case that every element of our mentality needs to be modifiable in order to qualify as a morally responsible agent – for otherwise none of us would qualify. But if just some elements need to be modifiable, the question then is which one(s). Notice that the manipulator has millions of ways to program us and manipulate us. So, it is very unlikely any specification of this rational ability could actually block all cases of manipulation. Second, and more importantly, this approach is unable to handle cases like Direct Manipulation. In cases of step-by-step direct manipulation, whatever rational ability one claims to be necessary for responsibility, a sophisticated manipulator can manipulate Plum in a way that he possesses that ability. Therefore, though I believe certain rational capacities are indeed necessary for responsibility, I do not think the appeal to the possession of any particular rational capacity is the right way to draw the distinction between manipulation and ordinary determination for compatibilists.

If mere determinedness or lack of a particular rational ability is not the right way to explain why manipulation, but not ordinary causation, usually undermines responsibility, what would be the reason for compatibilists to believe that only manipulation undermines responsibility? Where can we find this feature that distinguishes manipulation from ordinary determination? The answer, I believe, must lie in the actual causal history of how an attitude or judgment is formed. When we compare a person who forms an attitude or judgment through exercising his rational capacities in an ordinary way with a person who, as a result of external brain manipulation, goes through the same pattern of exercise and forms the same attitude or judgment, we notice an important difference: the attitude or judgment formed in the case of manipulation is
imposed upon that person by an alien will – the very formation of that attitude or judgment in that person’s head is a result of the manipulator’s act of imposing his own attitude or judgment. Obviously, we are repugnant to such imposition of an alien will. So, perhaps the reason why we think manipulation undermines ownership and responsibility is this aversion to \textit{the imposition of an alien will}. If this is correct, then we may have a compatibilist way, based on this idea of imposition of an alien will, to distinguish manipulation from ordinary determination.

But two challenges lie ahead. The first challenge is to specify, in a compatibilist way, what exactly this \textit{imposition of an alien will} means. The second challenge is to provide a compatibilist explanation of why the imposition of an alien will undermines the victim’s moral responsibility. If, in the end, the reason why we think imposition of an alien will undermines responsibility is that we believe moral responsibility requires robust control and we lack such control when an alien will is imposed upon us, then this project leads us nowhere.

Let me start with the first challenge. The best way to articulate an imposition of an alien will is by looking at the details of how the relevant attitude or judgment is caused. If the very formation of an attitude or judgment is causally determined by a process which is, essentially, a process of merely copying the content of some existing attitude(s) or judgment(s), then that attitude- or judgment-formation is obviously a result of an imposition of an alien will. For example, when the manipulator installs an evil judgment, “Bill should be killed”, directly into Plum’s brain to make him act on it, this is a case of an imposition of an alien will: the formation of that judgment is causally determined by the manipulator’s act of installing that judgment, and the process of installation is a
process of merely copying the content of an attitude that already exists in the manipulator’s mind.

In some cases, the attitude or judgment that is formed in Plum’s brain need not be already entertained by the manipulator. For example, the manipulator may install three different pieces of information into Plum’s brain, an attitude that "All bad people should be killed", a belief that "Bill is bad", and a reasoning pattern that "If X is bad and all bad people should be killed, then X should be killed", which, together, lead Plum to form the judgment that Bill should be killed. In this case, the manipulator may never actually entertain the thought “Bill should be killed” in his own mind. But this is also a case of imposition of an alien will (or rather alien wills), since the formation of that judgment, “Bill should be killed”, is causally determined by the process of copying those three constitutive elements from the manipulator’s brain. For this reason, the formed judgment is also a result of imposition of alien will.

But I need to be more careful with the phrase *imposition of an alien will*, which, by invoking the word “alien”, seems to presuppose an account of ownership. Using this notion as part of a condition for ownership may arouse the suspicion of circularity. However, my actual definition of *imposition of an alien will* – a formation of an attitude or judgment is a result of imposition of an alien will if *that formation is causally determined by a process which is, essentially, a process of merely copying the content of some existing attitude(s) or judgment(s)*, contains neither “alien” nor any other notion that presupposes ownership. So, there is no circularity.

Based on this notion of imposition of an alien will, I propose the following condition for the kind of ownership necessary to ground responsibility: An attitude or
judgment can be properly said *one's own* in a way that grounds his or her responsibility
*only if* it is not the case that the formation of that attitude or judgment is causally
determined by a process which is, essentially, a process of merely copying the content of
some existing attitude(s) or judgment(s). Notice that this condition does not rely on any
incompatibilist notion like the power to do otherwise; all it involves is a description of
the causal history of how an attitude or judgment is actually formed.

The next challenge is to explain, in a pure compatibilist way, why imposition of
an alien will usually undermines moral responsibility. I believe we can find such an
answer in the normative Strawsonian framework. On the one hand, according to this
framework, every member of the moral community, who has the requisite capacities to
understand and comply with the norms and has the proper exposure to the norms, is
expected to obey them. Thus, the default expectation is that people ought to behave in
accordance with the norms. When the state of normalcy is disturbed, that is, when a
normative expectation is violated or transcended, we are frustrated or impressed by this
disturbance, and naturally we want find out the source of that disturbance – we want to
figure out what kind of thing would cause such a disturbance to normalcy. Our reactive
attitudes are a natural expression of such frustration or appreciation, and thus our reactive
attitudes are ultimately reactions to the source of the disturbance. On the other hand,
according to this framework, whether a normative expectation is violated or not is
determined ultimately by the attitude or judgment that causes behavior. Thus, when an
apparent disturbance to normalcy happens, our natural reaction is to look for the attitude
or judgment behind it. For example, when a moral evil is committed, we will naturally
look for the motivation behind it, and figure out where this plan of blatant violation of the
norms comes from.\textsuperscript{66} Therefore, when a disturbance to normalcy happens, it is the
original attitude or judgment, the source of the disturbance, that our reactive attitudes
really track. Just as, when we appreciate the beauty of a poem, our homage is paid not to
the person who merely recites it, but to the person who created it – the true producer of
that quality of beauty, our reactive attitudes are reactions to the moral flaw or excellence
in the attitude or judgment from which the disturbance to normalcy originates.

