Responsibility & Reincarnation: A Gaudiya Vaishnava Approach to Compatibilism and Sourcehood

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Arts (MA)

at

The University of Waikato

by

Kalindi Fletcher

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2024

Responsibility & Reincarnation:

A *Gaudiya Vaishnava* Approach to Compatibilism and Sourcehood

Abstract

This work explores the possible integration of *Gaudiya Vaishnava* theology to fortify the case of a compatibilist understanding of free will and determinism. As the debate surrounding moral responsibility persists, the compatibilist approach stands as a promising and popular position among philosophers, yet has not been able to successfully overcome its shortcomings concerning the ability to provide a satisfactory account of sourcehood that is required for the possession of moral responsibility for our actions. In response to the limitations of modern compatibilist theories, this work illuminates how adopting a few relevant *Gaudiya Vaishnava* metaphysical concepts – particularly, the beliefs of an eternal soul capable of reincarnation – enhances the compatibilist stance by offering a compelling account of sourcehood that preserves both our intuitions and rationality. Specifically, this work offers both a *soft-line* and *hard-line reply* to Derk Pereboom's Four-Case Manipulation Argument, and a response to Galen Strawson's Basic Argument. By tapping into the insights of Indian philosophy that is rooted in the Ancient Vedas, this study also demonstrates the significant value that non-Eurocentric philosophy can provide for our contemporary philosophical discourse.

Table of Contents

Introduction

The following study delves into the potential integration of *Gaudiya Vaishnava* theology into modern theories of *source compatibilism*. Amidst the ongoing debate regarding a reconciliation of free will and determinism, the compatibilist stance has grown popular among philosophers despite the many problems concerning the theory's coherence, particularly, the difficulty for compatibilist theorists to provide a satisfactory account of sourcehood that is required for moral responsibility.[1] In response to the limitations of modern compatibilist theories, this work illuminates how from adopting the *Gaudiya Vaishnava* conceptions of an *eternal soul*, the doctrine of *karma*, and the process of *reincarnation*, a compelling account of sourcehood can be developed in order to enhance the overall coherence of the compatibilist theory of moral responsibility.

My goal with this work is not to assertively delineate the exact position on free will that all *Vaishnavas* must hold, for at present there is plenty of room for interpretation. Nor is it my goal to cherry-pick elements from the spiritual tradition without the consideration of doctrinal coherence and consistency. My intention is rather to present a Gaudiya Vaishnava approach to compatibilism that is, hopefully, compelling enough to entice proponents of compatibilism into considering the adoption of certain features of Vaishnava philosophy into their own conceptions of responsibility.

To preface my analysis and arguments, the first two chapters of this thesis will be overviews required to *set the scene* for the discussion, and will both function as a type of literature review of both the East and West: chapter one will offer a brief summary of the current field of free will, responsibility, and determinism; and chapter two will be an introduction to the essential *Gaudiya Vaishnava* theology and metaphysics that pertains to my conception of freedom. In chapter three I will begin to raise the central issues around a compatibilist's struggle to overcome objections of manipulation. But rest assured, in chapter four I will attempt to provide a *Gaudiya Vaishnava* approach to avoiding the manipulation objections, specifically, one that offers both a *soft-line* and *hard-line reply* to Derk Pereboom's Four-Case Manipulation Argument, and a response to Galen Strawson's Basic

[1] According to a 2020 survey among philosophers globally, nearly 60% self-identified as holding a preference towards a compatibilist view (Bourget & Chalmers, 2023, p. 7).

Argument. Lastly in chapter five, I will address some of the potential objections that may arise concerning the notion of karma and the role of God.

Finally, I shall note here that for practicality sake, from now on I will mostly use the abbreviation *GV view* to refer to the position I am developing forthwith.

Chapter One

Overview of Free Will and Moral Responsibility

The Free Will Problem

Daniel C. Dennett once said that 'when we consider whether free will is an illusion or reality, we are looking into an abyss. What *seems* to confront us is a plunge into nihilism and despair' (Overbye, 2007, para. 7). Both attempting to affirm free will or deny it is an exhausting task. The idea that we possess free will is a belief that runs extraordinarily deep in the human psyche, for it underpins our senses of morality and responsibility, our sense of right and wrong, our senses of achievement and failure. Yet, philosophers have been contemplating what it *means* to have free will for many thousands of years. In the modern era too, neuroscientists, quantum physicists, and psychologists have made their contributions, and whilst distinctive schools of opinion have been established over the decades, a consensus among academics has hardly been found.

Free will is generally described as an ability to control one's actions according to one's own discretion. As I stand in front of my pantry, pondering what to eat for breakfast, I intuitively believe that I am completely free to choose either toast or porridge, and I intuitively believe that if I choose porridge rather than toast, I did so *freely* in that there was nothing preventing me from choosing toast if I had wished to eat toast instead, and also that there was nothing *forcing* me to choose porridge, but that I chose that from my *own* intelligence and ability to deliberate between the choices.[2]

One threat to this firmly held belief is the opposing intuition that everything has a cause for its existence or occurrence – that everything has some kind of explanation as to how it came about. Our experience in this world is that things follow reliable patterns and laws: a jar of honey does not randomly appear out of thin air, but we believe there is some causal chain in how it came about: it was placed there by a friend – who brought it from her beehives – where the bees developed it from nectar – which the bees collected from the flowers that the gardener friend planted. Even on a very local level we think it's rational to believe that everything has some explanation: externally it may appear that I chose at

[2] The alternative approach of defining free will as being an *ultimate source* of one's actions will be addressed later in this chapter.

random to eat porridge instead of toast, but we can theorise easily that perhaps I decided to roll a die to decide, odd number porridge, even number toast, and that it landed on a 3 and thus I ate porridge. The die landing on a 3, though, was caused, among other things, by the wind blowing through from the kitchen window, and the window being open was caused by a decision to open it after somebody had burnt their breakfast and smoke filled the room, and the reason why they burnt their breakfast was because they took a phone call, and so on. Even if, say, a religious man sees a jar of honey appear out of thin air and exclaims 'by the grace of God this has appeared! A mighty miracle!' there is still an implication that God played a *causal* role in the jar of honey coming into existence. The controversial point here is that if we hold it to be true that something cannot exist without a cause, then that implies that 'that the course of the future is entirely determined by the conjunction of the non-relational past and the laws of nature' (Timpe, 2008, p. 12). And that given the past and the laws of nature etc, there is only one possible future. The future is not open like a garden of forking paths, but instead like a singular track we can't help but walk through. What I'm describing here is *causal determinism*, which, in more specific terms, is understood as the principle that *every event is necessitated by antecedent events and conditions together with the laws of nature* (Hoefer, 2023, para. 1).[3]

The point of contention is, of course, how can these two facts about the universe be simultaneously true: how can somebody maintain their free will, if it's true that there are causal conditions that necessitate a particular choice being chosen? Though we may be under the impression that we are deciding *freely* between a range of options of actions, determinism holds that there can actually be only one course of action available to us due to the causal factors operating prior to your deliberative decision at a point in time, and there is nothing to prevent that outcome from occurring.

At this point, we may be contemplating a move to reject the truth of determinism in order to preserve our free will. Or perhaps our intuitions pull us to a rejection of free will and an acceptance of determinism. Both these positions are under the umbrella category of *incompatibilist* theories of free will – theories that deny the compatibility of freedom and

[3] Causal determinism is not to be confused with *logical determinism* which holds no real threat to free will. Nor should it be confused with *theological determinism*, which though it *can* hold a threat to creaturely free will, it often operates in a similar causal role as causal determinism. In chapter five, I will touch on Gaudiya Vaishnava perspectives on God's causal role.

causal determinism. In technical terminology, *libertarianism* is the type of incompatibilist position that affirms free will and denies determinism, and at the opposite end of the indeterminist spectrum is *hard determinism*, the position that rejects the existence of free will and affirms causal determinism. Both types of incompatibilist positions, however, have been victim to serious and persuasive objections raised by philosophers that appear to undermine one's ability to accept either of them.

Say we reject determinism in order to affirm free will; this rejection of the truth of determinism involves a necessary acceptance of some *indeterminism*. *Indeterminism* holds that it is *not* the case that all events are necessitated by antecedent events and conditions. The truth of this, however, appears to imply a type of randomness, chance, or luck involved in the actualisation of events. We can ponder like this: if an agent M has two options of A and B before her, and her choice is completely free within an indeterministic world, then it could be conceived that with 'exactly the same prior deliberation, the same thought processes, the same beliefs, desires, and other motives (not a sliver of difference!) that led to' agent M favouring and choosing option A, it might have issued in her choosing B instead (Kane, 2005, p. 16). Indeterminism here appears to *undermine* freedom just as much as determinism does, due to the lack of ability for the agent to *control* their actions. If an agent's choice comes down to chance or mere randomness, then it no longer appears *up to them* in an adequate sense, so can the agent be said to be *responsible* for the outcome? R. Kane illustrates this dilemma in what he calls *Incompatibilist Mountain*: the point that if we are to assert that free will is incompatible with determinism (the ascent up the mountain) then we must also offer an explanation as to how free will is compatible with indeterminism instead (the descent down the mountain) (Kane, 2005, p. 34).

If we follow these lines of argumentation, then it appears that free will is incompatible with both determinism *and* indeterminism, thus we may end up considering the position at the extreme end of the spectrum of free will theories, a position that affirms causal determinism and rejects the existence of free will altogether. The main question we would then have to answer is: what does that look like? And can we live with that? As I mentioned earlier, it is this belief that humans possess free will that underpins our very conceptions of ethics and morality. Free will underpins our conceptions of holding people accountable and responsible for their decisions, and distributing blame and praise accordingly. This perspective of the world in which we live in is in great threat in light of a

rejection of free will. If we hold the view that people lack the freedom to decide and actualise the events and actions performed by them, then it appears that this truth of determinism strips the agents of their deservedness. Consider an agent Sam who one day steals a car, normally we would wish to punish Sam for his crime as it was cruel – perhaps we imagine a few months in prison will suffice. However, because agents like Sam lack control over their actions due to the truth of determinism ruling out free will, then Sam could not be held responsible for the stealing of the car.

What we'd have to do, according to philosophers like Honderich, is swap our *retribution* theory of punishment for a *deterrence* or *reform* theory of punishment (Kane, 2005, p. 75). So whilst we're no longer able to punish agents like Sam with the justification that they *deserve* it, we could still justify punishing Sam in order to deter the performance of these types of actions in the future, or justify it in how Sam's experience of punishment will reform his character – despite not being deserving of any punishment in the first place. In a similar line of reasoning, another determinist, D. Pereboom, argues that just as societal leaders have the right to quarantine those who are carriers of severe infectious diseases in order to protect the wider society, then similarly it would also be justified to imprison somebody who has acted as a serial killer – quarantining them from society; 'Even if he is not morally responsible for his crimes, it would be as legitimate to detain him as it is to quarantine a non-responsible responsible carrier of a deadly communicable disease' (Pereboom, 2007, p. 116). In this way, we might be able to cling on to some of our intuitions around the wicked suffering some consequence for what they've done, but arguably, the requirement to give up the intuitive sense of somebody *deserving* a consequence is perhaps too ingrained in our psyche to accept a denial of free will.

We encounter the same undermining of intuition at the opposite end of our emotive reactions when wanting to praise and thank those who have performed actions we would normally consider worthy of gratitude, for according to this type of incompatibilism, they wouldn't ultimately be responsible for their good deeds. According to philosophers like Pereboom, however, 'we could go on believing that acts of certain kinds, say, of generosity and heroism, are admirable and that acts of other kinds are despicable even if we [do] not believe that persons are ultimately responsible. Gratitude, for example, he [Pereboom] says, "typically involves joy occasioned by the beneficent act of another. But hard incompatibilism fully harmonizes with being joyful and expressing joy when others are considerate and

generous on one's behalf"' (Pereboom, 2002, p. 485, as cited in Kane, 2005, p. 77). Perhaps some may feel that this adjustment to our outlook is a perfectly reasonable next step following an illumination of determinism, however, I can't help but anticipate a wave of difficulty in justifying our reactive attitudes – emotional responses such as resentment and gratitude that presuppose agent responsibility – without our foundational conception of free will ensuring deservedness.[4] There have been many creative and impressive strategies to make this incompatibilist position more comfortable to accept, however, I do not have the space here to engage with them.[5] For now though, considering the great sacrifice this position requires, let's continue to consider the other ways in which philosophers have rationalised about free will.

At this point, we may have gathered a sense in which neither position offers an epistemological advantage over the other. If we reject determinism, then we're left with a free will that is undermined by randomness. And if we accept determinism, then it appears we're forced to concoct new and peculiar ways to deal with our intuitions around justice, fairness, guilt, and responsibility, etc. Neither theory offers an account of free will that comes close to our intuitive beliefs and how we employ this idea of free will in our everyday lives. And with neither theory possessing a clear advantage over the other, we might perhaps consider turning to, or waiting on, scientific discovery in order to decide which theory of free will and determinism is more plausible.

Some philosophers, however, have argued that the truth of determinism is totally irrelevant – that no matter what science tells us about the universe's causal laws, free will could *never* be possible on *any* account. One prominent philosopher who has fought for this point is Galen Strawson who argued that due to an inability for anybody to be an *ultimate source* of their actions 'it makes no difference whether determinism is true or false. We cannot be truly or ultimately morally responsible for our actions in either case' (1994, p. 5). Strawson's argument will be examined more closely in chapter four, but in sum, he argues that because all human actions have a causal chain that extends outside of the agent's scope of control (for example an agent cannot control who their parents are or the country they

[4] The term '*reactive attitudes'* was coined by P.F. Strawson in his essay *Freedom and Resentment* (1962). Proceedings of the British Academy 48:187-211.

[5] For an interesting incompatibilist approach that involves the concept of reincarnation, see Ayon Maharaj's '*Hard Theological Determinism and the Illusion of Free Will: Sri Ramakrishna Meets Lord Kames, Saul Smilansky, and Derk Pereboom'* Journal of World Philosophies 3 (2):24-48 (2018)

were born in), then the agent can never be truly morally responsible for their actions as they can never be the ultimate source of all their actions or *self-determining*. Though the merits and flaws of this philosophical position will be examined more thoroughly later on in this work, what I'd like to point out here is the advantage the theory offers in its avoidance of relying on scientific discoveries to inform which ship to board. It doesn't deny that science might be able to offer a clear answer to if determinism is true or not, but by accepting a theory like Strawson's, we are not 'held hostage to the possible scientific discovery that determinism is in fact true' (Fischer, 2002, p. 307). Though these theories may feel less problematic because of this advantage, it would be unreasonable to accept this type of theory because of this factor alone. However, what is relevant to our contemplation is the question of the *relevancy* determinism has over our conception of free will and moral responsibility.

Though so far we have discussed some ways in which determinism and indeterminism undermines freedom, there are some strains of philosophy that deny it has a relevant place in our discussions on responsibility. Though Strawson defends the position that determinism is irrelevant because moral responsibility can never been true on either case, some philosophers hold the position that the truth of determinism is in some sense irrelevant because moral responsibility can exist *regardless* of the scientific status of determinism. One prominent proponent of this theory is J.M. Fischer – a type of *compatibilist* or '*semi-compatibilist*' that believes causal determinism is compatible with the existence of moral responsibility.

Unlike the previously described incompatibilist positions of *libertarianism* and *hard determinism* that commit to rejecting one of the beliefs – either in the existence of free will or the existence of determinism, *compatibilism* is the position that holds that there is no inconsistency with holding free will and determinism simultaneously true. Consequently, whilst the task for incompatibilists is to convince us to reject one of the strongly held beliefs, the equally difficult challenge for compatibilist proponents is to offer an explanation as to how the two beliefs can be simultaneously true without undermining freedom or undermining the causal capabilities of determinism.

Whereas a compatibilist position affirms the simultaneous truth of determinism and the existence of free will, on the *semi-compatibilist* position of Fischer, however, one can remain both agnostic towards the truth of determinism *and* agnostic towards the truth of

whether or not free will is compatible with indeterminism or determinism (Fischer, 2007, p. 56). Instead, what a *semi-compatibilist* affirms is that moral responsibility is compatible with both determinism and indeterminism. Though moral responsibility is often regarded as something that requires a strong sense of free will, philosophers like Fischer and Ravizza propose that there is an alternative, weaker type of control that is sufficient for moral responsibility, which is independent from the control that's required for free will.[6] The position is a *compatibilist* stance in the sense that it affirms that determinism is compatible with moral responsibility, however, it is *semi* in the sense that it doesn't commit to the truth of determinism like compatibilists do, and it also places more emphasis on the possession of *moral responsibility* than of *free will*.

Such a theory, if it were proven reasonable, would certainly be an appealing choice to commit to. It would not require us giving up our intuitions regarding individuals being responsible for their actions, nor our intuitions regarding the causal powers of the universe. And it would seem to possess an upper hand in how proponents 'need not 'flipflop' in this weird and unappealing way' depending on the scientific consensus on determinism (Fischer, 2007, p. 47). What Fischer and Ravizza's theory does require, however, is a very specific type of control that not all philosophers have agreed is adequate, though if it were, it would not be attainable regardless of whether the world functions deterministically or not (this control criteria will be discussed more thoroughly in chapter three).

As I have teased in so far, these beliefs in free will, moral responsibility, determinism, and indeterminism have ignited a myriad of methods to resolve the supposed tension between the concepts. And as philosophy has progressed, the various positions presented and considered by philosophers have developed into three major schools of reasoning: *libertarianism*, *hard determinism*, and *compatibilism*. In this thesis, the discussion around compatibilism, particularly, compatibilist theories of *sourcehood*, is at the heart of my proposal. Over the last five decades, compatibilist philosophers have developed strong and impressive defences of their position. However, there is a strong sense in which these theories are continually undermined or fall short at some point due to the metaphysical worldview upon which their theories are applied. Some philosophers are happy to bite the bullet, whereas some take it as evidence for the incoherence of the position. What I'm

[6] This proposed type of control will be discussed in chapter three.

interested in here, though, is building upon current compatibilist theories and assessing their success when applied to a different philosophical worldview – one that involves some of the key metaphysical beliefs of *Gaudiya Vaishnavism*.

The Requirement for Free Will

The intuition that we have free will flows deep through the human psyche, and from it blooms the inescapable emotions of guilt, embarrassment, remorse, pride, gratitude, joy, and so on. What's worth contemplation here is how these sensations often imply an underlying sense of ownership and responsibility. Guilt that penetrates the mind may provoke the discomfort of feeling like we have done something wrong, and regretting having done it; perhaps we are weary of future consequences because we know we are blameworthy; perhaps we feel pathetic for not doing better; perhaps we feel depressed for letting others down. What's significant here is that we have these sensations as a result of an underlying belief that we had control over our actions and decisions. We do not feel guilty for an earthquake destroying our neighbours home, for we can play no role in bringing that event about; but we may feel guilty if we vandalised it in graffiti during the night, for that action stemmed from our own decisions to perform it. Similarly, the way we interact with others is underpinned by a belief that others are also in possession of free will, hence why I may feel a sense of gratitude for the supermarket employees, for they are actively deciding to serve me so nicely, yet I lack all compassion for the self-checkout machines who operate without choice.

In other words, we often associate having free will with *being in control* of our actions, and our decisions *owing to us*. Having free will means possessing the liberty to decide how to act, and being responsible for what results from it. Along these natural lines of thinking, some philosophers in this discourse invoke the imagery of a garden of forking paths. We feel as though, so long as we aren't being manipulated or coerced, that when we stand in a garden staring out into a field of pathways winding through the flowers, that we possess the power to choose whichever path to walk down, if we so desired to walk down it. Freedom appears to imply necessarily that the future is open in this sense, that there is nothing preventing me from choosing any of the paths ahead of me. For if there was something that prevented my choosing path A, then I was not genuinely free to choose path A. Nor would we feel free if there was something or someone ensuring that I picked path A,

for if it were guaranteed that I would choose path A, then I wasn't genuinely free to pick any of the other paths.

