UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

GABRIEL DE ANDRADE PAGNOZZI MARUCHI

# RESPONSIBILITY AND CAUSATION: FROM EXTENSIONALITY AND CAUSAL SELECTION TO THE FOUR-CASE ARGUMENT

# RESPONSABILIDADE E CAUSAÇÃO: DA EXTENSIONALIDADE E SELEÇÃO CAUSAL PARA O ARGUMENTO DOS QUATRO CASOS

CAMPINAS
2024

GABRIEL DE ANDRADE PAGNOZZI MARUCHI

# RESPONSIBILITY AND CAUSATION: FROM EXTENSIONALITY AND CAUSAL SELECTION TO THE FOUR-CASE ARGUMENT

# RESPONSABILIDADE E CAUSAÇÃO: DA EXTENSIONALIDADE E SELEÇÃO CAUSAL PARA O ARGUMENTO DOS QUATRO CASOS

Dissertação apresentada ao Instituto de Filosofia e Ciências Humanas da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Filosofia.

Dissertation presented to the Institute of Philosophy and Human Sciences of the University of Campinas in partial fulfillment of the requirements for the degree of Master in the area of Philosophy.

*Orientador:* Prof. Dr. Marco Antonio Caron Ruffino
*Coorientador* Prof. Dr. Pedro Merlussi

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO DEFENDIDA PELO ALUNO GABRIEL DE ANDRADE PAGNOZZI MARUCHI, E ORIENTADA PELO PROF. DR. MARCO ANTONIO CARON RUFFINO.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Filosofia e Ciências Humanas
Neiva Gonçalves de Oliveira - CRB 8/6792

M368r

Maruchi, Gabriel de Andrade Pagnozzi, 2000-
Responsibility and causation : from extensionality and causal selection to the Four-Case Argument / Gabriel de Andrade Pagnozzi Maruchi. – Campinas, SP : [s.n.], 2025.

Orientador: Marco Antonio Caron Ruffino.
Coorientador: Pedro Merlussi.
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Filosofia e Ciências Humanas.

1. Responsabilidade . 2. Causalidade. 3. Livre arbítrio e determinismo. I. Ruffino, Marco, 1963-. II. Merlussi, Pedro. III. Universidade Estadual de Campinas (UNICAMP). Instituto de Filosofia e Ciências Humanas. IV. Título.

Informações complementares

**Título em outro idioma:** Responsabilidade e causação : da extensionalidade e seleção causal para o Argumento dos Quatro Casos
**Palavras-chave em inglês:**
Responsibility
Causality
Free will and determinism
**Área de concentração:** Filosofia
**Titulação:** Mestre em Filosofia
**Banca examinadora:**
Marco Antonio Caron Ruffino [Orientador]
Beatriz Sorrentino Marques
Fabio Dal Conti Lampert
**Data de defesa:** 28-02-2025
**Programa de Pós-Graduação:** Filosofia

**Objetivos de Desenvolvimento Sustentável (ODS)**
Não se aplica

**Identificação e informações acadêmicas do(a) aluno(a)**
- ORCID do autor: https://orcid.org/0000-0001-7760-5903
- Currículo Lattes do autor: http://lattes.cnpq.br/3909796174697261

**UNIVERSIDADE ESTADUAL DE CAMPINAS**
**INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS**

A Comissão Julgadora dos trabalhos de Defesa de Dissertação de Mestrado, composta pelos Professores Doutores a seguir descritos, em sessão pública realizada em 28/02/2025, considerou o candidato Gabriel de Andrade Pagnozzi Maruchi aprovado.

Prof. Dr. Marco Antonio Caron Ruffino
Profa. Dra. Beatriz Sorrentino Marques
Dr. Fabio Dal Conti Lampert

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertações/Teses e na Secretaria do Programa de Pós-Graduação em Filosofia do Instituto de Filosofia e Ciências Humanas.

# Acknowledgements

Any merit this dissertation may have is evidence of the amazing people I have near me more than anything else. Without their support – both academically and personally –, these last two years would not have been as fruitful – both academically and personally – as they were. Then, I should express my gratitude.

I am thankful for the supervision provided by Professor Marco Ruffino and Professor Pedro Merlussi. During the two years of the Master's, they helped me navigate many aspects of philosophy, the discipline, and philosophy, the profession. Thanks to them, I am closer to becoming a professional philosopher than I was two years ago. I also thank Beatriz Sorrentino Marques and Fábio Lampert for accepting the invitation to be part of the qualifying examining board and the final examining board for the Master's. Their comments on earlier versions of the dissertation really helped improve it.