When we see the unique importance of being the source of a disturbance to the
norm-complying interpersonal relationship, we understand why \textit{an imposition of an alien
will} is of such importance to moral responsibility: our resentment, indignation, gratitude,
appraisal and so on are directed towards the source of the disturbance to normalcy; and if
the attitude or judgment in question is merely a copy of some other existing attitude or
judgment, it is then not the source of the disturbance, and thus not the focus of our
reactive attitudes.

So, we have a compatibilist explanation, based on the normative Strawsonian
framework, of why imposition of an alien will would make Plum an inappropriate target
of reactive attitudes even if he forms the relevant attitude or judgment in his head. There
is reason, independent of any incompatibilist intuition, to believe that imposition of an
alien will undermines responsibility. Thus, we have an account of ownership that is
strong enough to distinguish manipulation from ordinary determination and, at the same
time, moderate enough so as to avoid any incompatibilist control.

\textsuperscript{66} In his book, \textit{Acts of Meaning} (1990), psychologist Jerome Bruner makes the observation that
human beings tend to search for reason or meaning whenever others' behaviors violate our
expectation, or whenever they do not adhere to the accepted social norms. Our natural disposition
to search for the cause of a violation of our expectation may be part of the psychological ground
for why our reactive attitudes, which are just a kind of social psychological phenomena, are
naturally directed towards the very source of a manifested ill will.
 Needless to say, my proposal of ownership is not a complete account; it only provides an exclusion clause – why manipulation undermines the relevant ownership necessary for responsibility. There may be other situations in which the relevant ownership and responsibility can be undermined. But my main concern here is to answer the Problem of Brain Manipulation. Thus, though a complete account of the kind of ownership necessary for responsibility would be desirable, I should not attempt that further project here.

Some Objections

Now, let us consider some objections to this proposal. The first objection is this. When Plum learns an evil attitude from someone else, the formation of that attitude seems to be determined by a process of copying an existing attitude. But, intuitively, we would not exempt Plum from his responsibility in this case. How can my account explain this apparent conflict?

I agree that when Plum forms an attitude, say, “Bill should be killed”, by learning it from another person, the content of that attitude is copied from that of an already existing attitude. But I do not think this attitude-forming process itself is determined by a process which is, essentially, a process of merely copying something else. Assuming that Plum is a qualified agent for responsibility, for a person like Plum to learn and accept an attitude like “Bill should be killed”, which is in violation of the norm against killing, is to manifest another implicit attitude about that norm – an attitude of disrespecting that norm or an attitude of not taking it seriously enough. This is because any person with the proper understanding of the norm and the proper ability to comply with it should be able
to recognize the inappropriateness of this attitude and thereby to denounce it – his ability provides him with the opportunity, a “checking point” in the learning process (so to speak), to sift those received beliefs, attitudes or judgments. The failure to block them out, either by giving those inappropriate received beliefs or attitudes a free pass or by consciously accepting them, automatically manifests a flawed attitude towards the norm – if Plum really took the norms seriously enough, as he should, he would have blocked those beliefs or attitudes out at the “checking point”. So I think, in the current case, there is something other than the salient attitude-copying process that is also involved in Plum’s learning: his failure to block out the evil attitude at the "checking point" manifests a flawed attitude that is already in him, and it is not copied from some existing attitude.

But what if Plum learns both the salient attitude, i.e. “Bill should be killed”, and the implicit attitude, i.e. the disrespect to the relevant norm, from someone else? This seems to make the learning process a process of merely copying something else; yet, still we will not exempt Plum from his responsibility if he acts on those attitudes. My response is that I do not think the implicit attitude – the attitude of disrespecting the norms, can really be acquired through learning. The reason for thinking so is that any person who has the requisite knowledge of the norms and the ability to comply with normative reasons would not accept an attitude that goes against the norms, unless he has already had such a disrespecting attitude. The failure to block such an attitude out in an ordinary learning process must manifest a disrespecting attitude that is already in the agent. In short: no disrespect to the norms can be learned unless the agent has already denounced his proper respect to the norms.
This is, of course, not to say that the agent may never fail to recognize inappropriate views or fail to denounce them. The “checking point” can get bypassed if one’s ability to comply with normative reasons is deprived or undermined. For example, if during the process of receiving a belief or attitude from the outside, the agent accidentally suffers from a brain trauma which temporarily deprives him of his ability to comply with normative reasons, then the "checking point" is bypassed. Or one may be brain manipulated and thus loses his ability to comply with his normative reasons. In these cases, the formation of the relevant attitude or judgment is indeed determined by a process which is, essentially a process of merely copying something else. But this is not a problem for my account because in these cases the agent is indeed not morally responsible for his behavior.67

Here is another objection. Consider the following revision to the Plum case. Suppose that Plum’s brain is connected to a machine which sends signals directly to every neuron in Plum’s brain and thus creates its brain states from second to second. Under the control of this machine, Plum forms the judgment that “Bill should be killed” and acts accordingly. In this case, the formation of the attitude is not determined by a process which is a process of simply copying some existing attitudes or judgments – the machine, which has no mind or consciousness, entertains no such attitude or judgment itself. But our intuition is that Plum is not responsible for either the judgment or the behavior. What do I say to a case like this?

First of all, if the mindless machine simply sends out signals in a completely random way, then it is extremely unlikely that Plum’s brain will have the general rational

67 Notice that there is no appeal to the local ability to comply with normative reasons. The reason why one is not morally responsible is simply that his attitude is a mere copy of some existing attitude.
capacities to understand, comply with and respond to different normative reasons. Therefore, if this is the case, Plum will not be a responsible agent at all because of his lack of the general rational capacities.\textsuperscript{68} Second, if the machine operates in accordance a rational plan that was programmed by an intelligent designer, then its operation is indeed a process of copying some existing attitudes or judgments – namely, the plan that the designer had in mind when he was doing the programming. Therefore, it is the designer, not Plum, who is the appropriate target of our reactive attitudes.

But suppose the mindless machine operates based on a rational pattern, but there does not exist any designer: it simply operates in that orderly way to make Plum’s brain possess the general rational capacities. If this is indeed the case, no matter how crazy it sounds, then I insist that Plum is indeed responsible for his attitudes and behaviors. The mere fact that Plum’s brain activities are completely determined by an external machine should not give compatibilists reason to think that Plum is not responsible. Nature can be viewed just as such a machine, sending out signals to our brain and causally determining our brain states from second to second. Since compatibilists have no problem accepting responsibility in cases of ordinary determination by Nature, I see no reason why we should hesitate to hold Plum responsible in the current case. Thus, my response in this case is that Plum does own his attitude in a way that can ground his moral responsibility.