For centuries, free will has been understood in this sense of *ability to do otherwise* – meaning that, although there will be one exact path that I will decide to walk down, this decision is a free decision because I possess the *ability to do otherwise* – to walk down a different path if I decided to do that instead. In 1969, philosopher Harry Frankfurt described this concept as the *Principle of Alternative Possibilities*: that 'a person is morally responsible for what he has done only if he could have done otherwise' (Frankfurt, 1969, p. 829). Intuitively, or at first glance, this principle may feel uncontroversial. As a consequence, however, some philosophers have employed this principle and conception of free will to argue against the compatibilism of free will and determinism.

As discussed earlier, casual determinism can be conceptualised as a principle that '*every event is necessitated by antecedent events and conditions together with the laws of nature*' (Hoefer, 2023, para. 1). If this were to be true, then it would appear to rule out an *ability to do otherwise*, for in this sense, if determinism is true then the garden of forking paths aren't legitimately open; though it may feel as though we are conducting a decision, the path we will choose has already been guaranteed and fixed to happen due to the history of the universe and the laws of nature. Peter Van Inwagen puts it straightforwardly that: 'if determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our own acts) are not up to us.' (van Inwagen, 1983, p. 16). In Inwagen's discussion of this topic, he develops the influential *Consequence Argument* which can be formulated in simple terms as the following points:

(1) There is nothing we can now do to change the past.

(2) There is nothing we can now do to change the laws of nature.

(3) There is nothing we can now do to change the past and the laws of nature.

(4) If determinism is true, our present actions are necessary consequences of the past and the laws of nature. (That is, it must be the case that, given the past and the laws of nature, our present actions occur.)

(5) Therefore, there is nothing we can now do to change the fact that our present actions occur. (Kane, 2007, p. 10).

If valid, the argument appears to rule out one's *ability to do otherwise* in any scenario, for if we cannot control the past or the laws of nature, which according to determinism, guarantees a particular future, then we cannot do contrary to what the future will be.

Despite the strength of the argument, some philosophers have presented objections to the validity of it. Fischer describes how some compatibilists, *Multiple-Pasts Compatibilists*, defend a position that argues that we 'can sometimes so act that the past would have been different from what it actually was' rejecting the truth of the first premise of the argument; whereas other compatibilists called *Local-Miracle Compatibilists* have argued that 'we can sometimes so act that a natural law that actually obtains would not have obtained', rejecting the second premise of the argument (Fischer, 2007, p. 55). These two approaches don't posit an actual ability to change the past or universal laws, but a counterfactual power over the past and/or laws of nature. What the compatibilist proponents of these positions argue is that one's *ability to do otherwise* can be understood not just in the strict sense of: if an agent freely willed A then, at the time of acting she could have done B and not A – where determinism is ruled out according to the *Consequence Argument*, but that it could also be interpreted by a conditional analysis of a counterfactual circumstance: we can suppose that in the case of an agent X who freely willed A, this means that *had* agent X wanted to do B and not A at that time, *then*, agent X would have done B and not A (McKenna & Coates, 2024a). In simpler words, these compatibilist theorists propose that when we act freely, we possess the ability to do something such that the past of laws of nature would have been different. It's not a power that is actually executed, but if it was, then the past or laws of nature would have been different. And in this way, some compatibilists argue that this preserves the *ability to do otherwise*, but does not fall victim to the *Consequence Argument*.

There has been, however, a great deal of debate over whether or not such a conditional analysis of the *ability to do otherwise* is a viable route of avoiding the conclusion of the *Consequence Argument*. Even within the compatibilist camp, not all compatibilists believe that the *Consequence Argument* is such a threat to compatibilism that is requires a reply at all – Fischer, for example, points out that it is a viable option for philosophers to remain completely agnostic to whether the *Consequence Argument* succeeds or is undermined somehow, as there exists another way of defining free will and moral responsibility that does not fall victim to such arguments.

At the present, it may seem essential here for a compatibilist to explain how determinism doesn't rule out our *ability to do otherwise* as the consequence argument appears to hold. However, the other route open for the compatibilist is to explain why the *ability to do otherwise* isn't essential for free will or moral responsibility in the first place at all. A compatibilist (or semi-compatibilist) can agree with the plausibility of the consequence argument in its success of affirming that determinism rules out alternative possibilities – but disagree on its relevancy regarding what constitutes our decisions and actions being *up to us*. In order to do this, what I shall do now is return to Frankfurt's discussion of the *Principle of Alternate Possibilities*, and raise some of the influential cases that question the principle's plausibility and relevancy.

The Frankfurt-Style Cases

Earlier, I mentioned Frankfurt's coining of the *Principle of Alternate Possibilities* to describe the belief that 'a person is morally responsible for what he has done only if he could have done otherwise' (Frankfurt, 1969, p. 829). Within the same paper, though, Frankfurt goes on to explain how it is a false understanding. It is true that an agent's ability to act otherwise is undermined or eliminated entirely when an agent is force to act under some coercion or compulsion. However, Frankfurt explains that there could also exist circumstances such that an agent cannot act otherwise than how the agent decides to act, but still that 'do not actually impel the person to act or in any way produce his action'; Frankfurt writes that a 'person may do something in circumstances that leave him no alternative to doing it, without these circumstances actually moving him or leading him to do it – without them playing any role, indeed, in bringing about that he does what he does' (Frankfurt, 1969, p. 830). In sum, although, for example, the inability to do otherwise may be a consequence of an agent being manipulated into doing *x*, it is not the state of being unable to do otherwise than *x* that eliminates their freedom, but rather, in this case, it is the manipulator themselves.

This point made by Frankfurt in 1969 has since developed into an entire genre of literature within the field of free will, often called *Frankfurt-Style Cases* or *Frankfurt-Style Examples*, that offer possible scenarios where agents lack alternate possibilities, yet should still be held morally responsible for them. To further our discussion on what constitutes free will, consider the following Frankfurt styled example offered by Fischer:

> Black is a nefarious neurosurgeon. In performing an operation on Jones to remove a brain tumor, Black inserts a mechanism into Jones's brain which enables Black to monitor and control Jones's activities. Jones, meanwhile, knows nothing of this. Black exercises this control through a computer which he has programmed so that, among other things, it monitors Jones's voting behavior. If Jones shows an inclination to decide to vote for Carter, then the computer, through the mechanism in Jones's brain, intervenes to assure that he actually decides to vote for Reagan and does so vote. But if Jones decides on his own to vote for Reagan, the computer does nothing but continue to monitor—without affecting—the goings-on in Jones's head.
>
> Suppose Jones decides to vote for Reagan on his own, just as he would have if Black had *not* inserted the mechanism into his head. Then Frankfurt claims that Jones is responsible for voting for Reagan, regardless of the fact that he could not have done otherwise. His voting for Reagan is something we can charge to his credit or discredit; it expresses something about Jones's character. But the principle of alternate possibilities does not seem to be satisfied; there is no way that Jones could have avoided voting for Reagan – if he had been on the verge of doing so, the computer would have intervened and ensured that he vote for Reagan. If the computer *had* intervened, then I think it is obvious that Jones would not have been responsible. But since the computer did *not* intervene, it is plausible to think, along with Frankfurt, that Jones is responsible. (Fischer, 1982, p. 26)

What Frankfurt-styled cases like the one above attempt to show is that the agent's actual ability to do otherwise is not always relevant or essential for considering them responsible for their decided action. Some of those affirming determinism will make the connection that just as the neurosurgeon blocks Jones's ability to do otherwise, which doesn't diminish Jones's responsibility, similarly, determinism operates in the world such that it can eliminate an agent's ability to do otherwise, yet (at least in some cases) play no role in diminishing the agent's responsibility.

The literature that has been developed surrounding the success and limitations of these styled arguments are vast, detailed, and complicated. Though many compatibilists will employ these cases to defend a view that though determinism eliminates alternate possibilities, alternate possibilities aren't required for responsibility, there are some compatibilists who accept that the ability to do otherwise is essential for free will and reject it's incompatibility with determinism – often by offering subjunctive or conditional

definitions of an ability to do otherwise (Timpe, 2008, p.26).[7] On the other side too, there are many incompatibilists who raise arguments to show that Frankfurt-styled examples aren't genuinely free from alternate possibilities – in order to affirm their theory that responsibility requires alternate possibilities which is incompatible with determinism (see descriptions of *Leeway Incompatibilism* in Timpe, 2008, 72-73). Yet, there are some other libertarian incompatibilists who accept, to some degree, the ability of Frankfurt-style cases to demonstrate how alternate possibilities *alone* is insufficient to be all that is required for freedom and responsibility (yet still, as libertarian incompatibilists, affirming that alternate possibilities do exist) (most influential among these philosophers being Kane who argues for a further requirement of *sourcehood.* See Kane, 2017).

Rising out from the literature and the nearing impasse, there's a sense that, at the very least, the *Principle of Alternate Possibilities* is *not* the complete picture of free will and responsibility. Whether we believe that alternate possibilities are completely irrelevant, or that they're partially relevant, there is a need for further discussion on what constitutes free will and moral responsibility. There has been a great deal of detailed study regarding how much can be deduced from Frankfurt-style examples, and though I have no room here to give an analysis of the various objections, I can agree with Timpe's point that 'even if alternative possibilities are a necessary condition for free will, not all of them will be equally morally relevant; so the focus simply on their presence (or absence) will not get at the heart of the matter' (2008, p. 72). We have to look outside of the debate of alternate possibilities and begin considering what it means for an agent to possess control over their actions if we are to get to the heart of what is required for free will and responsibility: '"[f]or the agent to have control, in the relevant sense, there must be an alternate sequence in which the agent does otherwise as a result of *an appropriate sort* of chain of events."… The mere existence of alternative possibilities does not entail the moral relevance of such alternatives beyond indicating that causal determinism is false' (Fischer, 1982, p. 31, in Timpe, 2008, p. 73).

What we have been examining thus far is a conception of free will centred around an ability to do otherwise, however, this is just one of the two main approaches – the second approach sees free will 'primarily a function of an agent being the ultimate source of his

[7] These types of compatibilist views are often called '*Strong Compatibilism'* and '*Conditionalism'*. See Timpe, 2008, Chapter Two for a further description.

actions and thus free will is grounded in a sourcehood condition.' (Timpe, 2008, p. 71). As Frankfurt-style arguments start to show the weakness of an alternate-possibilities-only approach, 'much of the literature has begun to focus more directly on the issues of "sourcehood," "ultimacy," and "origination"' (Timpe, 2008, p. 73). In a basic sense, the idea is simply that a 'person chooses freely only if he is the source (or originator or initiator) of his choice' (Timpe, 2008, p. 75). Intuitively, this feels non-controversial: when we are under the impression that we have performed an action freely, we also possess an understanding that we are the *causes* of the action; an action is *mine* because *I* brought it about, *I* made it occur.

On this sourcehood conception, when Jones votes for Reagan, he does so freely because he is the origin and initiator of the action, whereas, if Jones attempts to vote for Carter and is manipulated by Black's mechanism to vote for Reagan, then Jones is not the adequate source of that action of voting for Reagan – thus, he lacked free will.

The incorporation of a type of sourcehood condition required for free will and responsibility has been developed in various theories across libertarianism, compatibilism, and hard determinism. In Kane's conception of libertarianism, for example, he incorporates a necessary criteria of *Ultimate Responsibility* (UR) that holds that '[*t*]o *be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient cause or motive for the action's occurring*' (Kane, 2007, p.14). Kane argues that though we often make decisions and perform actions based on wills within ourselves that have already been formed, these wills have developed and risen by virtue of our past free choices and actions (Kane, 2005, p. 121). For agents to possess *ultimate responsibility* – which is required for Kane's conception of free will – our actions must be traceable back to one or more of our will-setting actions or '*self-forming* actions' as 'they would be the actions in our lives by which we *form* our character and motives (i.e., our wills) and make ourselves into the kinds of persons we are' (Kane, 2005, p. 130).

In a similar way, Frankfurt himself responds to his rejection of the *principle of alternate possibilities* by developing a type of sourcehood criteria involving a meshing of *first and second-order desires*. According to Frankfurt's theory, an agent has free will so long as their desires or volitions are rooted in a deeper second-order desire that moves them to perform the action (1988, chapter two). It is a hierarchical conception of free will, with the basic idea that our actions must have a proper source – which are deeper wills that also play

a role in the action-sequence. But whilst Kane argues that alternate possibilities must still be required at some point in will-setting, Frankfurt maintains that his conception of sourcehood doesn't require any alternate possibilities at all for free will; further, Frankfurt's position is ahistorical in that he does not regard the causal history of an action as relevant at all – so long as the agent *wholeheartedly identifies* with the action, which involves a necessary meshing of desires, then the agent is an adequate source and can therefore be considered free and responsible *without* an *ability to do otherwise*.

Fischer and Ravizza's conception is slightly more unique, positing that what is essential for moral responsibility is an agent's ability to be *reasons-responsive* and that this mechanism is owned by the agent appropriately. According to them, whilst we can remain agnostic as to whether or not an agent can possess what Fischer and Ravizza call *regulative control* – control over what future of genuinely open possibilities becomes actual, agents can definitely possess *guidance control* – the ability to 'act on a mechanism that is regularly receptive to reasons, some of which are moral reasons'; the conception that essential for moral responsibility is that the agent holds the ability to rationalise, reflect upon reasons, and deliberately decide to perform an action (Fischer & Ravizza, 1998, p. 82). If an agent's action results from coercion or manipulation such as hypnosis or brainwashing, for example, then the agent wouldn't possess the ability to guide their actions in a way that they can be responsible for. If, however, an agent's action results from the agent's own psychological decision-making process in the mind that is responsive to reasons and free from external manipulation, then they do possess an adequate *reasons-responsive* mechanism that enables them to be considered morally responsible for their actions. It's a type of sourcehood approach in that it isn't concerned with actual alternate possibilities, but rather that the actions have an origin and proper source in the agent as being a result from the agent's *reasons-responsiveness* process, and that this deliberative mechanism of the agent originates in them in an appropriate sense.[8]

Even some hard incompatibilists like Pereboom are also proponents of the idea that for an agent to be free, they must be the proper source of their actions. Like Fischer, Pereboom is agnostic to whether or not determinism is true, but unlike Fischer, Pereboom

[8] What exactly is meant by the mechanism being owned 'appropriately' in Fischer and Ravizza's theory will come under scrutiny in chapter three.

shares Strawson's view I previously mentioned that free will is incompatible with both determinism and indeterminism (Pereboom, 2007, p. 85). Pereboom's reason for this judgement that free will is never attainable owes *not* to the lack of an ability to do otherwise, but due to the inability for agent's action to ever possess 'a causal history of a sort that allows the agent to be the source of her action in a specific way' (Pereboom, 2007, p. 86).

I have listed briefly here only a few examples of popular theories that depend less upon an alternate possibilities definition of free will, and adopt one based upon sourcehood instead. Throughout this thesis, these theories will be analysed and discussed with a higher degree of scrutiny than I provided so far. But for now, though, I hope to have set the scene adequately enough with regard to the challenge that faces both incompatibilists and compatibilists, in order for me to begin now to introduce a new metaphysical worldview and develop from there an additional theory of free will and moral responsibility to throw into the ring and consider. In the next chapter, you will hopefully be able to trace the similarities between the conceptions of free will, responsibility, and causal determinism that make up the current discussions in the philosophy of free will mentioned so far, inside the philosophy and theology of the *Gaudiya Vaishnava* tradition, despite being significantly separated by geography and time.

Chapter Two

Introduction to Gaudiya Vaishnava Theology and Metaphysics

The Philosophy of Gaudiya Vaishnavism

Before I embark on an integration of Gaudiya Vaishnavism into compatibilism, I should first give a brief overview of the relevant philosophical beliefs of the spiritual tradition and what it entails when I speak of a *Vaishnava worldview*.

The origin of the employed term *Vaishnavi*sm is the ancient *Vedas*. The *Vedas* serve as one of the oldest religious scriptures of the world, and the basis for what we refer to now as the religions and practices of *Hinduism*. Filled with rich descriptions and glorifications of the many thousands of Gods, Goddesses, and personalities, the *Vedas* offer many paths of religious worship for spiritual and material advancement according to the needs and desires of the individual. Out of this variety that the *Vedas* offer, denominations of Vedic thought and interpretation have evolved since the Vedic inception, and there exists now various competing claims of who the supreme sole God of Gods is. Some believe it is Shiva, some think Ganesh, Vishnu, an impersonal Brahman, and so on. Most followers of Vedic religion, though, take Vishnu to be the supreme deity, and this monotheistic denomination is thus known as *Vaishnavism*.[9] *Gaudiya Vaishnavism*, however, is the religious movement founded in the 16th century in the *Gauda* region of Bengal, hence the term *'Gaudiya'*, established by *Sri Chaitanya Mahāprabhu* who is revered as both a saint and an incarnation of God Himself by followers – for this, the tradition is also commonly known as *Chaitanya Vaishnavism.*[10]

For around 700 years prior to the birth of the Gaudiya Vaishnava sect, philosophical dialogue and debate thrived, giving rise to the tradition of *Vedanta* – an umbrella term for schools of Indian philosophy primarily concerned with interpretations of the Bhagavad-Gītā*, Upaniṣads,* and the *Brahma Sūtras*. The heavy and rich debate generated distinct schools varying from strictly monist views (*Advaita Vedānta*), to dualist views (*Dvaita Vedānta*), and combinations of the two (Viśiṣṭādvaita, Bhedābheda, etc). By the time Chaitanya appeared, though, 'Vedanta [had] eroded to some extent during the fifteenth and sixteenth centuries due to the rise of devotional traditions' where religious emphasis was placed less on

[9] A 2010 study showed that an estimated 67.6 % of Hindus were Vishnu followers (Johnson & Grim, 2013).
[10] Bengal is known in present day West Bengal and Bangladesh.

philosophical coherence, and instead emphasised general cultivation of *bhakti* (love, or devotional service) (R.M. Gupta, 2007, p. xi).[11] Transforming the landscape, Chaitanya Mahāprabhu's unique contribution was that 'emerging from a period of intense devotional activity in North India, yet grounded in the Vedantic philosophical tradition, the Caitanya tradition combined and transformed the nature of both' (R.M Gupta, 2007, p. 1). Consequently, the Gaudiya Vaishnava tradition today is one that heavily relies upon and incorporates sophisticated and coherent Vedic philosophy in order to emphasise the growth of bhakti in the heart, with the understanding that the coherence of vedantic philosophy culminates in this position of *love of God*. It is this *bhakti* or *love of God* which ultimately leads the soul to liberation and salvation, and therefore, we may naturally expect the spiritual tradition to contain a secure conception of freedom that underpins such a qualifying love.

First, though, I will list some of the core features of this theology that are essential to the development of an account of freedom. Firstly, being the tradition's conception of *Krishna* as a *Personal OmniGod*. In the *Bhagavad-Gītā*, Krishna's omnipotence is described as possessing '*ananta-vīrya*' meaning unlimited potency (Prabhupāda, 1983, 11.40), '*amita-vikramaḥ*' unlimited force (11.40), and having '*ananta-rūpa*' an unlimited form (11.38).[12] In the *Śrīmad Bhāgavatam* His omniscience is evident in descriptions of Him as *'sarva-jñah'* – the knower of everything (Prabhupāda, 1995, 10.16.59). The *Muṇḍaka Upanisad* also describes Him as '*Yaḥ sarva-jñaḥ sa sarva-vid yasya jñāna-mayaṁ tapaḥ',* meaning 'He who is all-knowing, from whom the potency of all knowledge comes — He is the wisest of all' (Prabhupāda, 1995, 10.87.2 purport). His omnipresence also, described as '*sarva-gatam'* meaning all-pervading, and '*tvayā tataṁ viśvam'* that 'this whole cosmic manifestation is pervaded by You' (Prabhupāda, 1983, 13.33 & 11.38). Krishna's omnibenevolence and fairness is also described as *bhūri-karuṇa* 'meaning the one whose compassion (*karuṇa*) is superabundant or great (*bhūri*)' (Gupta, 2020, p. 2). He's also understood as the 'self's well-wisher (*suhṛt*) and is equal to all living beings (*sama*)' (Gupta, 2020, p. 3). This conception of God is also extremely personal, not only in His ability to literally take on the body of a person

[11] G. Schweig offers these translations for 'bhakti' in his *Bhagavad Gītā Concordance: A Comprehensive Word Reference with English and Sanskrit Indexes* (2024), p. 335.