\textsuperscript{68} Likewise, suppose that the machine only does local manipulation to Plum’s brain – that is, Plum’s brain is, by and large, a normally functioning brain, but the machine manages to insert certain attitude or desire into it (e.g. the desire to eat tomatoes). I believe, in this case, we can also appeal to the lack of proper rational capacities to explain why Plum is not responsible for the relevant attitude or desire. If the machine simply randomly inserts such an attitude or desire (without following any reasons-responsive pattern), that attitude or desire is not likely to be reasons-responsive. And thus, compatibilists can account for such cases by appealing to the lack of the relevant rational capacities.
To further support my claim of ownership here, I invite the reader to think about a case of *Brain Separation*. Suppose that a small but essential part of Plum’s brain, a part where all the important decisions are made, is taken out of his skull by scientists and placed in an external experimenting device. Scientists also manage to have the remaining brain grow back to its original volume, but the mental activities of the remaining brain are under direct control of the part that is taken away (through some advanced tele-biochemical control technology). In this case, we would have no problem viewing the external controlling part as *part of Plum’s operating brain*. And we view it as part of Plum’s operating brain not because it has a history of being a part of it – suppose that the part within Plum’s skull and the part outside of it operate independently (e.g. they run on two independent sets of rules of computation and do not causally interact with each other), we would have no problem rejecting the claim that the external part is still part of Plum’s operating brain even if it used to be a part of it. This shows that the real reason why we view the external part as part of Plum’s mind is by virtue of the essential role that it plays in Plum’s mental activities – it contains the set of rules on the basis of which Plum’s brain activities operate, and every brain activity of Plum’s brain directly results from that external system’s computation. Therefore, just because a decision is made outside of one’s skull, it does not follow that the decision is necessarily not that person’s own. If this is right, then given the essential role that the mindless machine plays in Plum’s brain activities, we also have good reason to view that machine as part of Plum’s brain. It might seem very counterintuitive to take a mindless machine as a constitutive part of a rational agency. However, we also need to keep in mind that this mindless machine is the one that operates a brain with competent rational capacities to understand and respond to
reasons. A group of brain cells may not have any awareness or “understanding” of the broader propositional content that our mind is entertaining, but it does not follow that this group of brain cells are not an essential part of what constitutes our rational agency.

Compare this case to a case in which the machine is operated by another person to carry out his will. In this case, the manipulator’s will is what determines Plum’s mental states. The presence of an existing will reverses the ownership ascription that we had earlier: in this case, the attitude or judgment formed in Plum’s head is a result of an imposition of another will – the attitude or judgment is a mere copy or reflection of the will of the manipulator. Thus, the proper owner of that attitude or judgment is the manipulator.

Let us consider a further objection. Suppose that the mindless machine is operated by a person; however, this manipulator is not entertaining any plan, attitude or judgment in his mind either – perhaps, he is simply randomly clicking buttons on the control board of the machine. In this case, whatever attitude or judgment formed in Plum’s brain does not reflect any attitude or judgment in the manipulator’s mind. Would my theory, therefore, say that Plum owns the attitude or judgment formed in his head?

First, I want to point out that the reason why the Problem of Brain Manipulation is such a serious challenge to compatibilism is that no matter what rational ability compatibilists propose as the pre-condition for responsibility, a sophisticated manipulator can manipulate his victim’s brain in a way that meets that pre-condition. But such sophisticated manipulation requires the manipulator to carefully attend to the environment of his object and form proper rational responses in his own mind. Thus, if a manipulator simply clicks buttons on a control board randomly without forming the
relevant attitudes, beliefs or judgments himself, it would be extremely unlikely, if not utterly impossible, that he can manipulate a brain in a way that the brain is, in general, able to understand, comply with and respond to various reasons properly.

However, for the sake of the argument, suppose that very strange things happen and the manipulator, without attending to any relevant reasons and forming any relevant beliefs or judgments in his own mind, manipulates the buttons in a way that miraculously enables Plum to have the general rational capacities. In this case, presumably one would say that I can no longer claim that Plum’s attitudes or judgments are a mere copy of the manipulator’s attitudes or judgments. There are two distinct minds: Plum’s mind with a particular set of attitudes, beliefs, desires, dispositions and judgments; and the manipulator’s mind with a particular set of attitudes, beliefs, desires, dispositions and judgments; and neither necessarily reflects the other. Since reactive attitudes are ultimately reactions to the will or mind (especially the relevant attitudes or judgments), there must be two distinct subjects for reactive attitudes. Consequently, it seems that my theory would say that Plum is an agent responsible for attitudes or judgments formed in his head just as the manipulator is an agent responsible for attitudes or judgments formed in his own head. But, of course, this seems wrong: if anyone should be responsible for the attitudes or judgments formed in Plum’s head, it should be the manipulator.

Let us grant that there are indeed two distinct minds or wills. However, what does the so-called “Plum’s mind” consist of? The mental activities in Plum’s brain, sure; but let us not forget the activities of the manipulator, which play the essential role in generating all the mental activities in Plum’s brain. If we have good reason to take the mindless machine as part of Plum’s mind in an earlier case by virtue of the essential role
it plays in generating Plum’s mental activities, then we should also have good reason to
take the manipulator’s mental activities as part of Plum’s mind given the identical role it
plays. Thus, when we react to “Plum’s mind”, what we have as the target of our reactive
attitudes is not just the body and brain of Plum, but also the body and brain of the
manipulator, who is now also part of this “Plum’s mind”. And given the ultimate role that
the manipulator, as part of “Plum’s mind”, plays in forming the relevant attitude or
judgment, it is reasonable to think that he, as an object, deserves a greater part of our
blame.

It may seem quite counterintuitive to think of the manipulator as part of another
mind. This is partly due to the fact that the manipulator has already had a unique mind of
his own. But, again, we shall keep in mind the very unusual role that the manipulator
plays in creating an individual mentality that has its own character, attitudes, dispositions
and so on. Our ordinary intuition about personal identity is attuned for ordinary
circumstances. Given how extremely unusual the current case is, our ordinary intuition is
perhaps not very reliable.

**Conclusion**

It is a plain fact that our beliefs, attitudes and judgments are constantly shaped by causal
influences from the outside: we read books, watch movies, listen to wise speeches, and
observe how greater moral characters act in difficult times … So, it is no surprise to find
out that we do not really “own” most of our beliefs, attitudes and judgments. But that
being said, there is still a sense to say that we do own those beliefs, attitudes and
judgments – we take them in or shut them out at the “checking point”, the *sifting*
opportunity arising from our ability to understand and comply with normative reasons. The failure or success at the “checking point” explains why we can still be at fault even if we are under causal influence from the outside – they manifest in us a disregard of the moral norms or an attitude of not taking them seriously enough.

When one’s mind is manipulated by something else, however, the “checking point” is no longer available. It is no longer available because, when we look into the actual causal history of the formation of the relevant attitude, belief or judgment, the formation is determined by the external manipulation, as opposed to its being a feature that already exists in that mind. More specifically, in an ordinary learning process, the external causal influence merely provides some information, while whether the mind will take it in or shut it out is to be determined by some cognitive feature that has already existed in the brain (e.g. the disregard to the norms); whereas, in a manipulation process, not only does the external causal influence provide the information, but also it determines whether that information will be taken in or shut out (e.g. the disregard to the norms is also installed in the brain by the manipulator). This is the key difference in the causal history of the formation of an attitude or judgment between manipulation and ordinary determination.

The reason why manipulation undermines the victim’s responsibility, on my theory, is to be explained by the ascription of the proper *subject* for reactive attitudes. When the relevant attitude or judgment is merely a copy of some attitude or judgment that already exists in the manipulator’s mind, the will of the victim merely reflects the will of the manipulator. Since our reactive attitudes are ultimately reactions to the source of the evil or good will, it is the manipulator who is the proper subject for our reactive
attitudes. When the manipulated attitude or judgment does not reflect any existing attitude or judgment (e.g. the formation of the attitude or judgment is determined by a mindless machine or a person who simply causes different brain states to occur in a clueless manner), then I argue that we have good reason to view the manipulating machine or person as part of “the subject” for our reactive attitudes. The reason why people usually think that the victim is not responsible under such manipulation is that they tend to view the victim as a distinct person as the manipulating mechanism. But if all the important psychological features of the victim’s mind actually directly result from the operation of the manipulating mechanism, I think there is good reason to view that mechanism, be it a machine or a person, as part of “the subject” (or the person) which we hold responsible. Of course, people may disagree with me on a correct theory of personal identity – though I believe unity and continuity of computational operation and other psychological features is a reasonable and widely accepted criterion for personal identity. However, the crucial point is that, if the problem of manipulation turns out to be an issue of personal identity, then compatibilists are relieved of the worry that we may need to appeal to any robust incompatibilist control to explain manipulation – for obviously there is no clear connection between robust control and personal identity.

Now, part of my theory builds on the idea that the proper subject for reactive attitudes is the causal source of the disturbance to normative expectation, and thus one may wonder how my account of ownership is different from source incompatibilism. According to source incompatibilism, a person under manipulation is not morally responsible for the manipulated behavior because the person is not the ultimate source of

---

According to a popular theory of personal identity, the identification of a person is determined by such psychological features such continuity and unity of memory, intention, beliefs, desires and characters. See, for example, Derek Parfit (1984).
that behavior. According to my account, the reason why a person like Plum is not responsible for his behavior is also that he is not the. In one sense, there is indeed an agreement between my account and source incompatibilism: we both think that being the causal source of the behavior is important for moral responsibility. However, there are also crucial disagreements.

First, we disagree on what exactly it means to be the source. For source incompatibilist, being the source means being the starting point of the causal chain leading to the behavior, and the start of that causal chain is an event whose occurrence is an undetermined event. By contrast, on my account, being the source means simply that the formation of the attitude or judgment is not determined by a process which is essentially a process of merely copying some existing attitude or judgment. The event of forming that attitude or judgment could indeed be a causally determined event.

Second, we disagree on why being the source of the behavior is important for moral responsibility. For source incompatibilist, being the ultimate source of a behavior is *metaphysically* necessary for moral responsibility – it is a metaphysical feature of this world that there is such a relation between being the ultimate source of a behavior and being morally responsible for that behavior. And our knowledge of this unique metaphysical feature is solely based on intuition. By contrast, on my account, being the source is relevant for moral responsibility because of the socio-psychological nature of our reactive attitudes – it is a socio-psychological feature of our human nature and our interpersonal relationship that being the source of the relevant attitude or judgment matters for moral responsibility. And our knowledge of this socio-psychological feature is based on our observation rather than mere intuition. As I will explain later in Chapter
Five, there is also a rational justification for why the relevant interpersonal relationship should work the way it does.
CHAPTER FIVE

A Non-Volitionist Account of Responsibility

We started with the Strawsonian concept of moral responsibility, which takes being morally responsible as being the appropriate subject for reactive attitudes. Then, we discussed an important development of the Strawsonian theory: the normative Strawsonian framework suggested by people like Wallace and Sher, according to which reactive attitudes are natural expression of our frustration or appreciation when our norm-complying interpersonal relationships are disturbed by some behavior, and these reactive attitudes are subject to certain constraints internal to the norm-complying interpersonal relationships. In the last chapter, we saw how this normative framework provided the resources for a compatibilist account of ownership and how that account is able to answer the Problem of Brain Manipulation. Now, it is time to put together a complete non-volitionist account of moral responsibility.

A Non-Volitionist Account of Moral Responsibility

There are two general conditions for moral responsibility. The first condition concerns the general qualification for being a morally responsible agent. In particular, there are two of them. First, the individual must have the general capacities to understand and comply with normative reasons. Second, the individual must have some proper opportunity to be acquainted with the relevant norms – an opportunity such that an
individual with the requisite capacities would become to know the relevant norms and understand them.

The second condition concerns the agent’s relation to the target behavior. The individual must have proper “control” over the behavior and thus own the behavior in the proper way. To properly own a behavior one must meet two requirements: (1) the behavior is directly caused by some, explicit or implicit, attitudes or judgments in the person’s head, and (2) he owns the attitudes or judgments in a proper way. A person is said to own an attitude or judgment only if the formation of that attitude or judgment is not determined by a process which is, essentially, a process of copying some existing attitude(s) or judgment(s).