[12] Referencing for religious texts won't be given by page number, but instead by verse specification, e.g., canto 10, chapter 87, verse 2, will be shortened to 10.87.2. And in cases where there is no canto, such as the Bhagavad-Gītā, the numbers will refer to just the chapter and verse.

when he wishes , but also in Gaudiya Vaishnava literature, it is even described that if one is spiritually qualified, one can develop a relationship in particular moods, e.g. *vatsalya-rasa* is mood of worshipping God in a mood of parental affection, and *sakhya-rasa* being worship in the mood of friendship.

The second important aspect of Gaudiya Vaishnavism is the philosophy that we are immaterial souls which do not die when the body dies, but are a distinct self that is different from the body in which it is temporarily contained. Though most Abrahamic notions of the soul similarly refer to an immaterial self of spiritual substance, that is as far as the similarity goes. Whereas most Abrahamic faiths hold the understanding that the souls were at one time *created* by God, the dharmic conception of the soul is much different in its belief of *eternality*. In the *Bhagavad-Gītā*, Krishna explains the point that for the soul (known as the *jīva* or *ātman* in Vedic literature), '*Never was there a time when I did not exist, nor you, nor all these kings; nor in the future shall any of us cease to be'* (Prabhupāda, 1983, 2.12). So whilst the body that the soul inhabits is prone to eventual death, Krishna emphasises that '*that which pervades the entire body you should know to be indestructible. No one is able to destroy that imperishable soul*' (2.17) and that '*this individual soul is unbreakable and insoluble, and can be neither burned nor dried. He is everlasting, present everywhere, unchangeable, immovable and eternally the same*' (2.24). And importantly for our later discussion on sourcehood, the soul was never created: '*For the soul there is neither birth nor death at any time. He has not come into being, does not come into being, and will not come into being. He is unborn, eternal, ever-existing and primeval. He is not slain when the body is slain*' (2.20).

This conception of the jīva (soul) is unified across Hinduism, however, the philosophy of the relationship between the jīva and God is a largely contested topic. Some schools of Vedanta hold a monist view (Advaita Vedānta) that the individual soul and God – or the Absolute Brahman – is non-different from God/Brahman, meaning they are ultimately one and the same. Other schools hold that God and the individual jīvas are completely and eternally separate (Dvaita Vedānta). The philosophy of Gaudiya Vaishnavas, though, is known as *Achintya Bhedābheda Tattva* – the inconceivable one-ness and difference: literally difference (*bheda*) and non-difference (*ābheda*). Gaudiya theologians posit that whilst the individual souls are *qualitatively one* with God, in that we are plenary portions *of* God, we are *quantitatively different* in that we are finite, limited beings – compared to Krishna who is

infinite and unlimited. So whilst the Abrahamic conception claims that God brought about the soul's existence from non-existence, the Gaudiya understanding is that the jīvas have *always* existed as extended *parts* of God Himself – though we maintain a difference between us in our inherent *less-than-ness* and finiteness compared to God who is infinite in power, benevolence, knowledge, etc.

If it is true that we are, in a certain sense, *one* with God, then why does there appear to exist a degree of separation between the jīvas and God? If God is a personal being, then what is the cause for our lack of relationship? The answer that most Gaudiya Vaishnavas hold (though not all) is a conception of a fall doctrine.[13] The understanding is that by nature of being finite – not all-powerful or all-knowing – there is an inherent *risk* of deviation from an originally pure state, a risk that due to our inherent fallibility – we possess the potential to make mistakes. And due to this potential within the jīva, whilst originally they were together with God, at some point, *some* of the jīvas are victims of their own imperfection and independence, and reject God – this leading to a transition from living in the spiritual world, a realm outside of time and space, into living in the material realm of time and space. So, though the souls are *eternal*, it is *not* the case that they have existed within this 'material' realm for eternity – this point will become especially relevant to my later discussion of an agent being an adequate *source* and *initiator* of their actions.

Once fallen to the material world, one is subject to the cycle s*aṃsāra*, another import aspect of Gaudiya Vaishnava philosophy. 'Saṃsāra', translating to '*wandering from one existence into another',* refers to the continual cycle of reincarnation according to one's karma. *Reincarnation* being the notion that one does not die when the mortal body dies, but rather the soul transmigrates into a new body at the time of death. As Krishna propounds in the Bhagavad-Gītā, 'one who has taken his birth is sure to die, and after death one is sure to take birth again' (Prabhupāda, 1983, 2.27). The process is described that 'as the embodied soul continuously passes, in this body, from boyhood to youth to old age, the soul similarly passes into another body at death. A sober person is not bewildered by such a change' (2.13), with the body also being compared to an external dress: 'as a person puts on new

[13] It is an on-going debate within the Gaudiya Vaishnava tradition whether we were originally with God, or whether we were originally in another in-between state, or whether we have existed eternally in this realm. The predominant conception is the first conception, and is the one accepted for my GV view.

garments, giving up old ones, the soul similarly accepts new material bodies, giving up the old and useless ones' (2.22).

This cycle of rebirth goes about according to the agent's accumulated *karma*. The functions and abilities of karma, though, are not universally agreed upon across the dharmic religions, including the variety of Hindu traditions. At its most general and basic core, however, is the understanding of karma possessing the power of retribution – a law governing the world that plays a causal role in one's future experiences and circumstances: 'the basic idea of the theory of karma is this: besides its purely physical effects, our actions have also moral and psychological effects, which might occur either in this life or in future ones, and which affect our environment, genetic make-up, physical characteristics, social status at birth, length of life, etc., as well as our psychological dispositions and tendencies to act, desire, etc' (Silvestre, 2017, p. 38).[14] In the Vedic literature, it is accepted that 'the Self takes on a body with desires, attachments, and delusions, and is born again and again in new bodies to work out the karma of former lives. The embodied self assumes many forms, heavy or light, according to its needs for growth and the deeds of previous lives. This evolution is a divine law' (Easwaran, 2007, Śvetāśvatara Upaniṣad 5.11-12).

Related to this notion of *karma* are the *guṇas* – the modes of material nature. H.D. Goswami explains that similarly as to how 'human color vision is trichromatic, based on three primary colors, so worldly life exists within a tri-modal system of goodness, passion and darkness' (Goswami, 2015, p. 12). Our performance of actions in this world are usually tied to a particular mode or a blend of them: for example, giving gifts in the mood of kindness without selfish intent is the mode of goodness (*sattva*), giving with a desire for something in return is in the mode of passion (*rajas*), and giving gifts with the intention of offense of harm is the mode of darkness (*tamas).* Importantly, especially to our later discussion on determinism, these guṇas play an influencing role in the future actions we are inclined to perform: as Goswami writes: 'Good, passionate and dark acts are all habit-forming. As we choose friends, places, music, food, work, etc. in various modes, we give those modes power over our lives… Virtuous acts beget virtue; passion engenders passion;

[14] Considering the great variation on views of *karma* that exists, Silvestre writes that 'one cannot speak of the theory of karma outside a specific Indian religious tradition. You have to pick up one tradition or a specific set of texts and see what it says about the doctrine; this will give you the only thing you can have: *a* theory of karma' (Silvestre, 2017, p. 37).

and dark behavior such as addiction, wanton violence, etc. traps one in darkness' (Goswami, 2015, p. 14). It is because of this that *saṃsāra* appears cyclical, due to the difficultly in breaking free from the effects of the modes of nature that bind us to future actions. However, one type of action that is *not* bound by the guṇas are *spiritual activities* – guṇa-atīta. By performance of these types of activities, the cycle of s*aṃsāra* can end for the individual.

Although this cycle of s*aṃsāra* is often described as a prison house, liberation from this repeated birth and death is possible and available to all souls. According to Gaudiya Vaishnava philosophy, the highest form of salvation is the return to Krishna in the spiritual realm, and in order to do so, one must give up their material attachments and develop *bhakti* (love for God). By performing '*bhakti-yoga*, one offers, and thus spiritualizes, all of one's life, clearing one's path to highest liberation in Kṛṣṇa' (Goswami, 2015, p. 123). As Goswami explains, 'after describing the character and consciousness of a transcendent soul, Kṛṣṇa concludes that by serving God with undeviating *bhakti-yoga*, one fully transcends nature's qualities and achieves spiritual existence.' (2015, p. 116). Naturally, we may infer that to be so qualified as to receive salvation, one must be adequately free to choose that path, for there is some sense in which qualification here involves *agent responsibility* in order to be deserving or a fine candidate. And in a similar sense, *karma* also tends to indicate to, or be founded upon, a belief of *freedom* and ability for an agent to be *morally responsible* – otherwise it appears that the moral retributive function of karma would cease to be fair. Affirming free will in the presence of karma, though, appears a problematic task despite the law appearing to require it – this will be touched on in the next section.

The Presentation of *Agency* in Gaudiya Vaishnavism

Woven throughout the discourses on these concepts of the soul, karma, and rebirth in Vedic literature are the related descriptions of control and agency. In fact, understanding the soul's agential ability is presented as an essential aspect of spiritual advancement.[15] Yet its paradoxical presentation is easily disorientating. Let me begin by including here a few examples from the *Bhagavad-Gītā* and *Śrīmad Bhāgavatam* that appear to infer a type of

[15] In the Bhagavad-Gītā, God explains: 'A seer who perceives no other doer but modes, and sees what is above modes, attains My state of being.' (Goswami, 2015, 14.19).

causal determinism operating in the world, which I will follow with evidence implying the existence of free will.

Remembering the previous description of *guṇas* – the modes of material nature, observe here how early on in the *Bhagavad-Gītā* Krishna explains that 'everyone is forced to act helplessly according to the qualities he has acquired from the modes of material nature; therefore no one can refrain from doing something, not even for a moment' (Prabhupāda, 1983, 3.5). Similarly in chapter thirteen, Krishna exclaims that 'actions are being fully done by nature alone. One sees, who thus sees self as non-doer' (Goswami, 2015, 13.30). In the *Śrīmad Bhāgavatam*, the same freedom-undermining philosophy is woven throughout, with statements such as 'an unintelligent person situated within the body created by his previous fruitive activities thinks, "I am the performer of action." Bewildered by false ego, such a foolish person is therefore bound up by fruitive activities, which are in fact carried out by the modes of nature' (Prabhupāda, 1995, 11.11.10). Though we may believe that we are deliberating and deciding our own actions, actually, these actions are being fully controlled by the modes of nature that we have acquired previously – and these mental states/moods/desires have conditioned our consciousness so strongly that their influence can force us to perform actions – so much so that Krishna urges us to not even consider ourselves the doer of our actions, despite our impression that we are a deciding doer in our everyday lives. Repeatedly, statements like these are propounded which give off the impression that at least *some* of the actions that we believe we've performed freely, aren't truly free.

Not just due to the acquired modes, though, freedom also appears to be undermined due to the presence of God. There are numerous statements regarding the inability to go against fate and providence, e.g. 'but still I never lament or rejoice, because I know this world is driven by time and fate' (Prabhupāda, 1995, 10.54.14). Even in a more *causal* sense, it is stated in the *Bhagavad-Gītā* that 'the Lord of all beings resides in the region of the heart, Arjuna, making all beings wander as they ride in illusion on the [bodily] machine' (Goswami, 2015, 18.61). Indeed, elsewhere, Krishna urges Arjuna (the main protagonist of the Bhagavad-Gītā) to fight in the forthcoming battle because in a significant sense, Krishna has already slain the opposing fighters due to His providential plan: 'therefore stand up! Gain glory conquering the enemies! Rule a flourishing kingdom! I alone have already slain them; you, Savya-sācī, be but an instrument!... Slay those slain by Me: Droṇa, Bhīṣma,

Jayadratha, Karṇa and still other battle heroes. Do not waver! Fight! You will conquer the foes in battle.' (Goswami, 2015, 11.33-34).

Paradoxically, in other sections of the religious literature, however, freedom of the individual agents is repeatedly affirmed to exist. In chapter six of the Bhagavad-Gītā, Krishna preaches to Arjuna that 'one should uplift self by self; one should not degrade self. Indeed self alone is self's friend; self alone is self's foe… Self is friend to that self by whom self alone is conquered; but that very self can work as enemy, due to non-self's enmity… For the peaceful who conquer self, Higher Self… is fully established in cold and heat, joy and sorrow, and so too in honor and dishonor', indirectly suggesting that the agent self *does* have some degree of control over the influences that may be influencing the self – an ability that is *up to us* whether to degrade or uplift (Goswami, 2015, 6.5-7). Similarly, in the *Śrīmad Bhāgavatam* it is stated that 'although the intelligence of a learned person may be bewildered by the modes of passion and ignorance, he should again carefully bring the mind under control. By clearly seeing the contamination of the modes of nature, he does not become attached' (Prabhupāda, 1995, 11.13.12). Piecing the philosophy together, it seems that whilst we shouldn't consider ourselves the doer of activities that are actually brought about by the modes of material nature, we do retain strength to overcome the bewildering modes.

Still though, it would be fair to wonder if the reason for the paradoxical nature of the statements is due to the statements being taken out of context. To eliminate this doubt, though, let me include here one of the final instructions that Krishna gave in the *Bhagavad-Gītā*: 'Relying on egotism, if you think, "I shall not fight," that decision is wrong. Nature will impel you… Kaunteya, you are bound by your work born of your nature. That which, from illusion, you do not wish to do, you will do even against your wish… The Lord of all beings resides in the region of the heart, Arjuna, making all beings wander as they ride in illusion on the [bodily] machine… Thus I have taught you knowledge more secret than the secret. Reflecting on this fully, do as you wish.' (Goswami, 2015, 18.59-61 & 18.63). Despite nature and providence *ensuring* that Arjuna will fight, Krishna still tells him to 'do as you wish' after reflecting – indicating that though there is the existence of some causal or theological determinism, this does not undermine Arjuna's free will *to do as he wishes*. Importantly also, the word 'reflecting' translates from the Sanskrit '*vimṛśya*' which others have translated to *deliberate* and *consider* – this point will play an important role in my later analysis of Fischer

and Ravizza's *reasons-responsiveness* approach to freedom which ascribes moral responsibility not according to an ability to do otherwise, but according to the agent's ability to reflect and deliberate upon moral reasons.[16]

It's becoming, perhaps, increasingly clear here that the sense of freedom described in Vaishnava scripture does not indicate a strict sense of *ability to do otherwise*, but rather an ability of reason and deliberation. In the following chapter of this work, a closer inquiry into what we mean by an agent possessing 'free will' will be offered. But before moving onto that, let me first finish this section with a few more words on how *Gaudiya Vaishnava* theologians have interpretated this apparent paradox of *doership*.

As explained above, Vaishnava scripture encourages us to give up the idea of ourselves being doers and reinforces the message that we are not the ones performing the actions, that instead, the *guṇas* are. At the same time, individual responsibility plays a vital role in the theology, it appears essential to the function and fairness of karma, and also for the cultivation of *bhakti* – essential for spiritual advancement. This *is* a topic that Gaudiya theologians have addressed: Baladeva Vidyābhūṣaṇa (1720-1790) addresses this topic in his *Govinda Bhasya* commentary on the *Brahma Sūtras*, and he explains the soul's non-doership in terms of the soul being distinct from the body. When we identify with the senses, beliefs, and desires that fill our consciousness, we therefore think of ourselves as the controllers of our actions; this is foolish, however, because actually it is us and our actions that are under control of the modes of nature that have infiltrated our mind (Vidyābhūṣaṇa, 1718/2018, p. 262-266). In this sense, we should understand that we are actually the soul – which is a *separate entity* to the senses and actions performed by the body and mind under the influence of the guṇas: 'This explains why there are statements in scripture calling the guṇas the agent. One who sees the jīva as the sole agent is called a fool (BG 3.27) because, though action depends on five factors, the person thinks of himself as the only factor.' (Vidyābhūṣaṇa, 1718/2018, p. 264).[17] How we are the *doers*, by contrast, is presented by two metaphors. Firstly, Vidyābhūṣaṇa references the *Bṛhad-āraṇyaka Upaniṣad 2.1.18* which states that: 'Like a king moves with his subjects in his kingdom, the ātmā , taking the prāṇas

[16] In Goswami's translation he translates the word *vimṛśya* as 'reflecting' (2015). In Prabhupāda's translation he translates it as 'deliberating.' In the Monier-Williams Sanskrit dictionary *vimṛśya* is defined as 'having deliberated or considered.'

[17] The five factors that action depends on are: 'place, doer, diverse means, diverse efforts—and, of course, Providence' (Goswami, 2015, 18.14).

with him, moves about in his body', indicating that though there exists a difference between the soul (ātmā) and the senses (prāṇas), the soul is the doer because it is the jiva that is 'the grasper of the prāṇa'. (Vidyābhūṣaṇa, 1718/2018, p. 262).

Explaining further, following the example given in the *Brahma Sutra 2.3.38* of a jiva being compared to a carpenter, Vidyābhūṣaṇa writes that the 'The carpenter in doing his carpentry is an agent in two ways-by using his own śakti in holding his axe and through the instruments like the axe. The jīva is the agent by his own śakti in grasping the prāṇas and by using the prāṇas to grasp other things. Though the action originates from the pure jīva, it is said figuratively that the material body is the agent because of the strong influence of the guṇas.' (Vidyābhūṣaṇa, 1718/2018, p. 264-265).[18] In this sense, the scriptural contradiction of doership is avoided by delineating the soul's non-doership due to being a separate entity distinct from the body and its actions, but identifying the soul's doership role in being the *source* of the influencing prāṇas. This concept of being the *source* of one's actions will play a significant role later in analysing what we mean by an agent possessing free will, alongside my previous hint regarding *reflecting* and *deliberating.*

The contradiction of doership, though, still needs to be addressed regarding the controlling aspect of God that was previously mentioned: remembering back to *Bhagavad-Gītā*, Krishna says that 'the Lord of all beings resides in the region of the heart, Arjuna, making all beings wander as they ride in illusion on the [bodily] machine' (Goswami, 2015, 18.61). Addressing this philosophical point, in the *Brahma Sūtras* 2.3.40, it is described that '…the Lord acts after considering the jīva's efforts, the injunctions and prohibitions are not useless and the Lord is not liable to the charge of cruelty' (Vidyābhūṣaṇa, 1718/2018, 2.3.40). So whilst it does appear that God can play a causally determining role in our lives, *we* as agents play a causal role in how God responds to our actions and efforts. This point, though, will be addressed more thoroughly in chapter five.

The Central Paradox

What isn't entirely clear yet is if the world according to Gaudiya Vaishnava scripture is wholly deterministic. However, I don't believe that it's clear that it's relevant at all anyway. Rather the point that is relevant, is that there does seem to be an ambiguity within the literature.

[18] The Sanskrit term 'śakti' used here translates to *power, ability, energy, capability*.

Take the case of Krishna in chapter eighteen telling Arjuna what the future will be, yet still encouraging Arjuna to deliberate and decide on the future. On one hand, we could assert that this prophetic statement indicates that the future is fixed and determined, however, it could be that God is simply speaking from His omniscience of knowing *x* will occur, which plays no *causal* role in Arjuna actually bringing about *x*. On the other hand, Krishna tells Arjuna that Arjuna's nature will impel him to bring about *x* (Bhagavad-Gītā, 18.59), and that there is nobody free from these nature-born guṇas (Bhagavad-Gītā, 18.40). From this, we could judge it thus indicates that all our 'free' decisions can be traced back to acquired guṇas that play a causally determining role in how we act.

In this thesis, I will be presenting a compatibilist view, however, my analysis of the success of a GV compatibilist approach will *not* be in how it compares to non-compatibilist approaches, but simply in the success of its internal coherence. In concluding this brief literature review of the scriptural presentation of agency, another important *relevant* point that I'll make is that when agency or free will is alluded to within the Gaudiya Vaishnava literature, it's usually *not* in terms of *an ability to do otherwise*, but instead, responsibility is mentioned in relation to *sourcehood* and engaging with the intellect's ability of *deliberation and reflection* – it is this conception of 'free will' that strikes many similarities with modern compatibilist approaches to defining free will – which will be centre stage in my next chapter.

Following this analysis, there are two main questions that need to be tackled: 1. How exactly is causal determinism compatible with free will? And 2. Is the doctrine of *karma* compatible with the answer to 1? Similar to Kane's *incompatibilist mountain*, there appears to exist also a dilemma with karma – for the GV view to be successful, karma can't just be compatible with determinism *or* free will, but it has to work with *both*.

Chapter Three

Responsibility and Manipulation

In this chapter, the reason behind the introduction of Gaudiya Vaishnava theology in chapter two will become clear following a return to the modern discussions of free will to examine the many strengths, but more pertinently, the *limitations* that popular branches of compatibilism possess. Specifically, what will be addressed is the struggle that compatibilist theories of free will face in overcoming objections that take form in examples of manipulation. Although at first glance these theories may appear to naturally correspond with our intuitions and perspectives, it becomes evidently difficult for compatibilist proponents to provide an adequate account of sourcehood that justifies responsibility on some cases, whilst still able to rule out responsibility on occasions where an agent is manipulated into deliberating and performing actions. The manipulation objections that will be put forth in this chapter aren't designed as examples of possible situations where the formula for identifying free will fails to produce the correct answer which in turn renders the conception useless *only* on rare and hypothetical scenarios, but instead, they are scenarios which give compelling evidence that the formula is incoherent and unable to identify responsibility on *any* occasion – these objections are stubbornly undermining and should be treated as such. In chapter four I will present a new approach to tackling charges from manipulation stemming from *Gaudiya Vaishnava* philosophy, but for now I shall illustrate the complexity of the challenge facing compatibilists with an overview of modern theories.

Compatibilist and Semi-Compatibilist Approaches

Prior to our discussion of agency in the Gaudiya Vaishnava tradition, we finished chapter one noting a divergence among philosophers away from defining free will and moral responsibility in terms of an *ability to do otherwise*, and rather on an agent being an adequate *source* of one's actions to necessitate possession of responsibility – this change largely owing to Frankfurt's presentation of counterexample scenarios where an agent appears to retain responsibility for an action that they could not stray from performing.

Instead of the *Principle of Alternate Possibilities*, Frankfurt suggested that what free will actually consisted in and required was that our hierarchical desires and volitions meshed

together: in simple terms, Timpe describes Frankfurt's theory simply that 'freedom of the will is being able to have the kind of will that one wants to have' (2008, p. 76). On Frankfurt's account, our desires can be grouped into two categories: first-order desires being 'simply desires to do or not to do one thing or another', whereas second-order desires are our underlying motivations desires for our desires – our desires of being different or having different urges, preferences, or motivations than what we actually do which is a 'capacity for reflective self-evaluation' that separates mankind from animals (Frankfurt, 1988, p. 12). Following this distinction, what matters for free will according to Frankfurt is that 'if someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then – so far as his moral responsibility for doing it is concerned – it really doesn't matter how he got that way' (Frankfurt, 2002 p. 27).

Frankfurt's presentation is appealing to me for two reasons. Firstly, although determinism might be true, it feels correct in positing that so long as the agent genuinely and wholeheartedly identifies with and desires to perform the action, then they should be held responsible for it, despite not being able to do otherwise. Secondly, on another intuitive level it makes sense with regard to how we sympathise with those whose first and second order desires don't mesh – take for example somebody addicted to smoking, on the first-level, they have the desire to smoke, but this perhaps does not mesh with their second-level desire to quit smoking and not possess the desire to smoke at all. This conception of hierarchical desires does seem to capture nicely why we consider somebody who suffers addiction is suffering, in some sense, a lack of sufficient freedom of will.

This presentation of free will and responsibility has been very influential, spanning out into literature developing different theories regarding the 'meshing' of desires. G. Watson, for example, attempts to account for freedom and responsibility in terms of 'a kind of harmony or mesh between different elements in the mental economy of persons qua agents' (McKenna & Coates, 2024a). Whereas S. Wolf's conception of meshing 'require[s] an anchor external to the agent's internal psychological states (the True and the Good)' (McKenna & Coates, 2024b). Some philosophers, however, express their doubts over the success of Frankfurt's conception, not just from incompatibilists in their attempt to reaffirm the *Principle of Alternate Possibilities*, but also from his fellow compatibilists that believe Frankfurt's account simply isn't comprehensive enough. The

most popular argument that I will raise here is the doubt that a Frankfurt conception of free will can withstand an objection from manipulation. In Timpe's 2008 discussion on Frankfurt, he offers the following simple responsibility-undermining manipulation example:

> Consider a case involving Allison and her manipulative husband, Hal, who is a master hypnotist. Suppose that Hal hypnotizes her and that, as a result, Allison finds herself with the desire to make Hal his favorite dessert every night. However, this desire is implanted in her psychological hierarchy by her husband via hypnosis. Furthermore, Hal also causes Allison to have the second-order desire to have her first-order desire to bake become her will. And in case Allison has any conflicting first-order desire not to bake, as part of his plan Hal also causes her to wholeheartedly identify with those other desires that he has implanted in Allison's psychological hierarchy. It looks then that Allison has been manipulated into identifying with a particular second-order desire. In such an instance, even though Allison's desire to make dessert for her husband meshes with her volitional structure, it does not originate in that volitional structure since the mesh is solely the result of external manipulation. (Timpe, 2008, p. 77)

If we are to accept Frankfurt's theory of free will, then we would be forced to accept that Allison acted freely and responsibly, despite being so obviously manipulated. This, however, is a bullet Frankfurt is willing to bite – let me include here again the same quote I referenced a few paragraphs ago: Frankfurt writes that 'if someone does something because he wants to do it, and if he has no reservations about that desire but is wholeheartedly behind it, then – so far as his moral responsibility for doing it is concerned – it really doesn't matter how he got that way' (Frankfurt, 2002 p. 27).

Though Frankfurt may be willing to accept such a position, it is certainly not a satisfactory enough account of freedom to convince others – one such critic of Frankfurt's theory being J.M. Fischer, who, writing in response to the manipulation objection, wrote once that 'Harry Frankfurt really has not addressed this problem. He has discussed it in different ways and in different places and it doesn't add up to much', whereas Fischer's own account instead, he argues, 'has some promise' (2000, p. 390 & 391). Although I touched briefly on the compatibilist case of Fischer and Ravizza in chapter one, let me elaborate further now and discuss its supposed 'promise'.

As I described previously, technically, the position of Fischer and Ravizza is not a strict *compatibilist* view, but a view they call *semi-compatibilist*; the cause for this distinction being the agnosticism provided by their requirement for moral responsibility that does not require them to take a stance on the truth of determinism – the stance *is semi-compatibilist* in the sense that their presentation of what constitutes responsibility is compatible with *both* the truth or falsity of determinism. Following from Frankfurt's rejection of alternate possibilities as a requirement for free will, Fischer and Ravizza accept this move, but instead of developing a theory of hierarchical meshing, Fischer and Ravizza turn to a conception based on what they call *guidance control*. As I briefly mentioned in chapter one, Fischer and Ravizza argue that whilst we can remain agnostic as to whether individuals possess what they call *regulative control* – a control over which future of genuinely open possibilities becomes actual, individuals certainly have the potential to hold *guidance control* – an ability to guide our behaviour, reflect upon reasons, and deliberate on decisions that make up our future actions. Essential to this *guidance control* is, firstly, that the agent's behaviour must issue from the agent's own mechanism, and secondly, *guidance control* requires the presence of *reasons-responsiveness*: '[r]easons-responsiveness locates the control required for morally responsible agency in the capacity to respond to reasons for action. An agent is free in the required sense just in case her behaviour stems from elements of her psychological constitution capable of conforming, to some degree, with the vagaries of reasons for action.' (McKenna, 2000, p. 89). Or more simplistically put, reasons-responsiveness requires that in order for an agent to be responsible, 'the agent's action in the actual sequence must be intentional, that is, appropriately connected to his reasons.' (Fischer & Ravizza, 1998, p. 81).

Developing their theory of free will, responsibility, and determinism, in their 1998 work *Responsibility and Control*, Fischer and Ravizza put forward three ingredients that must be present for responsibility to ensue. The first criteria Fischer and Ravizza require is that 'an individual must see himself as the source of his behavior… [t]hat is, the individual must see himself as an agent; he must see that his choices and actions are efficacious in the world.' (1998, p. 210). Secondly, 'the individual must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts… the individual need not have any sort of theory that explains which contexts are the ones in which he is fairly praised or blamed; he simply needs to believe that he is an apt candidate for the reactive

attitudes in certain contexts, and that it is not an arbitrary matter what those contexts are' (1998, p. 211).[19] And lastly, '[t]he third condition on taking responsibility requires that the individual's view of himself specified in the first two conditions be based, in an appropriate way, on the evidence' (1998, p. 213). Timpe nicely summarises this condition of responsibility in his words as 'a person chooses freely only if he chooses as he does (i) because of an appropriately reasons-responsive mechanism and (ii) he sees that mechanism as his own in an appropriate way' (2008, p. 80).

Following this conception of moral responsibility, on the example previously given of Hal who hypnotises his wife Allison, Allison would fail to be morally responsible as the behaviour issues from Hal's act of hypnotism and *not* from a conscious reasons-responsive mechanism that belongs to Allison. Although Allison performs the action, it was not performed via an appropriately responsive-responsive mechanism that is appropriately her own, and therefore, according to this Fischer and Ravizza conception, Allison can be deemed not morally responsible.

Although one could certainly argue that the conception of responsibility that Fischer and Ravizza presents is more detailed and comprehensive than Frankfurt's theory, it isn't clear at all that they are able to escape the same objection of manipulation scenarios that Frankfurt's theory suffered earlier. The example case of Allison being hypnotised can easily be altered so that instead of Frankfurt's hierarchical meshing of desires, Hal manipulates Allison such that she possesses a reasons-responsive mechanism and view of herself as target for reactive attitudes – meeting the requirements for guidance control and thus, responsibility too. Thus just as Frankfurt's conception of freedom is forced to deem Allison as responsible, Fischer and Ravizza's account should also hold Allison responsible.

As E. Stump contends in response to a similar manipulation example involving electronical manipulation, 'no matter how we specify the way in which a person's beliefs about herself must be based on her evidence, a person can meet these very specifications in virtue of having electronically [or hypnotically] induced in her a certain (no doubt very complicated) neural state or set of states… Consequently, [Allison] can meet the Fischer and Ravizza conditions for acting on a mechanism which is her own when she is so controlled by

[19] As described in chapter one, reactive attitudes can be understood as emotional responses, such as resentment and gratitude, that presuppose agent responsibility.

a manipulator that her very formation of her beliefs about her own agency on the basis of her evidence is electronically produced in her by the manipulator' (Stump, 2002, 53-54). It appears obvious that this conception of responsibility requirements issues a false reading of Allison being morally responsible.

Key to minimising the threat of manipulation cases, however, according to Fischer and Ravizza, is their third condition I mentioned of an agent's view of himself being formed upon evidence in an appropriate way. In a case like Allison being hypnotised by Hal, Fischer and Ravizza might argue along the lines that Allison doesn't fall victim of qualifying as responsible because she doesn't meet the third standard of forming the previous two criteria in *an appropriate way*. Naturally, now we must enquire – what is meant by 'an *appropriate way*'? Fischer and Ravizza themselves acknowledge that this appears 'abstract and schematic', but still hold that '[t]his condition is intended (in part) to imply that an individual who has been electronically induced [or hypnotically induced in our case] to have the relevant view of himself (and thus satisfy the first two conditions on taking responsibility) has *not* formed his view of himself in the appropriate way' (Fischer & Ravizza, 1998, p. 236). However, this simply won't suffice. Without a further definition of what Fischer and Ravizza mean by '*appropriate'*, Timpe argues that their third requirement for responsibility amounts to essentially 'a mere stipulation that the agent cannot be manipulated' (2008, p. 83). And what's insufficient about this mere stipulation is the inability in many occasions to draw the line where somebody is being manipulated, coerced, compelled, etc, and when they're not. The underlying belief behind the intuition that Allison is not responsible for her actions when being manipulated by her husband Hal, is the understanding that Allison shouldn't be held responsible for actions that she wasn't a genuine source of – although it appears her actions were products of a reasons-responsive process on her behalf, what initiated this process was external to her – her husband Hal.

In response to this theory of Fischer and Ravizza, Pereboom puts forward an impressive critique which demonstrates how there's *never* an ability to render somebody responsible because an agent's actions will always have a causal influence external to the agent that is outside of their control. By arguing so, it also renders the third condition of Fischer and Ravizza's theory as useless – if all actions can be proven such that they come about as a result of external factors (like Hal hypnotising Allison), then no action could have formed in an '*appropriate way*' required for Fischer and Ravizza – thus we lack a method of

identifying people as responsible. This critique from Pereboom takes its form in a convincing argument that attempts to demonstrate 'not only that compatibilist accounts of sourcehood fail for reasons of manipulation but also that the truth of causal determinism would threaten free will and moral responsibility for reasons of the sort that are associated with manipulation' (Timpe, 2008, p. 86). And of course, for my later purpose here of developing a *Gaudiya Vaishnava* compatibilist account of sourcehood, we must first address the charge of Pereboom's argument.

Pereboom's Four-Case Manipulation Argument

Incompatibilists have a history of employing manipulation-based arguments to undermine compatibilist theories of free will moral responsibility, but it is Derk Pereboom's 2001 version that has the most impressive force and garnered the most attention in contemporary discussions. And therefore, it is certainly an argument than any compatibilist theorist must provide an answer for to prove the soundness of their position. In Pereboom's argument, he begins by setting out four very similar hypothetical scenarios:

> **Case 1.** Professor Plum was created by neuroscientists, who can manipulate him directly through the use of radio-like technology, but he is as much like an ordinary human being as is possible, given this history. Suppose these neuroscientists "locally" manipulate him to undertake the process of reasoning by which his desires are brought about and modified – directly producing his every state from moment to moment. The neuroscientists manipulate him by, among other things, pushing a series of buttons just before he begins to reason about his situation, thereby causing his reasoning process to be rationally egoistic. Plum is not constrained to act in the sense that he does not act because of an irresistible desire – the neuroscientists do not provide him with an irresistible desire – and he does not think and act contrary to character since he is often manipulated to be rationally egoistic. His effective first-order desire to kill Ms. White conforms to his second-order desires. Plum's reasoning process exemplifies the various components of moderate reasons-responsiveness. He is receptive to the relevant pattern of reasons, and his reasoning process would have resulted in different choices in some situations in which the egoistic reasons were otherwise. At the same time, he is not exclusively rationally egoistic since he will typically regulate his behavior by moral reasons when the egoistic reasons are relatively weak – weaker than they are in the current situation.
>
> **Case 2**. Plum is like an ordinary human being, except that he was created by neuroscientists, who, although they cannot control him directly, have programmed him to weigh

reasons for action so that he is often but not exclusively rationally egoistic, with the result that in the circumstances in which he now finds himself, he is causally determined to undertake the moderately reasons-responsive process and to possess the set of first– and second– order desires that results in his killing Ms. White. He has the general ability to regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and accordingly he is causally determined to kill for these reasons. Nevertheless, he does not act because of an irresistible desire.

Case 3. Plum is an ordinary human being, except that he was determined by the rigorous training practices of his home and community so that he is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1 and 2). His training took place at too early an age for him to have had the ability to prevent or alter the practices that determined his character. In his current circumstances, Plum is thereby caused to undertake the moderately reasons-responsive process and to possess the first– and second– order desires that result in his killing White. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances, the egoistic reasons are very powerful, and hence the rigorous training practices of his upbringing deterministically result in his act of murder. Nevertheless, he does not act because of an irresistible desire.

Case 4. Physicalist determinism is true, and Plum is an ordinary human being, generated and raised under normal circumstances, who is often but not exclusively rationally egoistic (exactly as egoistic as in Cases 1– 3). Plum's killing of White comes about as a result of his undertaking the moderately reasons-responsive process of deliberation, he exhibits the specified organization of first– and second– order desires, and he does not act because of an irresistible desire. He has the general ability to grasp, apply, and regulate his behavior by moral reasons, but in these circumstances the egoistic reasons are very powerful, and together with background circumstances they deterministically result in his act of murder. (Pereboom, 2001).

The challenge to compatibilists, as proposed by Pereboom, is to identify the *relevant difference* between cases 1-2-3-4 that draws the line where Plum is *not* morally responsible for killing Ms. White to when he is morally responsible for the killing, and according to Pereboom: 'If I am right, it will turn out that no relevant difference can be found among these cases that would justify denying responsibility under covert manipulation while affirming it in ordinary deterministic circumstances, and this would force an incompatibilist conclusion' (Pereboom, 2001, p. 112).

At first glance, this challenge to find a *relevant difference* amongst a series of cases appears to operate as a Sorites Paradox – a type of conundrum that arises when trying to analyse at what point the identity or aspect of a thing changes. The classical example is of a heap of sand: if we remove one grain of sand from the heap, then, because the difference is so minor, we would still consider it a heap; but if we follow this pattern of always retaining our previous judgement due to the apparent insignificance of the change, then we would end up with our 'heap' being just a single grain of sand. Of course, this conclusion is false; at one end we have one grain and at the other we have a heap. The paradox is the confusion over at what point the few grains becomes a heap – where do we draw the line? What is the *relevant difference*? It could easily be imagined that an immediate response to Pereboom's argument could be to argue that Pereboom is employing a Sorites Paradox to push for his conclusion: one may argue along the lines that although we may not be able to identify the exact point where the grains become a heap, or where Plum goes from not morally responsible (on case one) *to* morally responsible (on case four), that does *not* mean that there is *not* a definite difference between one grain of sand and a heap, or a definite difference between Plum on case one compared to case four.

According to Pereboom, however, this response would be incorrect. Although it does urge the compatibilist to identify a relevant difference between cases, what drives Pereboom's conclusion forward is *not* that the difference between them is so minor and incremental, it is that there is *no such difference between them* all as far as agency and moral responsibility is concerned. From the perspective of a compatibilist, it may appear as though there is an incremental change in the degree of responsibility that Plum holds on each case, however, Pereboom's argument is that that very perspective is incorrect – we should see all cases as equally lacking responsibility. As Pereboom writes, the strength of his argument lies in how 'between each successive pair of cases there is no divergence at all in factors that could plausibly make a difference for moral responsibility, and that we are therefore forced to conclude that all four cases exhibit the same kind and the same degree of an incompatibilist responsibility-undermining feature' (Pereboom, 2001, p. 116).