When one meets these two conditions, he becomes an appropriate subject for reactive attitudes. But this is only half the story. So far, we have only settled the general appropriateness issue – that is, how the agent becomes an appropriate subject for any kind of reactive attitudes as a result of performing the behavior in question. We have not yet settled the specificity issue – that is, which particular reactive attitude is appropriate in the given circumstance. To settle the specificity issue, we have to look at the specific norms that govern the interpersonal relationships in the community.

In particular, we need to look at two kinds of norms. One kind of norms determine what normative expectation is appropriate to have. Norms like “Never kill an innocent human being” and “Be truthful to your friends” tell us what to expect from members of the community. And these norms must themselves be morally justified so that every rational member of the community would consent to them. Thus, there is an internal constraint on the appropriateness of reactive attitudes: the normative expectation
must be based on a justified norm in the first place. This explains why it is inappropriate to blame, say, homosexuals for having homosexual relationship, because no norms that support homophobia can be morally justified. Consequently, those outrageous reactive attitudes that besiege the original Strawson’s account are properly illegitimated in this normative account.

The second kind of norms determine which particular reactive attitude is appropriate when a given normative expectation is frustrated, fulfilled or transcended. Norms like “Murder is a much greater evil than theft” and “It is honorable thing to sacrifice your own wellbeing to help others in dire need” tell us that a crime of murder, other things being equal, should incur a more severe blame than a crime of shoplifting, and a sacrifice of one’s own good for others’ wellbeing calls for our appraisal and respect. Norms like these specify which particular type of reactive attitude is appropriate in a given situation and how strong it should be.

In determining which reactive attitude is appropriate to have, we first look into the attitudes or judgments which cause the behavior in question and which are the agent's own, and examine these attitudes or judgments to find out if they violate, satisfy or transcend any of our normative expectations; second, if some normative expectation is violated, satisfied or transcended, then based on the particular normative expectation that is violated, satisfied or transcended, we determine which particular reactive attitude is normatively appropriate.

This is my non-volitionist account of moral responsibility. This account has several theoretical advantages over its alternatives. First, my account pays proper respect to both the practical aspect and the metaphysical aspect of the commonsense notion of
moral responsibility: regarding someone as morally responsible is a practice as well as a belief. Thus, my account captures the rightful insight in both the ownership-based concept and the practically-oriented concept. Second, this account provides a better theory of types of reactive attitudes and degrees of blame or praise, which are not unfamiliar features of the commonsense notion of responsibility, yet features by and large ignored by most other accounts of moral responsibility. Last, but not the least, this account is able to answer the Problem of Brain Manipulation in a purely compatibilist way.

On this account, moral responsibility does require some kind of control, but this ownership-generating control need not be anything as robust as any incompatibilist power to do otherwise – it means simply (1) having the rational capacities to understand and comply with normative reasons and (2) being the actual causal source of a moral flaw or excellence.

However, there are people who would insist that a robust kind of control is needed. Their insistence seems to be grounded on a principle that they view as very intuitive, the principle that an individual is not morally responsible for a behavior unless he could have avoided it. This avoidability principle, they believe, entails that the agent must have a more robust control over his behavior, such as the power to do otherwise. Does moral responsibility necessarily require such a robust control? This brings us back to the challenge we mentioned earlier in Chapter Three.

**The Avoidability Principle**
The avoidability principle is the principle that we are morally responsible for something only if we could have avoided it. A more familiar face of this principle is the so-called Principle of Alternative Possibilities: we are not morally responsible for an act unless we could have done otherwise.\textsuperscript{70} If we accept this principle, then we need a much stronger kind of control for moral responsibility, the power or ability to do otherwise.

Libertarians find this principle intuitively right, yet others disagree. Serious doubt arises from a series of cases called Frankfurt-style counterexamples. These cases usually involve a scenario in which an agent willingly performs an act, but there exists, at the same time, a counterfactual intervention that would prevent her from not performing that act should she decide not to do so.\textsuperscript{71} According to the opponents of the avoidability principle, there is a strong intuition that the agent is morally responsible for her act, despite the fact that she could not have avoided it. If this intuition is right, according to them, the principle must be false. Two types of response have been suggested. According to one type of response, the Frankfurt-style counterexamples presuppose something incompatible with the Principle of Alternative Possibilities, namely, determinism, and thus simply beg the question against that principle.\textsuperscript{72} According to the second type of

\textsuperscript{70} Frankfurt (1962), p. 829.

\textsuperscript{71} Ibid.

response, the counterfactual intervention does not deprive the agent of all the alternative possibilities, and thus one’s being moral responsible in those cases can still be explained by virtue of the existence of the alternative possibilities that remain. Various rebuttals to these responses have been generated in turn, and I shall not try to evaluate their force here.

It is fair to say that it remains a live controversy whether counterexamples of this kind succeed in refuting the avoidability principle. I, along with many others, believe what these examples really do is to show that we have two competing intuitions with regard to moral responsibility, and both are strong. So, instead of using one intuition as evidence to *disprove* the other equally strong intuition, the best way to counter that principle is to *dismiss* it. The proper question that we should ask is: “What is the theoretical ground for this principle?” Why do we need such a principle for a theory of moral responsibility? If we can show that, apart from its intuitiveness, there is no good theoretical ground for the avoidability principle, and, at the same time, there is a solid theoretical ground for its competitor, then we have good reason to dismiss this avoidability principle as unnecessary.

What kind of theoretical reasons, apart from its intuitiveness, do we have to believe that the avoidability principle is needed for a theory of moral responsibility? I can think of only two: first, it may be that the avoidability principle provides the best justification for our actual responsibility-involving practices; second, it may be that this

---

73 For responses of the second kind, see Rowe (1991), *Thomas Reid on Freedom and Morality*, and Van Inwagen (1983), *An Essay on Free Will*. According to this response even if the agent does not have the stronger form of alternative possibilities, namely, the ability to make a different decision, she still has a weaker form of alternative possibilities, namely, the ability to decide whether she is the real cause of her decision. One response given on behalf of Frankfurt-style challenge is by John Fischer, who argues that this weaker alternative possibility is merely a “flicker of freedom”, which is too weak to ground moral responsibility for an act.
principle is entailed by the very concept of moral responsibility. Let us consider them in turn.