Instead of arguing along these lines, the heart of the issue for compatibilists, as Pereboom argues, is the compatibilist's inability to offer a satisfying account of sourcehood. When an agent is manipulated, he fails to be morally responsible (even if the agent acts according to their reasons-responsive mechanism or a Frankfurt-styled appropriate meshing

of desires) because the manipulation was outside of his control. Because Plum was not the *source* of the radio-like technology that the neuroscientists use to manipulate him, he could not then be reasonably held responsible as he didn't possess adequate control over what would cause him to kill Ms. White. Intuitively, this feels non-controversial. Pereboom's argument, though, is that this failure to be an adequate source of one's action (that most compatibilist accounts take to be essential for moral responsibly in some shape or form) is evident on *any* case where determinism is regarded true. Timpe describes this argument as *Pereboom's Two-Step Strategy* where stage one is to 'argue that in cases of overt manipulation, the agent fails to be the proper source of his action in the way required for free will and moral responsibility' (the *manipulation undermining* claim), and stage two being to 'argue that causally deterministic cases are similar, with respect to the agent's failure to be the proper source of his actions, to cases of overt manipulation (the *similarity claim*) (Timpe, 2008, p. 89).

In case one, Plum's potential for moral responsibility is undermined by the factors beyond his control that determine his action, namely, the neuroscientists that create him and manipulate him directly through the use of radio-like technology, despite outwardly possessing some reasons-responsiveness process. In case two, Plum's potential for moral responsibility is undermined by the factors beyond his control that determine his action that take the form of the neuroscientists who, although they take a step back and no longer produce his every state moment to moment, have still programmed him to weigh reasons in an exact way that will produce their desired action. In case three, the causal factors outside of Plum's control have a much more realistic presence, where Plum has been conditioned by a rigorous upbringing which determines his decision, after undergoing a reasons-responsive process, to kill White. Employing the similarity claim, Pereboom points out that 'if the compatibilist wants to claim that Plum is morally responsible in Case 3, he must point out a feature of these circumstances that would explain why he is morally responsible there but not in Case 2' (Pereboom, 2001, p. 114). The compatibilist is challenged to explain the difference between the causal determinism coming from neuroscientists programming Plum and the causal determinism coming from parents and community programming Plum. Surely no relevant difference can be found, and thus in case four, where compatibilists most definitely wish to say that Plum *is* morally responsible, we are forced to concede that as there is no relevant difference between the determinism in case four, where Plum has simply

be programmed by the truth of physicalist determinism, and case three, case two, and case one, then Plum is not morally responsible and could never be morally responsible according to these compatibilist accounts.

The crux of the problem for compatibilists to defend themselves is to explain somehow how an agent can be held morally responsible despite there always existing a deterministic causal chain from an action that traces back to factors beyond the agent's control. Some compatibilists will take a *hard-line reply* and 'adopt the counter-intuitive position that the manipulated agent is responsible for his action, while accepting the no-difference thesis'; whilst other compatibilists will take a *soft-line reply* and 'accept the intuitive claim that the manipulated agent lacks responsibility for his action, and deny the no-difference thesis', and in the following section I will offer a few examples of such approaches (De Marco, 2023, p. 488).[20]

Replies to Pereboom

The approach taken by Fischer is a type of *soft-line reply* in how he argues *for* the existence of a difference between the cases, arguing that 'it is natural to suppose that Professor Plum is morally responsible for killing Ms. White, even if he is not blameworthy (or not fully blameworthy) for doing so' (2023, p. 233). But as Timpe points out: 'it isn't clear how an agent could be morally responsible for a morally wrong action (by virtue, in part, of exercising the appropriate kind of control over that action) and fail to be morally blameworthy for that action' (2008, p. 84). It is indeed possible for an agent to be morally responsible but not blameworthy, that is, when the agent's action is worthy of praise rather than blame, or if the action is morally neutral, but this is not the same sense that Fischer is describing. In a 2004 paper, Fischer writes that though he can accept that Professor Plum meets the threshold for being *responsible* on all four cases, he is only *blameworthy* on cases three and four, and *not* cases one and two (2004, p. 158). In a vague tone, Fischer attempts to make the distinction that to be blameworthy involves 'the circumstances under which one's values, beliefs, desires, and dispositions were created and sustained, one's physical and economic status, and so forth' (2004, p. 158). Yet this position is a puzzle; whilst it is

[20] Michael McKenna coined the terms *soft-line* and *hard-line replies* in his 'A Hard-line Reply to Pereboom's Four-Case Manipulation Argument' (2008), where he notes that the distinction is loosely based off the distinction between hard and soft compatibilism in Robert Kane's 'The Significance of Free Will' (1996).

possible to be morally responsible without being blameworthy, as I noted earlier, on this specific case of Professor Plum, Pereboom replies that 'for an agent to be morally responsible for an action in the sense at issue is for it to belong to him in such a way that he would deserve blame if he understood that it was morally wrong, and he would deserve credit or perhaps praise if he understood that it was morally exemplary…' and therefore on this case, Plum 'could not be morally responsible for committing this murder without also being blameworthy for it' (Pereboom, 2007, p. 99). At no point, however, does Fischer present an additional specific condition that is required for blameworthiness that can be identifiable to draw a line, instead positing that an agent could meet all the criteria for agency regarding a specific morally wrong action, yet somehow lack any blameworthiness for this action (Timpe, 2008, p. 85). Understandably, it appears reasonable to side with Pereboom's judgement that Fischer's account of moral responsibility remains undermined by his manipulation argument.

In a different approach, M. Mckenna objects to the force of Pereboom's argument by responding with a *hard-line reply*. Instead of making the claim that at some point there is a sufficient reason to believe that Plum crosses over a threshold, qualifying him for responsibility (or blameworthiness on Fischer's account), McKenna rather develops a response to show that it is 'not clearly established that Plum is not free and is not morally responsible for killing Ms. White' on all four cases – thereby undermining the conclusion of Pereboom's argument (McKenna & Pereboom, 2016, p. 168). To do this, McKenna makes a set of four moves, which I will briefly summarise here.

Firstly, McKenna urges us to consider the many relevant moral properties present in the Case Four version of Plum: just like ordinary moral agents we encounter every day, Plum possesses an advanced, complex, and sophisticated phenomenology capable of forming moral decisions and reacting to them, originating from a rich history of moral development through his existence. We could also allow that he has the relevant capability of emotional responses and reactions, such as regret and pride, etc. We can, in sum, see the Plum of Case Four as a 'fully-fledged member of the moral community' (2016, p. 168). The significant point McKenna makes though, is that 'if all this is true of Plum in case 4, It must be granted that it can equally as well be true of Plum in case 1.'

McKenna's second move is to make the straight-forward request that both compatibilists and incompatibilists 'must allow that in Case 4 it is an open question whether

Plum acts freely and is morally responsible for killing White' (2016, p. 169). The reason for this move being the idea that if one were to fixedly assume either a compatibilist perspective that Plum *is* free and responsible, or the incompatibilist position that Plum is *not* free and responsible on case four, either assumption at the outset would be 'patently question-begging' (2016, p. 169). The usefulness of this point becomes more obvious in McKenna's following move.

The third step McKenna takes, is to use Pereboom's own strategy against him in an attempt to prove that the intuitive pull and *similarity claim* can work in the compatibilist's favour too. Pereboom's argument begins with case one – a case of blatant manipulation and non-responsibility which, following Pereboom's philosophy, shouldn't be treated differently than case two where the conditions are only slightly changed – same for the change to case three, then case four – leading to the conclusion that Plum can't be morally responsible on *any* of the cases. McKenna argues, though, that if we start from case four with an open and undecided mind as to whether Plum is morally responsible or not (as established in McKenna's previous move), and progress through the cases treating each case no different than the last (as Pereboom does with his *similarity claim*), then we end with the conclusion that we can't entirely rule out that Plum *isn't* morally responsible on case one – our same position as case four (McKenna & Pereboom, 2016, p. 169). In face of this counter-argument, therefore, Pereboom's argument appears far less fruitful than previously thought. McKenna's move here, however, rests upon the assumption that a compatibilist can remain open to a possibility that Plum is potentially morally responsible on case one – which is not an easy bullet to bite – but is addressed by McKenna in his last move.

If a compatibilist is to argue in this way, the compatibilist is forced to accept the deeply unintuitive position that Plum might be morally responsible on cases one and two. In response to this, however, McKenna makes his fourth point in emphasising how, in the real world there are 'actual cases which are in many respects just like the imaginary outlandish manipulation cases featured in an argument like Pereboom's, and these cases are not all the uncommon' in the hopes that if we are willing to accept that sometimes these actual persons maintain responsibility, then we should feel more at ease with accepting Plum's responsibility on all cases (2016, p. 170). In quite a convincing tone, McKenna writes that:

> there's no end to the number of naturally occurring real-life cases of "manipulation" in which massive and unexpected alterations disrupt people's lives in ways that dramatically reconfigure their psychic constitutions: People suffer traumatic accidents; have a loved one die in their arms unexpectedly; are crushed during their youth by the weight of violent parents. Yet, if these types of changes leave the adult person otherwise sane, rational, and stable in ways that would allow satisfaction of pertinent compatibilist conditions on free action… very few are inclined to think that these kinds of "manipulation" cases undermine a person's free agency and her responsibility (2016, p. 170).

I'm not convinced at all that this suggestion will go the distance, though. Could one not argue the opposite way that if these real-life cases are examples of genuine manipulation, then perhaps our mistake is *not* in holding Plum not morally responsible when he is, but rather in holding these real-life people morally responsible at all in the first place! For some, this theory from McKenna is satisfactory enough to maintain a compatibilist position in the face of Pereboom's incompatibilist four-case manipulation argument. However, necessary for accepting McKenna's *hard-line* reply is a biting of the bullet in regards to maintaining Plum's possible responsibility in the first cases; even if one is able to appeal to real world cases where manipulation and responsibility may be simultaneously present, surely we can also uncover real world examples of manipulation where considering somebody to be responsible would be ludicrous – closer to the conviction Pereboom tries to evoke in his cases one and two. For some compatibilists, this requirement will be too untenable and they'll retreat to a *soft-line reply*. Incompatibilists also could consider this last move of McKenna as unsound, and thus remain resolute in the strength of Pereboom's argument. As for Pereboom and McKenna themselves, the impasse solidifies itself in this disagreement over whether the 'neutral inquiring audience [has] been given sufficient reason to move away from an open-minded agnosticism and toward an incompatibilist conclusion' in light of case one and case two, or whether 'the hardline compatibilist have sufficient resources to maintain that, once set out, it remains rational for such an audience to remain neutral' (McKenna & Pereboom, 2016, p. 171). Summarised best in their own words: 'McKenna thinks Pereboom is deluded, and Pereboom thinks McKenna is bat-shit crazy' (2016, p. 171).

Though McKenna has certainly, and thoughtfully, tried to present a *hard-line reply* as palatable as possible, it's not outrageous to deny its ultimate acceptability. There is one thing from McKenna's position that I won't easily dismiss, however, and that is McKenna's

judgement of *soft-line replies* being ultimately useless. McKenna explains how although a *hard-line reply* will apparently involve bullet-biting as demonstrated earlier, a *soft-line reply* will never be able to win the war, despite perhaps seeming to win battles – an incompatibilist is always able to adjust their manipulation argument such that the new proposed conditions for compatibilist responsibility is captured (McKenna & Pereboom, 2016, p. 164). McKenna writes that 'in adopting a softline reply to pertinent instances, compatibilists just forestall the inevitable, which is a revised case of manipulation, and a new instance of a manipulation argument for which compatibilists will not have the option of a softline reply—for which, they'll *have* to face the music and take on directly a hardline reply' (2016, p. 164). Whilst it is my position that McKenna's *hard-line* approach isn't ultimately successful, I do accept the belief that if any response is able to overcome Pereboom's manipulation argument, it will certainly have to be a type of *hard-line reply*.

Not all philosophers agree with McKenna's criticism of *soft-line* approaches, however. In response to McKenna's dismissal, Haji and Cuypers have argued that McKenna's interpretation of the four-case argument is 'not dialectically the most charitable way to understand Pereboom's argument' (2006, p. 25). McKenna makes a mistake, they say, when accepting Pereboom's presentation of cases where an agent can be simultaneously manipulated, thus not morally responsible, and also meet all the compatibilist's criteria for responsibility. This move necessarily involves presupposing that there cannot exist any such version of compatibilist criteria for responsibility that sufficiently rules out the presence of manipulation, it's the view that '*no matter what* compatibilist contender is on the table, "some credible manipulation case could be fashioned" that calls this contender into question' (Haji & Cuypers, 2006, p. 22). According to Haji and Cuypers, though, if we are to be charitable and accept Pereboom's intention of putting forth an argument where the first two cases *do* demonstrate a genuinely manipulated agent, then this would certainly rule out any potential success from a *hard-line* reply: the four-case sequence would not have any purchasing power if we were to reject the idea that the manipulation present is a sure threat to responsibility (as McKenna has argued), and thus Pereboom's argument loses its dialectical force. However, if we honour Pereboom's intentions and appreciate him to be saying 'something of this sort: "You, the compatibilist, agree that there are manipulation cases (such as cases 1 and 2 in the four-sequence argument) in which the agent, Plum is *not*

responsible *because* he is manipulated…', then it appears a *hard-line reply* of McKenna's type that is happy to maintain Plum's responsibility on cases one and two is no longer an option available to us, and thus our only hope left is returning to a *soft-line* reply (Haji & Cuypers, 2006, p. 28).

McKenna and Pereboom argue, though, that it is indeed possible to put forth a formulation of case one where it appears both manipulation and any compatibilist's criteria is present. In a new version of the argument, Plum acts mostly on his own accord and ability without intervention from the neuroscientists. The neuroscientists, though, still possess the *same ability* to manipulate moment by moment, however, on this new case, the neuroscientists only use their ability once:

> In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White (Pereboom, 2014, p. 76).

Regardless of whether or not this new presentation is convincingly puzzling enough to be accepted by Haji and Cuypers, if we deny this reply and instead stick with Haji and Cuypers' theory of freedom, we will still run into trouble elsewhere. Haji's *soft-line reply* holds a conception of responsibility centred on three conditions of what he calls *appraisability* – these include:

> a control constituent, which says that the sort of control required for moral appraisability is volitional control; an epistemic one, which— stripped down to its core— says that in order for an agent to be morally appraisable for an action, the agent must believe she is doing something wrong or morally amiss, or she is either executing her moral duty or at least doing what is morally permissible by performing that action; and finally, an authenticity constituent, which says that the agent's action for which she is appraisable must issue from actional springs that are "truly her own" (Haji, 1998, p. 237).

It is this third ingredient of ownership that springs up a deeper consideration of sourcehood. We can circle back to Pereboom's argument, and instead of questioning responsibility, we can question the ownership of actions that Plum has. What is the relevant difference between cases 1-2-3-4 that offers Plum ownership on some and rules it out on others? As

Haji argues, if Plum is genuinely manipulated on cases one and two, then his decisions aren't *truly his own*. But what, then, is the new element or ingredient found in case three that renders Plum a sufficient source and owner of his action? How is the 'rigorous training practices of his home and community' that determines Plum's action different to the neuroscientists who, 'although they cannot control him directly, have programmed him to weigh reasons for action so that he is often but not exclusively rationally egoistic' which leads to Plum killing Ms. White, such that Plum is a sufficient source of his actions on the former but not the latter?

According to Pereboom, the underlying cause for why such a relevant difference doesn't exist (and therefore, impossible to identify) is because on all cases, Plum fails to be an adequate *source* of the factors that influence his decision.[21] Even if we take case three as a stand-alone scenario, and even if we accept the compatibilist's belief that Plum possessed the *relevant control* to reflect upon and consider what he had been conditioned to value by his community upbringing, and that he had an adequate control to decide for himself whether or not to kill Ms White, still, the reasons for why he killed White – the reasons that entered into Plum's legitimate reasons-responsive mechanism could be traced back to a source external to Plum – his parents and community – a source that he could not control whether or not its influence entered into his reasons-responsive process. Plum could not, for example, decide who his parents are that raised him, or decide which community or location he was born into that played a role in shaping how he saw and acted in the world. And therefore, due to this lack of control, Plum cannot be held morally responsible on Pereboom's account. Indeed, even if he chose *not* to kill Ms White, such a decision would have a reason for acting as so, and this reason could be traced back to sources outside of Plum's agential structure.

Pereboom sums up his reasoning concisely in the following origination *Principle (O)* that: 'If an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she

[21] Even if we take McKenna's approach to start at case 4 and move backwards, yet still accept Haji and Cuypers' approach of maintaining genuine manipulation present on cases one and two, we'd then struggle to identify the reason for jumping from proper sourcehood on case three to improper sourcehood on case two.

has no control.' (Pereboom, 2001, p. 47).[22] Professor Plum is therefore *never* responsible for his action on any of the cases because his actions and decisions all result from a deterministic causal process that traces back to factors beyond Plum's control (Pereboom, 2001, p. 112). Even if we were to remain agnostic on the truth of determinism, this *Principle (O)* affects equally a deterministic theory and an indeterministic theory as it does not appear possible for *any* agent to ever be free from external influence or possess an adequate sort of control over their decisions and actions on either causal law of the universe. Considering even an incompatibilist libertarian position like Kane's that holds that we very *rarely* are ultimately responsible for our actions, the existence of these *rare* occasions are ruled out on Pereboom's *Principle (O).*

Haji and Cuypers, among others, have made many criticisms of this strict principle from Pereboom; however, this belief in the unattainability of the required ultimate sourcehood has certainly stuck around, and it appears to have caused a deadlock between parties with some philosophers suggesting that some slightly weaker form of sourcehood would suffice for responsibility, whilst others remain fixed that responsibility requires nothing less than a strict point of origination or initiation as expressed in Pereboom's principle. I do believe that there are many objections to this principle that are worth consideration, however, much like the other objections we have seen in this chapter, they are intricate and plentiful. What interests me more at this point, though, is the challenge of seeing how far one can go whilst accepting this sourcehood principle from Pereboom. Whilst other free will theorists have seemingly hit a brick wall in trying to avoid Pereboom's unattainable sourcehood requirement, in the following chapter what I will develop is an approach to answering Pereboom's four-case argument that does *not* rely upon a dismantling of the argument's coherence, but instead the approach takes the argument at face value and examines the potential success of an approach that contains a different metaphysical foundation, which consequently, I believe, doesn't fall into the same stalemate of Haji, McKenna, or Fischer and Ravizza.

[22] This principle formulated in Pereboom's words is a development from R. Kane's conception of *ultimate responsibility* (UR) in The Significance of Free Will (New York: Oxford University Press, 1996), p. 35.

Chapter Four

A Gaudiya Vaishnava *Soft-Line* and *Hard-Line* Reply

So far, the picture I have painted of the current state of compatibilism in the free will debate may appear gloomy and forbidding, however, my intention in this section is to return a sense of optimism into the debate. This chapter is a continuation on from chapter three, examining the coherence of a compatibilist approach to reconciling responsibility and causal determinism, and in particular, how a Gaudiya Vaishnava compatibilist theory holds against the compelling arguments that undermine compatibilist conceptions of sourcehood. Though I previously agreed with McKenna's judgement that a winning reply to Pereboom's four-case manipulation argument will have to be a *hard-line reply*, I will somewhat paradoxically begin with a *soft-line reply* that will eventually develop into a *hard-line* response. I should perhaps offer the following warning regarding my forthcoming replies, that though they tackle the problem straight-forwardly, it is done so certainly unconventionally; there are many possible objections that could be raised from those against the introduction of Vaishnava metaphysics, and whilst I may touch upon a few potential inconsistencies, this is not a defence of the existence of Indian metaphysics sustaining our universe, this is a defence of a Gaudiya Vaishnava compatibilist conception of responsibility and approach to answering Pereboom's arguments for incompatibilism.

To begin with a short reflection, what we found in chapter two was that *agency* had been often presented in Gaudiya Vaishnava scripture in relation to the term '*vimṛśya'* which implies an act of *deliberation*, *reflection*, and *consideration*, in quite a comparable way to a form of *reasons-responsiveness* that Fischer and Ravizza posit. And in other areas of Vaishnava literature, agency was presented in terms of being the *source* of one's actions and desires; firstly, in the philosophy that though it appears our agency is influenced by the senses (prāṇa), the soul remains an ultimate source of the influence because we as souls are 'the grasper of the prāṇa' (Vidyābhūṣaṇa, 1718/2018, p. 262); and secondly, the concept of s*aṃsāra* opens the door to seeing our casual history as something the extends far beyond this single body – meaning that though we may not remember our previous actions that were performed in previous lives, we are still experiencing the effect of them in this life –

what appears external to us, may truly be internal to us instead, and thus we are potentially a *cause* of (and thus, *responsible for*) much more than we think.

I should quickly note here that is due to this depiction of free will presented in Vaishnava theology that I am presenting here a *sourcehood* compatibilist approach, rather than a classical *leeway* compatibilist approach that is concerned primarily with *an ability to do otherwise* and *genuine open alternate abilities*. Though there certainly is merit in a *leeway* theory of compatibilism, the Gaudiya Vaishnava concept of a soul and reincarnation no doubt provides much *more unique* and *compelling* content for consideration regarding how we view and understand an agent's causal history, and thus, a *sourcehood* approach will be my primary focus. On a larger picture, I should also preface my argument by reiterating the point that though I am here presenting a compatibilist theory, much like Fischer and Ravizza I do believe that on my conception, one can remain agnostic towards to the truth or falsity of determinism; although I accept that there must exist a truth to the matter according to Gaudiya Vaishnava theology, I'm not convinced that such a truth is relevant to a definition of moral responsibility.

Returning now to my approach, in terms of the analytic terminology employed in current free will literature, a Gaudiya Vaishnava form of compatibilist conditions could perhaps be construed in a formula such as: An agent A is free with respect to an action X so long as they meet the *control principle*: the action X must have issued from the agent's appropriately reasons-responsive mechanism, and also meet the *sourcehood principle*: the agent is an ultimate source of the action occurring. Of course, if one is committed to a form of physicalism, then this theory of free will will certainly not be available to them, however, I am hopeful that perhaps if the conception proves coherent enough, it may indirectly serve as a tempting reason to join us on the supernatural side. And with that, to see how well this conception stands let us now put it to Pereboom's test.

A Gaudiya Vaishnava *Soft-Line* Reply

Beginning with a *soft-line* approach, and considering the Gaudiya Vaishnava worldview described in chapter two, I'd like to propose the thesis that Professor Plum is definitely morally responsible in case three and four, due to the statement that 'Plum is an ordinary human being' in the first sentence of each scenario, and that on cases 1 and 2, he is *not* morally responsible due to not being an ordinary human being but instead 'created by

neuroscientists'. On cases three and four, Plum's humanness gives us the chance to consider Plum as an eternal soul with a possible history of prior actions in previous lives that have played a causal role in his current decision making process to kill Ms. White. Whereas on cases one and two where neuroscientists have created Plum, we can deduce that Plum consequently does not possess an eternal soul with a possible history of prior actions in previous lives.

This idea of an eternal soul that can undergo the process of reincarnation problematises directly the hidden principle in Pereboom's manipulation argument that holds that all agents ultimately have causal chains originating *outside* of them. While other philosophers are forced to accept that with any action, there will be influences that affect the decision to action that originate outside of the agent, for it is not possible for the agent to play a causal role in who they are born to, in what country, in wealth or poverty, or the parenting style they are raised in; philosophers who accept the doctrines of reincarnation and karma can argue that it *is* possible for agents to play causal roles in the influencing factors that later appear as external to the agent. Whilst if the Gaudiya Vaishnava approach accepts the truth of determinism, and therefore affirming the existence of causal chains, the Gaudiya Vaishnava proponent is *not* also committed that these chains are necessarily *outside* of the agent. Plum on case three, for example, experiences influence from his home and community, and usually we would consider this an external influence as it is an influencing factor that does not originate in him and he had no control over who would be his parents and community. According to Gaudiya Vaishnava theology, however, one *does* have some control over who their parents will be when the soul, as an agent, moves from one body at the time of death into a new body. Though they may not consciously decide their parents from a catalogue, according to the spiritual tradition the parents or circumstances that one is born into is determined by the soul's accumulated karma and desires. Professor Plum's birth into the community that influenced him to commit the heinous crime has casual chains that originate *not* completely external to Plum, but can be traced back further from previous actions in previous lives. Just as an alcoholic suffering heavily from the freedom-undermining influences of addiction is still somewhat responsible for who they become because their original decision to over indulge. Professor Plum, though suffering heavy influence from those around him, is ultimately responsible for the association he has due to prior actions or desires that have created this environment.

In this undeveloped state, this theory is insufficient in two ways: firstly, it doesn't meet the sourcehood criteria necessary to maintain freedom and moral responsibility. The ability of reincarnation doesn't, by itself, actually infer the non-existence of a factor that is externally sufficient for the agent's action. With reincarnation, we can ask: does it lead to an infinite regress of sourcehood? Or if there was a beginning cause or creator, would this not be a final external influence which would hold the moral responsibility rather than the agents? And secondly, karma appears thus far as a type of fatalistic self-determination that *undermines* one's free will and moral responsibility.

Let's start by briefly addressing the first issue of sourcehood. We're correct in judging that it's not enough to say 'agent M is morally responsible for action X because the reason agent M was inclined to perform X is due to actions Y in his previous life' as we are then required to inquire what caused actions Y, perhaps it was attitudes Z, then what caused Z? and so on infinitely. Even accepting the concepts of reincarnation and the eternality of souls, in this undeveloped state my GV view would not overcome Pereboom's *Principle (O)* I previously mentioned that: 'if an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control' – so far I have simply pushed sourcehood into an infinite regress that requires an ultimate source, which, we may logically theorise to be, *external* to the soul (Pereboom, 2001, p. 47). It seems that even with the ability for the soul to reincarnate, the soul itself is not sufficient enough to overcome Pereboom's sourcehood requirement because the soul couldn't be a *causa sui* – cause of itself. To address this sourcehood principle, I'll move now to Galen Strawson's Basic Argument in order to flesh out these complexities more thoroughly and use it to springboard my development of an adequate theory of sourcehood on the reincarnation theory.

Strawson's Basic Argument is a compelling attempt to explain how it is impossible to be truly morally responsible for anything, no matter the truth or falsity of determinism, due to the inability for any of us to be a *causa sui*, and it goes as follows:

> **(1)** Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to 'reflex' actions or mindlessly habitual actions).

(2) When one acts for a reason, what one does is a function of how one is, mentally speaking. (It is also a function of one's height, one's strength, one's place and time, and so on. But the mental factors are crucial when moral responsibility is in question.)

(3) So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking - at least in certain respects.

(4) But to be truly responsible for how one is, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way.

(5) But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, 'P1' - preferences, values, pro-attitudes, ideals - in the light of which one chooses how to be.

(6) But then to be truly responsible, on account of having chosen to be the way one is, mentally speaking, in certain respects, one must be truly responsible for one's having the principles of choice P1 in the light of which one chose how to be.

(7) But for this to be so one must have chosen P1, in a reasoned, conscious, intentional fashion.

(8) But for this, i.e. (7), to be so one must already have had some principles of choice P2, in the light of which one chose P1.

(9) And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice.

(10) So true moral responsibility is impossible, because it requires true self-determination, as noted in (3). (Strawson, 1994, p. 6-7).

If we wish to thwart the argument, we have to answer a few questions: Does the eternality of the soul imply an infinite regress of sourcehood? If so, can people be held morally responsible? And if so, how can karma function fairly without an initiating cause? If the eternality of the soul does *not* imply an infinite regress of sourcehood, then surely there

must be something external that holds the moral responsibility, and so how can we hold individual agents morally responsible?

On the GV view, I believe, we can both deny an infinite regress and deny an externally sufficient cause that is completely morally responsible. I can happily accept premises 1-8 and deny the idea in the ninth premise that posits that it would entail 'a regress that we cannot stop' thereby also denying that 'true moral responsibility is impossible.' To justify this we simply require an elucidation on Vaishnava metaphysics, which I shall attempt to explain now.

I'll start by bringing the idea of God into the picture with a quote from Strawson himself who states that 'even if such a causal "aseity" is allowed to belong unintelligibly to God, it cannot be [sic] plausibly be supposed to be possessed by ordinary finite human beings' (1994, p. 15). However, according to Vaishnava philosophy, human beings aren't as ordinary as Strawson assumes. Though I have already offered a brief overview of the main theological aspects of Gaudiya Vaishnavism in chapter two, I will reiterate here some of the features of relevance for our discussion.

Firstly, as individual persons, we are immaterial souls (known as a *jīva* or *ātman*) that are distinct from our bodies. And as souls we have the potential for reincarnation – the transmigration between bodies at the time of death: *dehino 'smin yathā dehe kaumāraṁ yauvanaṁ jarā ; tathā dehāntara-prāptir dhīras tatra na muhyati* 'As the embodied soul continuously passes, in this body, from boyhood to youth to old age, the soul similarly passes into another body at death. A sober person is not bewildered by such a change.' (Prabhupāda, 1983, 2.13).

Secondly, as souls, we are eternally existing beings: *na jāyate mriyate vā kadācin nāyaṁ bhūtvā bhavitā vā na bhūyaḥ* ; *ajo nityaḥ śāśvato 'yaṁ purāṇo na hanyate hanyamāne śarīre* 'For the soul there is neither birth nor death at any time. He has not come into being, does not come into being, and will not come into being. He is unborn, eternal, ever-existing and primeval. He is not slain when the body is slain.' (Prabhupāda, 1983, 2.20).

Thirdly, there exists a law of *karma* that can play a causal and retributive role in the world that, in a very basic and simplistic sense, is the philosophy that 'besides its purely physical effects, our actions have also moral and psychological effects, which might occur either in this life or in future ones, and which affect our environment, genetic make-up,

physical characteristics, social status at birth, length of life, etc., as well as our psychological dispositions and tendencies to act, desire, etc' (Silvestre, 2017, p. 38).

Fourthly, there are two distinct realms that the soul can populate: the material world – the one within time and space in which we live in now, and the spiritual world/divine realm – a pure plane which is unaffected by time and space, in which the souls are together with God: 'Between the spiritual and material worlds there is one tattva suitable for both worlds... This is called jīva-tattva [the living souls]. Jīva's composition is a spiritual particle. Since it is weak (being a particle), it is suitable for being bound in the material world. Since it is pure by composition (conscious particle), if it obtains a little spiritual strength, it can live eternally in the spiritual world in complete bliss' (Thākura, 1896/2019, p. 47). Not only is the spiritual world the destination of ultimate liberation and salvation, it is also the original home of the eternal soul.

And lastly, the spiritual world being the original home of the eternal soul can be explained by the *bhedābheda* school of Gaudiya philosophy. As previously described in chapter two, *Bhedābheda*, literally translating to *difference (bheda)* and *non-difference (ābheda)*, is a description of the relationship between God and the eternal souls. The philosophy is that the individual souls are *qualitatively* one with God, in that we are literally plenary portions *of* Him, but we are *quantitatively* different in that we are finite, limited beings, whilst God is infinite and unlimited. *Gaudiya Vaishnava* proponents do not hold the Abrahamic conception of God at some point creating the finite being, instead, the theology is that we have always existed as parts of God himself, and yet we are different from God in our inherent *being less than* God in power, benevolence, knowledge, etc. From this *bhedābheda* relationship, there is also the doctrine of a fall – the belief that though originally the soul was together with God in this divine realm, at some point, due to the finiteness of the soul which inherently brings with it a risk of mistake or error, develops the desire to leave the divine realm.

Returning to Strawson's Basic Argument, we can start making replies to the obstacles that would undermine our ability to defeat the argument. If we were to concede that the eternality of the soul implies an infinite regress of sourcehood, then I do believe it would be difficult to hold one morally responsible. A quick argument would be that infinite regresses are physically impossible, therefore, there must be an initiating cause, thus, there is an

ultimate source that is external. Whether an actual infinite regress is physically possible or not, though, isn't actually relevant in Strawson's case, just that moral responsibility would not be compatible with it as 'true self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice' (1994, p. 7).

Aside from the soul, though, is the infinite regress of *karma* logically possible? Or a self-contradiction? My position is certainly that the doctrine of karma relies upon rejecting the idea of a physical infinite regress: karma would not be fair if it did not begin from a clean slate and initial causing action. For somebody to experience some karmic reaction, they must have performed a karmic accruing action – and this is a beginning. But if karma is beginningless and we have always existed in s*aṃsāra* for an eternal past as long as we as souls have existed eternally, then this would mean that for every karmic consequent we suffer there was a karmic accruing action, and for every karmic accruing action, the action has been at least partially influenced by a previous karmic consequent. Which came first, the chicken or the egg? Which came first, the karmic accruing action or the karmic consequent experience? If karma is beginningless, we have no answer.

On the GV view, however, this is avoided by rejecting the existence of an actual infinite regress of this material world, despite affirming the eternality of the soul.[23] Though there was never a point where the jīva (soul) was created, there was a point where the jīva was once outside of time and space, and then moved into the realm of time and space (recall back to my previous description of the fall form the divine realm). And thus, 'The root of all karma is the desire to perform actions. The root of the desire to perform action is ignorance (*avidyā*). *Avidyā* means forgetting that one is the servant of Kṛṣṇa [God]. That ignorance does not start within the material time. The *jīva's* root of karma (*avidyā*) arises at the junction in taṭastha. Therefore there is no beginning to karma in terms of material time. Thus karma is without beginning' (Thākura, 1896/2019, p. 222).

The answer is simply that though we have existed for eternity, accepting a fall doctrine means that we have not existed for eternity *in this realm of time and space*. Instead, we were in a spiritual realm and at some point, due to our inferior nature compared to God's infiniteness and supremacy (meaning, as plenary portion of God we are not

[23] I'm not rejecting that it's possible. But on the Gaudiya Vaishnava worldview, such a regress does not exist in actuality.

omnipotent, omniscient, and omnibenevolent), we rejected God and thus we fell/entered into the material world – this realm within time and space. Our karma (and therefore moral responsibility) can be traced back to this initial failure – and this is the root choice that initiates karma and saṃsāra.

But if the desire/decision to reject God started our karma, following Strawson and Pereboom we must inquire what was the prior cause of that rejection? The answer is that it all traces back to our constitutional position as parts and parcel of God. Because we are merely parts of God, we do not possess the entire abilities of God, just as the hand has certain bodily abilities, it cannot do the work of an entire whole body. According to *Gaudiya Vaishnava* theology, we have God's powers and abilities, but to a minute or lesser degree. Because of this, there is an inherent risk of deviation – stemming from avidyā, as mentioned in the previous quotation. Thus, this is the final root cause of our actions. Because we don't know everything, and because we can suffer from forgetfulness, this leads the jīva to desire X and thus reactions occur. And this condition of the soul is eternal and intrinsic. This is how I believe that on this GV view, we are not forced to accept steps nine and ten of Strawson's argument. We can conceive of a causal chain of responsibility back to a point that is inherent to the soul itself. Although it may not be a cause of itself, it is enough to undermine Strawson's argument. It does not work in completing an infinite series of choices of principles of choice, but it is a *self-determination* in how there exists an original source that is *not* external to the agent.

One may wish to argue, though, that because all actions can be traced back to essential aspects of the soul which have influence yet aren't caused *by* the soul, then the agent lacks moral responsibility because they did not make themselves in such a way – they never caused themselves to be that way and in that condition. Pereboom and Strawson may wish to argue that an inherent finiteness is exactly the eternal ultimate/final cause and thus we aren't responsible for who we are. Pereboom is perfectly clear, 'if an action results from a deterministic causal process that traces back to factors beyond the agent's control, then he is not morally responsible for it' (2001, p. 116).

My rebuttal to this, however, is that though our finiteness may be something that we cannot control, the finiteness alone is not a *sufficient condition* for an initial action. The law of gravity doesn't impel the ball to move, but once the ball is in motion, the law of gravity limits the abilities of the ball. The rules of chess are not *causally sufficient* for a game to

begin – one may decide the rules, yet not play the game; if one plays, though, then they are governed by the rules of chess, but the existence of the rules does not act as a cause *in itself* for the game to begin.

At this point, let me introduce some simple logic to develop my theory; consider the difference between these two conditions: **Necessary Condition:** 'A condition *A* is said to be necessary for a condition *B*, if (and only if) the falsity of (/non-existence/non-occurrence) [as the case may be] of *A* guarantees (or brings about) the falsity (/non-existence/non-occurrence) of *B*' e.g. Having four equal sides is a necessary condition for being a square (Swartz, 1997). And **Sufficient Condition: '**A condition *A* is said to be sufficient for a condition *B*, if (and only if) the truth (/existence/occurrence) [as the case may be] of *A* guarantees (or brings about) the truth (/existence/occurrence) of *B*' (Swartz, 1997).

Sufficient conditions aren't always necessary, though. A sufficient condition for travelling to Auckland from Australia would be taking a direct flight on Air NZ. However, taking a sailboat to Auckland would also be sufficient condition, a kayak would be sufficient too, if also brave enough. Consider now **Causal Dependence:** Where *A* and *B* are two distinct possible events, *B causally depends* on *A* if and only if, if *A* to occur *B* would occur; and if *A* were not to occur *B* would not occur. (Menzies & Beebee, 2024). What's essential to notice is that 'the definition requires the causally dependent events to be *distinct* from each other. Distinctness means that the events are not identical, neither is part of the other, and neither implies the other' (Menzies & Beebee, 2024). J. Kim points this out also that whilst it's true that if I had not written 'r' twice in succession, I would not have written 'Larry' the dependence is *not* a casual relation (1973, p. 571); 'neither dependence counts as a causal dependence since the paired events are not distinct from each other in the required sense' (Menzies & Beebee, 2024).

Similarly, our finiteness is an intrinsic property that can restrict and limit our actions, but it does not play a causally initiating role. The fact that a square has the intrinsic property *x* of having four equal sides can limit the square's abilities/other properties – in that because of that property *x*, the square can't also have the property *y* of having one side larger than the other sides. But it is not that the property *x* plays a *causal* role in a square being a square, the property *x* is a necessary condition – the *x* and *y* are not distinct enough. Certainly, we could say the same for chess: the rules of chess *x* are a necessary condition for the event *y* Mary winning a game of chess. For if *x* did not exist, then *y* could not occur.

Though there is a dependency here, it is not a causal dependency because the events/facts are not distinct from each other; *y* implies the existence of *x*. The rules of chess *x* alone are not a sufficient condition for *y* Mary winning a game of chess, for it does not play a causal role (in itself). As we can imagine a world in which the creator of chess writes the rules down, yet never plays a game.[24]

Let's imagine an eternal game of chess where the pieces have always been in motion. The activity of playing implies the necessary condition of the rules of chess existing. But the rules don't act *causally* to make the game occur. So, whilst the pieces moving and the game playing might be able to be reduced down to a dependence on the rules existing, this is not a causal dependence, but rather two eternally existing truths. Because nothing exists temporally prior to the eternal game of chess and its rules (being *eternal*), then there is no proper part of the distant past which is *causally sufficient* to bring about the game of chess and its rules.

Consider Timpe's simplified presentation of Pereboom's position on sourcehood:

'… let us define a causal chain *x* being *externally sufficient* for agent *A*'s decision to *d* if the following conditions are all met:

1. There is a proper part of *x*, *y,* involving *A*'s agential structure which is causally sufficient for *A*'s decision to *d*.
2. There is a proper part of *x*, *z,* which does not involve *A*'s agential structure.
3. *z* is temporally prior to *y.*
4. *z* is causally sufficient for *y*.' (Timpe, 2008, p. 90)

On my GV view, the final source – the soul's inherent finiteness – is never separated from the agent's agential structure. We can accept premise 1, and reject the rest as the two elements at play here – the soul and the soul's condition – are not *distinct*. Consequently, I believe this is how a compatibilist conception of moral responsibility can avoid the conclusion of Strawson's Basic Argument.