In practice, we usually exempt people from their responsibility if they are directly manipulated, mentally ill or under wrongful coercion. It might be argued that the avoidability principle best explains why these practices are justified. And if this is true, then there is good theoretical reason why we need this principle. However, I think not only do we not need the avoidability principle to explain these practices, but also the principle fails to provide an adequate explanation in most cases.

To start with, we have seen, in the last chapter, an alternative way to explain why brain manipulation usually cancels moral responsibility: there is no need to appeal to the avoidability principle. Furthermore, appealing to that principle does not provide a good explanation: a sophisticated manipulator could allow certain behaviors to be avoidable. For example, he could always throw a die to decide which behavior "his puppet" should perform. So, it is unclear that appealing to that principle even provides an intuitive explanation to cases involving manipulation.

What about mental disorder? Certain mental disorders will psychologically compel us to perform behaviors we would otherwise not want to perform. For example, a kleptomaniac is compelled to steal even if he does not “want” to. It seems that the avoidability principle gives a fairly straightforward explanation in this kind of cases. However, an alternative explanation that does not require alternative possibilities is also available. On my account, one of the qualifications for being a morally responsibility agent is the possession of rational capacities to understand and comply with normative reasons. Patients who are compelled by their mental illness usually lack the capacities to
follow their moral reasons, and this explains why they are usually not morally responsible.74

Some mental disorders do not actually compel us to do anything, but simply distort our understanding of what is going on. For example, an autistic patient cannot read other people’s feeling through their behaviors, and thus he often has no clue when his behavior hurts others. My account has a nice explanation of why patients like this are not morally responsible for their behavior: the behavior does not manifest a moral flaw. The patient simply is not aware that he is hurting someone.

On the other hand, the appearance that the avoidability principle offers a neat explanation does not stand a closer scrutiny. Consider the following case. Suppose a patient suffers from two different kinds of compulsion simultaneously, one compelling him to do A and the other compelling him to do B. The patient may actually be able to avoid either A or B, for it may be an undetermined event which compulsion will eventually prevail. But it does not follow that he is therefore morally responsible for his behavior. The avoidability principle does not seem to capture what is really missing in such patients.

The same thing can be said about cases of wrongful coercion. The practice that we do not regard a person as blameworthy for his act when he is wrongfully coerced can be explained by the fact that he violates no moral obligation. For instance, if you choose to help a bank robber get away from the scene when he points a gun at your head, you are not blameworthy for doing it. You are not blameworthy not because you could not have

---

74 Likewise, to explain why a person who was suffering from a traumatic brain injury or was under the influence of a powerful poison or drug is often not responsible, we do not need to appeal to any local capacity to X or not X. My explanation is simply that the person loses the general rational capacities to understand and comply with normative reasons in those circumstances.
avoided it – some people may actually be willing to sacrifice their life, but because there is no moral obligation that requires an ordinary person to sacrifice his life to stop a bank robbery. There is no such moral obligation because it would be *unreasonable* to put the value of a life below the value of stopping a robbery. Thus, we have an adequate way to account for wrongful coercion without any appeal to the avoidability principle.

Not only is the avoidability principle unnecessary for explaining cases of wrongful coercion, any appeal to that principle goes precisely in the wrong direction. Consider this revised bank-robbery case: the bank robber points a gun at your head, and instead of asking you to drive him away, he asks you to kill an innocent bystander this time. You might be as incapable of resisting his command in this case as you are in the original case, but you would be blameworthy for following his command. This shows that the explanation cannot lie in the fact that you are unable to avoid the behavior; rather, the explanation lies in the fact that you have a moral obligation not to kill an innocent human being even if your own life is at stake. Thus, your compliance would manifest a moral flaw in this revised case.

Therefore, not only do we not need the avoidability principle to explain our practices in cases of manipulation, mental illness and wrongful coercion, we have good reason not to appeal to the principle. My account of moral responsibility, which relies on a weak kind of control, provides a better justification for our practices than an account that relies on such robust control as the power to do otherwise.

What about the second possibility? Might the avoidability principle be needed because the concept of moral responsibility implies it? Here is how one might argue for such a claim. The normative Strawsonian framework says that being morally responsible
means being appropriately subject to some normative expectation. A normative expectation is an expectation that one ought to do such-and-such or have such-and-such an attitude. But if one is determined to act against the norm, then he should not be expected to comply with the norm, for that would violate the principle, "Ought implies can". So it follows that, for a normative expectation to be reasonable, the person must be able to avoid disobedience – that is, he must have the power to do otherwise.

But the principle, "Ought implies can", is ambiguous. Recall a distinction we mentioned earlier. There are two types of capacities: the general capacity to X and the local capacity to X – one’s capacity to X in a particular circumstance. The advocates of the avoidability principle seem to assume that the "can" here should be understood as the local capacity. Thus, for them, “Ought implies can” means that if one ought to X then he must have the local capacity to X in the given circumstance. But this interpretation is questionable. For example, we all believe that drivers ought to pay close attention to the road when they are driving. But we also know that, as a matter of fact, it is inevitable that our mind will wander off from the road now and then. So, if “can” means the local capacity, then the principle "Ought implies can" is simply false in this case. One may respond that perhaps the correct norm should be that we ought to pay close attention to the road most of the time, but not all the time. However, such a response is immediately repudiated by our purchase of the complaint, "You should have paid closer attention", when an accident really happened. Consider another example. It is said that teachers ought to make their teaching interesting and engaging; but, as a matter of fact, many teachers are unable to make every class interesting and engaging. Should "can" indeed
mean the local capacity, the principle, "Ought implies can", would be rather ill-grounded, if not blatantly false.

Richard Feldman suggests that the kind of ought-claims in the above cases should be understood as role-oughts: they are "oughts that result from one's playing a certain role or having a certain position."\(^{75}\) And such ought-claims are justified as long as the agent who takes the relevant role has in general the capacity to fulfill the requirements of that role. It does not require the agent to have the relevant local capacity on every occasion. Likewise, for the normative Strawsonian framework, the ought-claims in our normative expectations are role-oughts: they are oughts that result from being a participant in the norm-complying interpersonal relationships.\(^{76}\) In the determination of role-oughts, it is the general capacity, not the local capacity, that matters. As Wallace points out, insofar as an individual has the general capacities to understand and comply with norms, it is fair to expect him – a participant in the norm-complying interpersonal relationships – to comply with the norms. Clearly, this general capacity to understand and comply with norms does not imply the power to do otherwise.