The necessary consequence of accepting this conception of freedom and sourcehood, however, is accepting some loss of total moral responsibility. As we see on the

[24] We would still argue that *x* plays a restricting and limiting role over the game of chess, and thus in the event of a game of chess being played, the players are all limited in their freedom over the choices.

eternal chess game analogy, although we can conceive of how the act of playing chess both *implies* the existence of rules of chess and the rules of chess are a necessary *part* of playing chess, and therefore we might identify the root source as not being external due to its lack of distinctness. In an eternal game of chess, the players are still limited and restricted in their moves: a bishop can only move diagonally, a king can only move one square at a time, etc. Despite this limitation, though, we still hold winners and losers responsible. For although there may have been restrictions on what they could do, the players *did* possess the option of choice; players can go through a process of deliberation (or a reasons-responsive mechanism) and decide their move, and the opposing player responds to this – much like karma – influencing our next set of choices we have to choose from. Imagine somebody losing a game of chess and exclaiming that they only lost because of the existence of the rules of chess. The winning player would surely reply that whilst the rules of the game did restrict some freedom, there was still *enough* freedom of choice available for you to be held responsible for your loss, and mine for the win.

Similarly with our performance of actions in the world, though there is the existence of an inherent finiteness which can limit our abilities and the choices available to us, we are still responsible for our actions *enough* for our everyday use of praising and blaming. Even in cases where agents seem to lose their freedom completely, say an alcoholic in the pits of addiction, they are somewhat still responsible for the reduction in freedom because of their previous choices that brought them there – just as the player who is checkmated and consequently no longer has the freedom of choice, is still somewhat responsible for that position because of their previous chess moves.

At this point, I must pause to make the concession, though, that the conception of sourcehood I have presented here falls short of meeting Strawson's strict sense of a *causa sui* – a being that is a cause of itself: 'In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects' (Strawson, 1994, p. 5). Indeed, a *causa sui* is a completely unattainable position that I'm not sure even a God could fit; and certainly so according to a GV view where a soul is eternal and *without* cause. However, I do wonder if the GV conception would meet the relevant criteria of sourcehood that Strawson is implying with 'at least in certain crucial mental respects' (1994, p. 5). It is tempting here to argue, as others have done, that Strawson's *causa sui*

criteria is too strong and throwing it out entirely, doing this however, runs the risk of allowing weaker conceptions of sourcehood to be sufficient which, though it may allow room for some responsibility to be attained, may also inevitably lead to an inability to distinguish responsibility from non-responsibility in circumstances like manipulation. It *is* my opinion that the *causa sui* criteria is too strong, as not even a hypothetical being could be one. However, I do think that there may be a benefit derived from working backwards first — beginning with Strawson's criteria and weakening it just enough until we find a sweet spot somewhere that allows moral responsibility, yet not so weak that it falls victim to arguments from manipulation such as Pereboom's.

The crux of Strawson's argument is that, for moral responsibility 'one must have brought it about that one is the way one is, mentally speaking', and following this, it is not the case that the GV soul can meet this as, on the GV view, the soul's desires and decisions stem from the soul's eternally co-existing internal nature that, by being internal, is inseparable and non-different from the soul itself, and thus, not able to have been *brought about* in Strawson's sense (1994, p. 6). However, on the GV view, it is *not* the case that something or somebody other than the agent has 'brought it about that one is the way one is, mentally speaking', for the soul's nature that is the original influence is as internal and causeless as much as the soul itself is. And therefore, I wonder if the soul could still possess *some* moral responsibility 'at least in certain crucial mental respects' by virtue of being an agent that is causeless and lacks any original external influencing factor – an agent whose actions can be traced back to influencing factors that are solely internal to the agent (1994, p. 5).

And in a similar way, we could reflect back upon Pereboom's *Principle (O)*: 'If an agent is morally responsible for her deciding to perform an action, then the production of this decision must be something over which the agent has control, and an agent is not morally responsible for the decision if it is produced by a source over which she has no control.' (Pereboom, 2001, p. 47). Although the GV proponent could maintain that for all morally relevant decisions, the production of the decision was something over which the agent had control over, this control, however, was limited by the soul's finite intelligence – a finiteness that was *not* under the soul's control, in that, a soul cannot simply decide whether to be eternally finite or infinite, it just *is.* In my judgement, this conception does rule out total control and total responsibility, but it doesn't rule out partial control. And this partial

control is all that is required to satisfy our natural intuitions that people are variably responsible for their actions in the everyday sense.

A glaring possible objection to this aspect of my GV theory is that such a loss of complete control isn't *fair*, but with this question we may start teetering close to the *problem of evil*. Certainly, if we accept such a conception of freedom and responsibility, then we are forced to consider also the coherence of the associated metaphysical baggage the theory entails. Although I will provide a very short reply to a potential question over fairness in chapter five, there is a more obvious objection regarding my reply to Pereboom four-case manipulation argument that I will address first – which is the point that the first two cases can easily be adjusted to exclude the idea of Professor Plum being *created* by neuroscientists, and instead account for the GV conception of Plum being an eternal soul. This is a sound and easy objection, however, it would not cause me to erase the conception of responsibility I have developed so far, but rather, require me to adjust this *soft-line reply* into a *hard-line reply* to Pereboom's argument.

A Gaudiya Vaishnava *Hard-Line* Reply

Recalling a previous quotation from McKenna and Pereboom that writes 'in adopting a softline reply to pertinent instances, compatibilists just forestall the inevitable, which is a revised case of manipulation, and a new instance of a manipulation argument for which compatibilists will not have the option of a softline reply—for which, they'll *have* to face the music and take on directly a hardline reply', as I promised earlier, I will now attempt to 'face the music' and offer a GV *hard-line* reply to the argument (McKenna & Pereboom, 2016, p. 164).

Incidentally, Pereboom has already put forth a new version of his manipulation argument that accounts for the crucial aspects for Plum's responsibility on my GV account – the potential existence of an eternal soul. So far, I have been employing Pereboom's original 2001 presentation of the four-case argument. Pereboom's 2014 presentation of the argument, however, includes just enough changes to be mostly inconsequential for most compatibilists, but major for a GV reply. The 2014 version is as follows:

> Case 1: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins

to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort.

Case 2: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic (as in Case 1), and at times strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).

Case 3: Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to prevent or alter these practices. Due to the aspect of his character produced by this training, in his present circumstances he is causally determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White. While Plum does have the general ability to regulate his behavior by moral reasons, in virtue of this aspect of his character and his circumstances he is causally determined to make his immoral decision, although he does not decide as he does due to an irresistible desire. The neural realization of his deliberative reasoning process and of the decision is just as it is in Cases 1 and 2.

Case 4: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal

> circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1–3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum's reasoning process and decision is exactly as it is in Cases 1–3; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill. (Pereboom, 2014, p. 76-79).

Whilst on the 2001 version, cases one and two began with the remarks that 'Professor Plum was created by neuroscientists…' and 'Plum is like an ordinary human being, except that he was created by neuroscientists…', which left the door open for a *soft-line* reply. But with the above 2014 presentation, on all cases Plum is regarded as a human being, with no reference to being created by neuroscientist, and thus on this new version, a *soft-line* reply is not open at all for the Gaudiya Vaishnava proponent to present.

The difficulty of constructing *hard-line* response, though, is the way to explain satisfactorily how it could be true that Plum, despite our intuitions, is *just as* morally responsible on cases one and two as he is on cases three and four. Recall on McKenna's *hard-line* approach, McKenna attempted to make the thought of Plum being responsible whilst also manipulation more palatable by urging us to consider the many real-world examples of persons manipulated where we still hold them responsible. The issue I raised against this suggestion, however, is that we could easily argue in the opposite direction that if these real-life cases are examples of genuine manipulation, then perhaps our mistake is *not* in holding Plum not morally responsible when he is, but rather in holding these real-life people morally responsible at all in the first place; in this case, using Pereboom's similarity claim, we would find Plum not morally responsible on all four cases. It is true that in taking a *hard-line* reply, we are forced to accept something that appears initially as unintuitive, and I do agree with McKenna's approach in trying to make the Plum's responsibility in cases one and two more palatable, however, McKenna, by appealing to real-life scenarios has failed to make the case palatable enough. I do believe, however that with an inclusion of *Gaudiya Vaishnava* metaphysics and conception of sourcehood, Plum's responsibility on all four cases can be far more digestible than any compatibilist attempt I have seen.

Let me begin my *hard-line* reply by first considering a different manipulation scenario put forth by Timpe in order to develop my theory before putting it to the test against

Pereboom's four-case argument. On the previous, 2001 version of Pereboom's argument, we overcame the threat of manipulation on case three by explaining how it can be the case that Plum played a causal role in bringing about the influences he experienced that contributed to his decision to kill Ms. White. However, this method of avoiding the undermining effects of manipulation alone is not applicable to all cases of manipulation. Consider the follow freedom-undermining manipulation case offered by Timpe:

> Allison's husband, Hal, implants desires in her volitional structure via hypnosis (or via a neursoscope he's purchased from Frankfurt's *Neurostimulators Inc.*). Suppose that Hal wants his father-in- law to die because he stands to inherit a considerable sum of money from her father's estate. However, Hal doesn't want to wait for the money, as he owes his bookie for unpaid gambling debts. So he hypnotizes his wife into killing her father. Such a case could be described in such a way that Allison satisfies all of *Fischer/Ravizza's Condition*, thereby having free will on their account, although plausibly Allison should not be thought to be free in virtue of the manipulation involved. Suppose that when Allison is deliberating about their financial situation, Hal causes her to have a certain reasons-responsive mechanism of the sort that Fischer and Ravizza think is required for guidance control—that is, he induces in her a volitional structure that is different from the one that she previously had. This new reasons-responsive mechanism is so sensitive to their financial demands that it judges that killing her father to receive the inheritance immediately is the morally preferable action; nevertheless it satisfies the first half of *Fischer/Ravizza's Condition*. In addition, during the process of hypnosis Hal also causes her to take ownership of this new mechanism thereby also fulfilling the second half of their condition as well. Because of the inputs from her husband, Allison sees herself as the source of the choices that result from this newly implanted volitional structure, believes that she is responsible for how she chooses on the basis of this new structure in appropriate circumstances, and believes this on the basis of the evidence that she has at the time. Allison acts on the bases of these reasons and kills her father for the inheritance. I[t] appears that in such a case, Allison would meet all the requirements that Fischer and Ravizza think are required for guidance control despite the fact that her meeting all these requirements is the result of her being hypnotized or otherwise manipulated by her husband. But there is good reason to think that in this case Allison would not have the kind of control over her choices required for her to be morally responsible for killing her father since all of the relevant factors are the product, not ultimately of Allison herself, but of her being hypnotized by her husband, Hal. (Timpe, 2008, p. 81-82).

For other compatibilists, the first aspect of this threat to freedom that they would tackle is the point that *because* Allison is not the proper source of what's been implanted, then she cannot be held morally responsible – and this is where they get stuck trying to

develop a theory of sourcehood that *avoids* responsibility on these cases, but *maintains* responsibility in circumstances like cases 3 and 4 of Pereboom's Plum scenario. Whilst my theory can posit an adequate account of sourcehood, there still appears to be a roadblock elsewhere: although Allison may be the karmic source of *being implanted with Hal's desires* (and therefore seemingly the cause of the loss of her own freedom), why then should she be held morally responsible for the action of killing her father? To overcome this obstacle, I believe we need to flesh out a distinction of Hal's causal role in the killing of Allison's father.

Imagine a case where Hal implanted the desires within Allison but did not know for sure whether it would work successfully to lead Allison to perform the murder. This is a similar situation as to Plum in case three, where his parents and community implanted the desires within Plum, but didn't necessarily know for sure what the later effects of it would be or the events that would be produced. With Hal and Allison, on this case we might say that she was free due to her ability to deliberate (or due to her reasons-responsive mechanism) and being the (indirect) karmic source of the implanted desire (which she is able to reflect upon and reject). But the fact that Hal does know exactly what will happen doesn't mean that he 100% causes it to occur. *Knowing* that *if* he did *a, b* would happen, does not follow that his doing *a causes* entirely for *b* to happen, though it does make him morally responsible if he did. We can approach this scenario the same way some theists have approached God's blameworthiness: although God, from His omnipotence, *knows* precisely what will happen in the course of the universe, it does not mean that he necessarily *causes* exactly what will happen in the course of the universe.[25]

My point is that, so long as the process of deliberation (or relevant reason-responsive mechanism) is uncorrupted and legitimate, then Allison is free. Just because Hal *knows* what will happen if he did *x*, it doesn't mean he is the *total cause* of *x* producing *y*. He is morally responsible, and so is Allison. Allison is morally responsible her action issues from her deliberate decision and reasons-responsive process, and Hal is morally responsible because by knowing 100% the effect of his action, then he is, in a way, agreeing to allow that effect to occur.

[25] As far as I'm aware the only theists that can succeed with this line of argumentation, though, are those who reject the idea that God ever first created humans (such as my account), or perhaps those who argue a for a type of Molinism and Middle Knowledge.

Consider Agent A who punches Agent B and knows 100% that that event will lead to Agent B punching Agent A back. The fact about knowledge does not eliminate Agent B's freedom. Although we may say that Agent A played a causal role in B punching A (in that it was a direct karmic consequence from A's initial punching that A had to suffer for), Agent B is still responsible for their punch so long as the action issued from their appropriate reasons-responsive mechanism. Both are morally responsible.

Returning now to Pereboom's four-case manipulation argument, I believe it's possible to justify Professor Plum's state of being responsible on cases one and two in the following way. When first approaching these cases, we may have formed the natural judgement that, on case one, although Plum does undergo a deliberation process before acting that is adequately reasons-responsive, and with adequate meshing between first and second-order desires, etc, this process, however, was manipulated by the neuroscientists who, by pressing a button, produced in him a particular neural state that lead to the murder of White. And thus, intuitively, we may have deemed Plum *not* morally responsible because he wasn't an adequate source of the neural state. This then leads to the difficult task of explaining how we could find Plum simultaneously *responsible* on case three – we are faced with the question: what is the relevant difference between the case of neuroscientists instilling in him a particular neural state that he is not a source of, and the case where Plum's parents and community produce in him a particular neural state that he is also not a source of?

At first glance, perhaps we will notice that on the first case, the difference is that the neuroscientists are in full knowledge of what will necessarily happen when they install the neural state, whereas on case three we have no indication that the parents possess the same type of providential foresight. However, as I have established in the discussion of Allison and Hal, knowing about *x* happening does not, by itself, cause *x* to happen. So long as Plum's reasons-responsive mechanism itself remained uncorrupted, then Plum always possessed the option of refraining from murdering White. Accepting that, perhaps we still wish to argue though that maybe Plum is not responsible on either case because he wasn't a proper source of the neural state that influenced his action; considering a Gaudiya Vaishnava worldview, though, Plum *can* be considered a source of the implanted neural state. We can easily theorise that this experience Plum has on case one of being implanted with a neural

state that eventuates in murdering White is a type of karmic consequence from a previous action or desire from Plum.

We could speculate, for example, that Plum in his previous life was a woman named Camille who manipulatively persuaded her lover Louis into murdering her husband Arthur for the selfish desire of Camille and Louis running away together; and therefore in Camille's next life she took birth as Plum who was manipulatively persuaded by the neuroscientists (via the implanted neural state) to kill Ms White – forced to suffer the similar experience of Louis of being so heavily influenced into committing murder. In this way, we can see how Plum is a source of the neural state implanted by neuroscientists, as the event was caused by the karma accumulated in his actions as Camille. Considering this sourcehood of influencing factors and Plum's possession of adequate reasons-responsiveness, we should consequently find no difficulty in deeming Plum morally responsible for his action on case one.

Reflecting now on case three, as I have already discussed in the *soft-line reply*, on this case we can understand that Plum, as an eternal spiritual soul, can play a casual role in the conditions, family, and environment that he is born into, whilst also retaining responsibility in how he reacts and chooses to operate within the conditions of life he faces through reasons-responsiveness. Using Pereboom's similarity claim, we can hopefully see case two also in a new light that no longer appears unintuitive to regard Plum as morally responsible. Indeed, the scenario given in case two is perhaps the most straightforward analogy for the conception of rebirth on the Gaudiya Vaishnava perspective: whereas on Pereboom's story 'a team of neuroscientists programmed him at the beginning of his life so that his reasoning is often but not always egoistic', this aligns perfectly with the Vaishnava philosophy that the consciousness that one develops in their lifetime is transferred over at the time of death and migration into a body: 'one surely attains the very state that one remembers while quitting the body… whatever it may be; for one has always caused that state of being to be' (Goswami, 2015, 8.6). Although this transfer and implantation of a disposition isn't done by neuroscientists according to Gaudiya Vaishnava theology, it is similar enough to be considered analogous and for us to see the potential explanation as to how Plum could be seen as an adequate source and thus, ultimately, morally responsible. And again using the similarity method, we could return to case one and view the actions of the neuroscientists implanting the neural state as a type of delayed transfer of the soul's disposition which

happens naturally at every birth for all souls, and therefore, should not affect our ability to consider Professor Plum morally responsible on all four cases.

What I have attempted to present here is a convincing *hard-line* reply to Pereboom's four-case manipulation argument that, I hope, successfully avoids Pereboom's objection and attempt to undermine a compatibilist (or semi-compatibilist) theory of free will and responsibility. In addition to this, I'm also convinced that this reply could be similarly presented as a successful response to other arguments from manipulation that have been presented in free will literature. Consider the following freedom-undermining manipulation scenario offered by A. Mele:

> Diana creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E. If this agent, Ernie, has any unsheddable values at the time, they play no role in motivating his A-ing. Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. (Mele, 2006).

Whilst other compatibilists may struggle with conceptualising a view of sourcehood that works to reject and affirm responsibility effectively, we can apply a similar method here as on the previous cases discussed. If we were to interpret 'Diana creates a zygote Z' to mean that Diana *brings about* the existence of the Zygote, and thus the zygote lacks an eternal soul, then we could argue that the zygote Ernie lacks sufficient moral responsibility due to not having any control over who he is or the circumstances he's placed in, and rather Diana holds the moral responsibility here. If, however, we were to interpret 'Diana creates a zygote Z' to mean that she makes an arrangement for an eternal soul to enter into this zygote, then it's possible for us to imagine this arrangement being a karmic consequence to a previous action of his, and thus consider him a partial source. And if we also understand that though Diana is responsible for the final event E because of her acting in full knowledge that if she did made the arrangement then E would occur, she is not *alone* responsible for E occurring

because Ernie still maintained his reason-responsiveness (or his position as a *reliable deliberator* as Mele describes) and is thus also culpable.

Where manipulation would genuinely undermine free will and moral responsibility would be in cases where the agent lacked a legitimate reasons-responsive process (or equivalent satisfactory process of deliberation). For example, a case where a hypnotist didn't just plant a desire in an agent's mind, but hijacks the reasoning process entirely by manufacturing the states and considerations entirely, would be an example where the agent lacks the ability to deliberate and thus lacks moral responsibility. However, following this understanding of karmic sourcehood, could it not be argued that the hypnotised agent still retains partial responsibility for whatever the hypnotist makes him do, as according to the doctrine of karma it was due to his previous actions that the hypnotised agent found himself having the experience? What relevance then is a reasons-responsive mechanism? Perhaps, what we may have encountered here is a type of damning conception of sourcehood that means we can never escape holding at least some little responsibility on every case if it's true that we are the ultimate source of all our actions. But surely we can imagine some cases in which this appears false and unfair. This line of questioning leads into a larger potential objection that the law of karma functions as a freedom-undermining deterministic mechanism, but due to the seriousness of the threat, I will return to discuss it more thoroughly in the following chapter.