Therefore, the principle “Ought implies can” does not entail that normative expectation requires the power to do otherwise: we cannot derive the avoidability principle from the normative Strawsonian framework. Neither the conceptual framework nor an adequate explanation of our actual practice requires the avoidability principle. A reasonable theory of moral responsibility with an adequate explanatory power does not


\(^{76}\) As we shall see later (p. 93-5), philosophers tend to find the theoretical justification for reactive attitudes on a contractualist ground (some kind of fair-play principle). Wallace seems to hint at this in his discussion of fairness too. So, it is not over-reaching to say that the ought-claims in normative expectations are grounded upon people’s moral status as a member of the community who bears the duty of fair-play.
need such robust control as the power to do otherwise. Thus, I conclude that there is no good theoretical ground for the claim that we must accept the avoidability principle.

**Why We Need Ownership**

Moral responsibility does not require such robust control as the power to do otherwise. But objections can also come from the other side of the aisle: Why do we need an ownership-based concept of moral responsibility at all? Why do we need unattenuated blame or praise, such as reactive attitudes, in our interpersonal relationships?

This question is two-fold. On the one hand, the worry can be why there must be these emotionally charged reactive attitudes, such as resentment, indignation, gratitude and so on. Why cannot we simply form beliefs about other people without any of those emotional sentiments? On the other hand, the worry can be why we should have reactive attitudes that are directed towards an agent, not just towards their behavior or attitude. This is to assume that the agent somehow *owns* the behavior or attitude. But why do we need such ownership-based reactive attitudes at all?

The first worry is a worry about why the practice must take the emotional form as it does, which is a fair question but by no means a pressing one. For Strawson and Wallace, the answer lies simply in the nature of our psychology. Those emotionally charged reactive attitudes are simply natural expression of our frustration or appreciation, period. That being said, I tend to think that reactive attitudes need not always be emotionally charged. As I argued in Chapter One, an increase or decrease of respect could also be a form of reactive attitudes.
The second worry requires more serious treatment. Why should interpersonal relationships work in such a way that we always blame or praise the agent? Why cannot we simply blame or praise behaviors or attitudes? Is there any rational justification for the ownership-based reactive attitudes? Are they simply an arbitrary feature of our psychology?

For some compatibilists, there is a rational justification for the ownership-based concept from a contractualist point of view. For example, James Lenman suggests that being held responsible (in his word, accountable) for behaviors as if one owns them is a form of respect that beings like us would view as “desirable and choiceworthy”. It is fair to blame or praise people for a behavior as if they own it because this is what they would like to be treated as a member of the society. The reason why being held accountable for behaviors as if one owns them is a form of respect is that a being like us badly wants to be taken seriously, to be treated like a “grown-up” – he wants “his words and actions as reflecting on him, as expression of his character, of what sort of person he is”, and he wants to participate in “a party of codeliberation about how to conduct such relationships” as friendship, marriage and so on.

The basic idea behind this contractualist approach is that an ownership-based concept of moral responsibility is justified in the sense that, if we were behind the "veil of ignorance", we would choose to be treated in this way. I agree with this contractualist approach to justify unattenuated blame or praise. This justification, if successful, shows that the Strawsonian theory of moral responsibility, which treats moral responsibility as a social practice involving reactive attitudes, is not simply based on some arbitrary

---


78 Ibid., p. 21.
psychological features of human beings, but also independently and rationally justifiable. I also agree with Lenman that there is good reason to prefer a world in which such ownership-based concept exists to a world in which it does not. But, I think the reason needs to be further specified and strengthened.

In addition to what Lenman has said, I would like to add that the ownership-based concept of moral responsibility is preferable because it is fundamental to our sense of “self” and the meaning of life. To see this, imagine a world where no such ownership-based reactive attitudes exist:

*Indifferent World:* In this “indifferent” world, whenever you do something bad, you hear response like “This is another instance of X-type behavior, the eleventh occurrence in the past twenty-four hours”, and instead of criticizing you for your bad attitude or judgment, people simply look into your brain and say “Here is the neurophysiologic structure that causes that disturbing type of behavior, and it calls for immediate correction by Y-measure”; whenever you do something good, you hear response like “A desirable pattern of Z-behavior is observed; stimulus measure C is to be activated to raise the probability of re-occurrence”, and instead of praising you for your good deed, people simply point to certain features of your make-up and say “It is the combination of these particular features that causes the current desirable pattern of performance.”

It is not hard to imagine that our sense of *self* would be greatly diminished in such a world. A great part of our sense of self comes from how our behaviors and decisions are uniquely characterized, labeled and differentiated from other behaviors and decisions by the society, from how failures or achievements are attributed to *us*, from how valuable
relationships are directed towards us, and from how stories are told of us in the society. If people simply bypass us, the agents, and take an objective stance on the behaviors or attitudes, then we will not be constantly reminded of our existence as a whole individual; rather, we will be no more than a collection of certain types of behaviors and mental states.

Consequently, we would not find much meaning in such notions as autonomy and individuality, and the notion of self-respect would be simply meaningless – for all that is of worth to the society is this type of behaviors or that type of attitudes. A concrete sense of self is essential for the meaning of life. We value our life not just because of the bodily pleasure that a living organism can enjoy. We value it also because of those meaningful relationships that we can achieve and entertain as a member of the community. Without being treated as a whole through behavioral interactions, we would not develop such a sense of self, nor would we acquire a sense of meaningfulness as an individual being. This is why ownership-based reactive attitudes are preferable.

Besides their importance to the sense of self and the meaning of life, I believe there is a further reason why ownership-based reactive attitudes are justified. We know that, according to the Strawsonian concept, reactive attitudes are just natural reactions to the frustration or satisfaction of normative expectations; and such reactions do not require external justification. That is why even if a person's success or failure in meeting the expectation has been causally determined, it is still appropriate for us to have those reactive attitudes. However, this is not to say that reactive attitudes do not have any causal influence on people's future behaviors. Clearly, reactive attitudes can shape people's behaviors by providing incentives for them to comply with the norms in the
future. Having these reactive attitudes in the norm-complying community actually contributes to better overall compliance. In particular, these reactive attitudes help to re-shape people’s decision-making process on occasions like the “checking point” – having received negative reactive attitudes in the past, for instance, one would become more cautious about norm-violating views that he encounters in the future, which makes the conscious or subconscious mechanism that is at work at the “checking point” less likely to give problematic views a free pass. Therefore, our reactive attitudes are actually self-serving in the sense that they reinforce the norm-complying interpersonal relationships. What this shows is that reactive attitudes are not just an arbitrary psychological fact about our human nature: there is good reason why our social psychology works the way it does.