For now however, I will summarise that though there are many objections that have yet to be raised, this chapter has laid out a basic framework for how proponents of the Gaudiya Vaishnava tradition can go about a reconciliation between responsibility and determinism, and subsequently also laid a path out for a compatibilist to strengthen their position, provided they adopt a few of the key features of Vaishnavism into their worldview – specifically, the idea of an eternally existing soul and its potential for reincarnation. What I hope to have established thus, is that, provided I can satisfactorily address the objections that will be raised in chapter five, there is a way to reply to Pereboom's manipulation argument such that it avoids the charge that responsibility is incompatible with determinism; and in doing so, I also proposed a way to approach Strawson's Basic Argument by suggesting a causal history involving reincarnation and the belief that there can exist, instead of a *causa sui*, a type of agent that is causeless and that lacks any original external influencing factor – an agent where all actions can be traced back to influencing factors that

are solely *internal* to the agent. In the following chapter, I will raise and reply to a few of the most obvious potential objections that threaten to undermine the potential of responsibility on my GV compatibilist theory of free will.

Chapter Five

Objections and Replies

It is to be expected that with an integration of Asian philosophy into a field that is classically Eurocentric, some friction and clashing will surely arise as the Eastern tradition attempts to play by foreign rules. However, it was certainly proven in the previous chapter that such an integration and meshing of ideas is both possible and arguably fruitful too. Therefore, to preserve the theory's fruitfulness, I shall attempt now to address some of the potential concerns over a Gaudiya Vaishnava compatibilist conception of free will that threatens to undermine its coherence.

At this point, it could be suspected that someone might wish to raise the point that whilst we may be able to identify cases where we completely lack any control over our actions and are subsequently not responsible for them, as we established in chapter four it is peculiar that there is no point at all where we are able to consider anybody as totally in control and thus completely responsible for their actions, as due to the finite nature of the soul we are bound to make mistakes and accidents, etc. Consequently, we may wish to ask questions of if it is *fair* that we operate within a world in which we aren't 100% 'truly responsible for the way you are' to use Strawson's words (1994, p. 13). Here I am edging closer to the territory of theodicy considering this GV view incorporates the existence of a personal and all-powerful God, but as I cannot here offer an entire reply to the problem of evil, I will try to just elucidate some relevant features of Gaudiya Vaishnava theology that may explain the fairness of our restrictive abilities.

So far, I have been discussing karma only in terms of its *retributive* effects, in how karma functions as a process of reward and suffering responding to people's actions and desires. Another aspect of karma, though, is its *reformative* effect.[26] The notion is that via this process of rebirth suffering karmic consequences, the experience of suffering and reward has a purificatory effect on the soul and the soul eventually develops a morally perfect disposition. We can see here karma functions as a part of a soul-making process,

[26] I make this distinction stemming from Gupta and Gallagher's point in *Reincarnation and Universal Salvation* that 'We maintain that these *karmic* mechanisms (1) administer justice by giving individuals consequences to their actions that are proportional to the moral quality of these actions and that (2) the suffering administered through these mechanisms serves a soul-making purpose' (Gupta & Gallagher, 2023, p. 114).

where suffering can be considered not just *fair* in how it stems from a deserving action, but also ultimately *fair* in how it purifies the soul from an original state of moral imperfection to one of pure perfection. Understanding karma's *reformative* mechanism, we can now consider how although a limitation of free will can lead to our suffering, on the larger picture this suffering is never unjustified as it is always serving the higher purpose of character reformation / soul-making. A chess player may lose a valuable piece in a foolish mistake, but the player learns not to make such a move again. I cannot here dive into a thorough defence of this type of soul-making theories (and it has already been done elsewhere), but I hope this brief clarification can lead us to understand why holding agents morally responsible despite never being 100% responsible for who they are can still be considered *fair* and *justified* on my GV view of free will.[27]

Another objection of *fairness* regarding the law of karma is the question of whether karma itself is compatible with moral responsibility. The objection is this: it's not fair nor just at all to hold people morally responsible for actions that they are performing whilst under the influence or determining effects of their *karma* or *guṇas*. Take for example some agent *M* who performs some action *Z* which leads to the karmic reaction that *M*, in their next life, will have a particular disposition *D*, and in this next life, *M*'s disposition *D* will influence him to perform an action *X.* Is it fair, then, to hold *M* responsible for action *X*? Even if the agent is an adequate *source* of *X*, it still appears that he is helpless in *X* occurring. On cases like these, could it not be argued that these agents are simply acting *not* freely, but just deterministically according to their previous karma?

To answer this, we must return back to the element of control within the conditions of responsibility. Although it is true that one's karma and guṇas can causally influence an agent's experiences and decisions, it is always *up to* the individual in how they act or respond to the karmic influence. Consider a game of poker: though the player is not responsible for the cards they are dealt, the player *is* responsible in how they play them – it is not that the cards determine how the player plays, though the cards may influence or limit the player's decision, it remains sufficiently *up to me* how to play my cards. Even if causal determinism obtains, the player remains responsible for how they play so long as the next

[27] For a defence of Gaudiya Vaishnava Soul-Making, see: Gupta, Akshay & Gallagher, Alex (2023). Reincarnation and Universal Salvation. *Faith and Philosophy: Journal of the Society of Christian Philosophers*.

move comes about from their reasons-responsive deliberation process. It is not that guṇas and karma skip over an agent's reasons-responsive mechanism, but the agent retains the ability to execute his willpower as he wishes: when an alcoholic is presented with a drink, although he will no doubt be influenced by the intense desire to drink, the alcoholic being an agent retains the reasons-responsive ability to choose whether to let this intense desire win, or to decide to refrain. However, the GV theory of karma is much fairer than depicted as in the poker analogy, as on our case, the agent is also the source responsible for which cards they are dealt – in how the circumstances that an agent is put in is traceable back to their previous actions. In this way we can see that karma does not function deterministically in a way that eliminates moral responsibility; karma should be understood as the cards that one is dealt, whereas *responsibility* lies in how one plays them. I should also note here, however, that not all actions are necessarily influenced by karma and guṇas; according the Gaudiya Vaishnava theology, as an agent becomes purified his actions gradually become less under the influence of past karmic activity, and become more under the influence of God. Ultimately, the goal for Gaudiya Vaishnavas is to transcend this cycle of saṃsāra – of repeated birth and death according to the agent's accumulated karma – and take up spiritual activities that are completely free from material influence. This, however, naturally leads us to another possible objection concerning the role of God within the Gaudiya Vaishnava philosophy.

As I touched on briefly in chapter two, Krishna (God) is described as playing a causal role in our actions with scriptural statements such as: 'The Lord of all beings resides in the region of the heart, Arjuna, making all beings wander as they ride in illusion on the [bodily] machine' (Goswami, 2015, 18.62) and 'indeed, whether a person acts on his own or is forced by others, he is always under the Supreme Lord's control' (Prabhupāda, 1983, 11.11.10). Of course, this appears as a contradiction to the GV compatibilist theory that I have described wherein it is the agent who is in control by virtue of their ability to deliberate and decide, and their possession of adequate sourcehood. However, not only is there a threat of theological inconsistency between Gaudiya Vaishnava philosophy and the free will conception I have presented, but more seriously, there is the possible objection that in my effort to reconcile causal determinism and free will, what I have done instead is simply deferred the threat to now face a new issue of *theological determinism* versus free will.

In response, though, I will repeat the answer I offered back in chapter two, that though it is true that Krishna can and does play a causally influencing role in our lives, in other places of scripture it is explained that Krishna '…acts after considering the jīva's efforts, the injunctions and prohibitions are not useless and the Lord is not liable to the charge of cruelty' (Brahma Sutras 2.3.40). Though there is an element of theological determinism present in the Gaudiya Vaishnava literature, it is not necessarily the case that this interferes with an agent's control to decide to act for themselves – Krishna acts according to the desires and decisions of the agent – as the saying goes: *Man proposes, God disposes*. Of course, this objection requires a far more detailed defence than what I can provide here, but I hope that what I have briefly mentioned indicates that the Gaudiya Vaishnava scripture does contain some potential for a lengthier and more convincing reply.

The final concern I will mention briefly pertains to why advocate for a compatibilist theory specifically. Could it not be plausibly argued that these concepts of *karma*, *reincarnation*, and an *eternal soul* could be easily construed to support a *Hard Determinist* or *Libertarian* theory of free will? Indeed I do, these concepts of *eternal souls*, *reincarnation*, and *karma* are not limited to a compatibilist approach. However, I do not have the space here to offer a Gaudiya Vaishnava analysis of *Hard Determinism* or *Libertarianism*, though there would certainly be plenty to consider. For now, where I believe my conception possesses an advantage, however, is how this conception allows one to remain undecided regarding the truth of determinism – the basis for identifying moral responsibility is independent from the truth of if the universe functions deterministically or not. Alike the semi-compatibilist position of Fischer and Ravizza, this theory does not stand the risk of being confirmed or rejected as science progresses towards a definitive answer. Similarly, as I raised in chapter two, it remains largely debatable whether the Gaudiya Vaishnava scripture portrays a deterministic or indeterministic account of the universe, meaning that this GV conception I have presented does not depend upon a truth of a particular side.

Whether determinism is true or not, if this source compatibilist conception is coherent, then the ability for agents to be morally responsible is unimpeded on either case – even if there exists indeterminism in the deliberation or reasons-responsive process, with the function of karma and reincarnation, the agent would still be an adequate source of *all* the possible outcomes, and therefore remain morally responsible for whichever outcome is actualised. In the case that my theory is unsuccessful, however, then we may have to

concede to a loss of moral responsibility on either a libertarian or hard determinist approach; but in developing a compatibilist conception, I have tried, in a sense, to take the hardest route first in order to 1. Preserve the most of our intuitive beliefs, and 2. Leave the door open on the debate of the truth of determinism.

Conclusion

As this study draws to a close, my hope is that it has become evident how the integration of Gaudiya Vaishnava theology can strengthen the case for a compatibilist, or semi-compatibilist, theory of free will and determinism. In particular, the success demonstrated in how the Gaudiya Vaishnava conception of an eternal soul that is capable of reincarnation can offer convincingly satisfactory account of sourcehood that is essential for the possession of responsibility over our actions.

As we found in chapter three, many compatibilist approaches, including theories from Frankfurt, Fischer and Ravizza, McKenna, and Haji, were all forced to either *bite the bullet* at a certain point, or leave their theory vague and incomplete as they encountered insurmountable objections – particularly, the compelling Four-Case Manipulation Argument from Pereboom that spotlighted the limitations for these compatibilist theories to offer a satisfactory account of sourcehood that would not be undermined by occasions of manipulation. However, in chapter four it was illuminated how much further and more convincing a compatibilist account of moral responsibility could be following the adoption of Gaudiya Vaishnava metaphysics, especially with regard to offering a satisfactory reply to Pereboom that, I believe, escaped the damning conclusion of his argument without sacrificing our intuitions or rationality.

Admittedly, however, there is still a major sense in which a compatibilist theorist would have to *bite the bullet* when accepting my view with regard to 1. Giving up a strictly physicalist stance of the universe – for the ideas of reincarnation and immaterial souls are not currently embraced by modern physics, and 2. Giving up the idea that any agent can possess 100% total control – for nothing can be a *causa sui*. For some like myself, though, these are hardly concessions that are tough to swallow, and importantly, these concessions do not demand drastic changes to our intuitive beliefs regarding our everyday perspective of moral responsibility.

The difficulty surrounding the reconciliation of free will and determinism has been pondered for millennia, and no doubt the debate will subsist into the future. At the very minimum, though, what I have proven here is that Eastern philosophy, specifically that stemming from the ancient Vedas, has a deep wealth and value to contribute to modern

philosophy that has not yet been tapped into and taken advantage of. As I have tried to illustrate in a *Gaudiya Vaishnava* reconciliation of free will and determinism, if we broaden our horizon to include the intriguing worldviews of elsewhere, we may be surprised in how well we are able to surpass previous deadlocks and limitations.

Bibliography

Bourget, D., & Chalmers, D. (2023). PHILOSOPHERS ON PHILOSOPHY: THE 2020 PHILPAPERS SURVEY. *Philosophers' Imprint*, *23*(1), 1–53. https://doi.org/10.3998/phimp.2109

De Marco, G. (2023). Manipulation, machine induction, and bypassing. *Philosophical Studies*, *180*(2), 487–507. https://doi.org/10.1007/s11098-022-01906-2

Easwaran, E. (2007). *The Upanishads* (2nd ed.). Nilgiri Press.

Fischer, J. M. (1982). Responsibility and Control. *The Journal of philosophy*, *79*(1), 24–40. https://doi.org/10.5840/jphil198279159

Fischer, J.M. (2002). Frankfurt-type examples and semi-compatibilism. In Robert Kane (ed.), *The Oxford Handbook of Free Will* (pp. 281-308). New York: Oxford University Press.

Fischer, J. M. (2004). Responsibility and Manipulation. *The Journal of Ethics*, *8*(2), 145–177. https://doi.org/10.1023/B:JOET.0000018773.97209.84

Fischer, J.M. (2007). Compatibilism. In, ed. Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M (Eds.), in *Four Views On Free Will* (pp. 44-84). Blackwell Publishing.

Fischer, J. M. (2023). *My way : essays on moral responsibility* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780195179552.001.0001

Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control : a theory of moral responsibility*. Cambridge University Press.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829–839. https://doi.org/10.2307/2023833

Frankfurt, H. G. (1988). *The importance of what we care about : philosophical essays*. Cambridge University Press.

Frankfurt, H. G. (2002). Reply to John Martin Fischer. In, ed. Buss, S., & Overton, L. *The Contours of Agency: Essays on Themes from Harry Frankfurt*. The MIT Press. https://doi.org/10.7551/mitpress/2143.001.0001

Goswami, H.D.. (2015). *A Comprehensive Guide to Bhagavad-Gita with Literal Translation* (1st ed). Krishna West Inc.

Gupta, Akshay. (2020). Re-envisioning a Caitanya Vaiṣṇava 'Perfect Being Theology' and Demonstrating Its Theodical Implications. *Journal of Hindu-Christian Studies,* (33):42-52.

Gupta, Akshay & Gallagher, Alex (2023). Reincarnation and Universal Salvation. *Faith and Philosophy: Journal of the Society of Christian Philosophers*, Vol. 40: Iss. 1, Article 6. DOI: 10.37977/faithphil.2023.40.1.6

Gupta, R. M. (Ravi M. (2007). *The Caitanya Vaisnava Vedanta of Jiva Gosvami : when knowledge meets devotion*. Routledge. https://doi.org/10.4324/9780203500682

Haji, I. (2023). *Moral appraisability : puzzles, proposals, and perplexities*. Oxford University Press. https://doi.org/10.1093/oso/9780195114744.001.0001

Haji, I., & Cuypers, S. E. (2006). Hard- and soft-line responses to Pereboom's four-case manipulation argument. *Acta Analytica : Philosophy and Psychology*, *21*(4), 19–35. https://doi.org/10.1007/s12136-006-0002-5

Haji, I., & Cuypers, S. E. (2007). Magical agents, global induction, and the internalism/externalism debate. *Australasian Journal of Philosophy*, *85*(3), 343–371. https://doi.org/10.1080/00048400701571602

Hoefer, Carl. (2023). "Causal Determinism", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition). Edward N. Zalta & Uri Nodelman (eds.). https://plato.stanford.edu/archives/win2023/entries/determinism-causal/

Johnson, Todd M & Grim, Brian J. (2013). *The World's Religions in Figures: An Introduction to International Religious Demography*. John Wiley & Sons. ISBN 978-1-118-32303-8.

Kane, R. (1996). *The Significance of Free Will.* New York, US: Oxford University Press USA.

Kane, R. (2005). *A contemporary introduction to free will*. Oxford University Press.

Kane, R. (2007). Libertarianism. In, ed. Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M (Eds.), in *Four Views On Free Will* (pp. 5-43). Blackwell Publishing.

Kane, R. (2017). Responsibility, indeterminism and Frankfurt-style Cases: A reply to Mele and Robb. In *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (pp. 91–106). United Kingdom: Taylor & Francis Group.

Kim, J. (1973). Causes and Counterfactuals. *The Journal of Philosophy*, *70*(17), 570–572. https://doi.org/10.2307/2025312

Maharaj, Ayon. (2018). Hard Theological Determinism and the Illusion of Free Will: Sri Ramakrishna Meets Lord Kames, Saul Smilansky, and Derk Pereboom. *Journal of World Philosophies* 3 (2):24-48.

McKenna, M. (2008). A Hard-line Reply to Pereboom's Four-Case Manipulation Argument. *Philosophy and Phenomenological Research*, *77*(1), 142–159. https://doi.org/10.1111/j.1933-1592.2008.00179.x

McKenna, M. and D. Justin Coates. (2024a). *Compatibilism: State of the Art*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/compatibilism/supplement.html

McKenna, M. and D. Justin Coates. (2024b). *Compatibilism*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/compatibilism/#ContComp

McKenna, M. and Pereboom, D. (2016). *Free Will: A Contemporary Introduction (Routledge Contemporary Introductions to Philosophy)*. Routledge

Mele, A.R. (2006) *Free Will and Luck*. Oxford University Press

Menzies, Peter and Helen Beebee. (2024). *Counterfactual Theories of Causation*. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/causation-counterfactual/#CouCauDep

O'Connor, Timothy and Franklin, Christopher. (2022). *Free will*. *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/win2022/entries/freewill/

Overbye, Dennis. (2007). *Free Will: Now You Have It, Now You Don't*. The New York Times https://www.nytimes.com/2007/01/02/science/02free.html

Pereboom, D. (2001). *Living without free will*. Cambridge University Press.

Pereboom, D. (2007). Hard Incompatibilism. In, ed. Fischer, J. M., Kane, R., Pereboom, D., & Vargas, M (Eds.), in *Four Views On Free Will* (pp. 85-125). Blackwell Publishing

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life* (1st ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685516.001.0001

Prabhupāda, A. C. Bhaktivedanta Swami. (1983). *Bhagavad-gītā as it is : with the original Sanskrit text, roman transliteration, English equivalents, translation, and elaborate purports* (2nd ed., rev.enl.). Bhaktivedanta Book Trust.

Prabhupāda, A. C. Bhaktivedanta Swami. (1995). *Śrīmad Bhāgavatam.* Bhaktivedanta Book Trust

Schweig, G. M. (2024). *Bhagavad Gītā Concordance: A Comprehensive Word Reference with English and Sanskrit Indexes*. Columbia University Press. https://doi.org/10.7312/schw14132

Silvestre, R. S. (2017). Karma Theory, Determinism, Fatalism and Freedom of Will. *Logica Universalis*, *11*(1), 35–60. https://doi.org/10.1007/s11787-016-0154-z

Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies*, *75*(1/2), 5–24. https://doi.org/10.1007/BF00989879

Strawson, P.F. (1962). Freedom and Resentment. *Proceedings of the British Academy.* 48:187-211.

Swartz, Norman. (1997). *The Concepts of Necessary Conditions and Sufficient Conditions*. Simon Fraser University. https://www.sfu.ca/~swartz/conditions1.htm

Thākura, Śrīla Bhaktivinoda. (2019). *Jaiva Dharma: Two tales of spiritual seekers* (Swami, HH Bhanu, Trans.). Tattva Cintāmaṇi Publishing. (Original work published 1896)

Timpe, Kevin. (2008). *Free will sourcehood and its alternatives*. Continuum.

Van Inwagen, Peter (1983). *An Essay On Free Will*. Oxford: Clarendon Press.

Vidyābhūṣaṇa, Śrīla Baladeva. (2018). *Brahma Sūtras: With Govinda-bhāṣya commentary of Baladeva Vidyābhūṣaṇa*. (Swami, HH Bhanu, Trans.) Tattva Cintāmaṇi Publishing. (Original work published 1718)