Of course, this does not mean that my account is in any sense consequentialist, like the social-regulation view. The crucial difference to bear in mind is that while the social-regulation view takes reactive attitudes to be essentially forward-looking, that is, the appropriateness of reactive attitudes depending on their outcome, my account takes reactive attitudes to be essentially backward-looking, that is, the appropriateness of reactive attitudes depending on what people have done. Therefore, the appeal to consequence in my theory functions merely as a justificatory foundation of the theory, not as an immediate criterion used in determining whether or not one is morally responsible in a particular circumstance.

**The Moral Outcome Luck Problem**
Recently, Scanlon argues that any attitude-based account of blame suffers the "moral outcome luck problem". Suppose that two drivers are both guilty of careless driving. But one driver’s careless behavior, as a matter of pure bad luck, leads to an accident and kills a young boy. If the manifested attitude or judgment is the basis for our blame or praise, how can we explain it that we tend to blame the driver who happens to have the accident to a much greater degree, even if both drivers have an attitude of the exact same type?^80

Scanlon’s own solution is that to blame someone is not only to judge him or her as blameworthy, but also "to take your relationship with him or her to be modified in a way that this judgment of impaired relations holds to be appropriate."^81 For instance, if a good friend of mine betrays me, his action makes it inappropriate for me to remain in the friendship-type relationship with him; his action provides ground for me to modify my relation with him – to blame him. What explains the "moral luck outcome problem" is that blame, unlike blameworthiness, is "a function not only of the gravity of a person's faults but also of their significance for the agent's relations with the person who is doing the blaming."^82 The reason why it is appropriate for, say, the mother whose son is killed in the accident to blame the unlucky careless driver to a greater degree is that her relation

---

^79 Scanlon (2008), pp. 128-129.

^80 This case is originally discussed in Thomas Nagel's "Moral Luck," in Mortal Questions.

^81 Scanlon distinguishes an account of blameworthiness from an account of blame. On his view, to judge individuals to be blameworthy is to judge that their conduct shows something about their attitudes that impair the relations one has with them; while to blame someone is actually to hold modified attitudes toward him or her. He thinks that blameworthiness is still a judgment on the basis of one's attitudes, but the practice of blaming someone does not require a culpable attitude – an impaired relation is enough for blame. See Scanlon (2008), pp.128-29, p.141.

^82 Scanlon (2008), p.150.
with that driver is impaired by his act of careless driving, but her relation with the other driver is not impaired by the act of his careless driving, at least not in the same way.

I do not think this moral outcome luck problem presents any real challenge to my account of moral responsibility, since my account, unlike many of its alternatives, also distinguishes between the general appropriateness issue – when one is blameworthy or praiseworthy, and the specificity issue – which particular blame or praise is appropriate in a given circumstance. What Scanlon’s discussion highlights here is that the consequence of the actor’s behavior – how the behavior affects the members of the community, also plays a role in determining which particular reactive attitude is appropriate and how strong the reactive attitude should be. If a behavior causes grave suffering to members of the community, it would justify a more severe kind of blame than if it causes just a minor suffering (even if the manifested attitude or judgment is the same in both cases). This should be no big surprise: there are obviously norms telling us that the loss of a life is a greater evil than the loss of a limb.

**Living a Responsible Life**

In this dissertation, I proposed an account of ownership that does not require any robust notion of control, such as classic libertarian free will. To own a behavior in the morally relevant way, I argue, is simply to be the original producer of the attitude or judgment that causes the behavior in question. Based on this compatibilist account of ownership, I further propose a non-volitionist account of moral responsibility, which, I argued, offers the most cohesive explanation of various features of our practice of taking people to be morally responsible for their behavior. According to this non-volitionist account, we can
be morally responsible for behaviors that result from either a conscious, deliberative mental process or an automatic, non-conscious mental process, as long as that mental process involves some exercise of rational capacity, such as explicit judgments or implicit attitudes, and we own that judgment or attitude to a proper degree.

One may worry that, by downplaying the importance of conscious choice in our moral life – that is, by removing conscious choice from the center of our norm-complying interpersonal relationship, my account may have worrisome moral consequences. For example, it may cause members of the moral community to be less careful with their choices. Quite to the contrary. I believe knowing that even a small incident of laziness or insensitivity would lead to some later behavior for which we could be held morally responsible, we would more often engage in conscious introspection and self-examination, and our conscious choice would play a more significant role in our moral life than it does right now. By accepting this non-volitionist account of moral responsibility, we would be more likely to pay greater attention to our behaviors, to be more cautious about our implicit attitudes and prejudices, and to make greater effort in forming the right sort of characters and habits. Consequently, such effort would reduce unwanted behaviors resulting from our moral negligence. We would live a more responsible life.

In this sense, my account of moral responsibility implies certain changes to our current practice. Nevertheless, the main purpose of this project is not to advocate radical changes to our practice. Rather, the primary goal of this project is to provide a theory that best reflects what is actually going on in our real life and gives the best theoretical explanation to it. Traditional theories of moral responsibility, which built upon antique
notions like control and free will, seem to be at odds not only with many of our actual practices, but also with the narrative of contemporary natural science. I hope that I have succeeded in showing that a naturalist account is not only possible, but also preferable.
BIBLIOGRAPHY


VITA

Xiaofei Liu was born in Lengshuijiang, a small town in the central part of Hunan Province in south China. He received his bachelor degree from Beijing University, where he majored in law. After a few years of working outside of the academia, he went to Scotland and studied philosophy at the University of St. Andrews, where he received his Master of Letters. He entered the University of Missouri-Columbia in the year of 2006, and received his M.A. (philosophy) in 2008 and PhD (philosophy) in 2012 under the supervision of Dr. Peter Vallentyne.