The companionship of Alan René, Alex Bispo, Ariston Sócrates Guimarães, Angie Paola Hugueth Vásquez, Daniel Diniz, Deiver de Melo, Gabriel Eugênio, João Felipe Rasi, Mahan Vaz, Miguel Nunes, Rafael Ongaratto, Rafael Souza, Renato Valvassori, and Sofia Abelha was essential in these two years, both as friends and as fellow philosophers. I also thank Gabriel, João, and Renato for reading early drafts of some parts of the dissertation. I thank my parents for their unrelenting support during these last two years.

Foremost, I have to thank Mariana Fernandes for being by my side these two years. She provided a lot of insightful philosophical discussions that helped shape my philosophical views and the arguments present in this dissertation. More than that, however, she was always there for me. The Master's would have been way more stressful without her to take my mind away from it when needed. She filled these two years – that could have been simply a stressful period – with precious memories. For that, and everything else, I am extremely thankful.

Last but not least I express my gratitude to the staff of the Octavio Ianni Library and the Philosophy Graduate Studies Office for helping deal with the Master's bureaucracy.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES)- Finance Code 001.

# Resumo

O Argumento dos Quatro Casos, proposto por Derk Pereboom, é um argumento da manipulação pelo incompatibilismo, a tese de que responsabilidade moral e determinismo são incompatíveis. Neste trabalho, eu argumento contra o Argumento dos Quatro Casos a partir de certas teorias específicas sobre causação. Para tanto, eu discuto os debates sobre causação que motivam essas teorias. Após apresentar o Argumento dos Quatro Casos, a dissertação se divide em duas partes. O ponto de partida da primeira parte é o debate sobre extensionalidade causal. Entre as visões motivadas por essa discussão, encontramos o contrastivismo, segundo o qual relações causais não são binárias. Assumindo o contrastivismo, eu proponho uma definição de Sorte-C, um tipo de sorte moral. Com a noção de Sorte-C em mãos, defendo que o primeiro caso do Argumento dos Quatro Casos não motiva sua conclusão incompatibilista. O agente manipulado do primeiro caso está sujeito a Sorte-C, e isso impede o caso de motivar a conclusão incompatibilista. A segunda parte da dissertação foca na seleção causal. Seleção causal é um fenômeno encontrado em nossas atribuições ordinárias de causação. Quando buscamos pela causa de um evento, não nos limitamos às condições metafísicas que filósofos tradicionalmente associaram à causação, como dependência contrafactual e processos físicos. Normalidade estatística e normas morais também afetam nossos julgamentos sobre causação. Esse fenômeno motiva a tese de que normalidade tem um papel em determinar fatos causais. Assumindo essa ideia, eu proponho uma teoria compatibilista que oferece uma resposta ao Argumento dos Quatro Casos.

**Palavras-chave**: Responsabilidade moral; Causação; Argumento da Manipulação; Determinismo.

# Abstract

The Four-Case Argument, put forward by Derk Pereboom, is a manipulation argument for incompatibilism, the thesis that moral responsibility and determinism are incompatible. In this work, I argue against the Four-Case Argument taking as a starting point certain theories about causation. With this end, I will discus the debates about causation that motivate those theories. After presenting the Four-Case Argument, the dissertation divides into two parts. The starting point of the first part is the debate about causal extensionality. Among the views motivated by this discussion, we find contrastivism, according to which causal relations are not binary. Assuming contrastivism, I propose a definition of C-Luck, a kind of moral luck. With the notion of C-Luck at hand, I defend that the first of the Four-Case Argument does not motivate its incompatibilist conclusion. The manipulated agent in the first case is subjected o C-Luck, and that precludes the case of motivating an incompatibilist conclusion. The second part of the dissertation focuses on causal selection. Causal selection is a phenomenon found in our ordinary attributions of causation. When we search for the cause of an event, we do not limit ourselves to the metaphysical conditions that philosophers traditionally associated with causation, like counterfactual dependence and physical processes. Statistical normality and moral norms also affect our causal judgments. This phenomenon motivates the thesis that normality plays a role in determining causal facts. Assuming this idea, I propose a compatibilist theory that offers a reply to the Four-Case Argument.

**Keywords**: Moral Responsibility; Causation; Manipulation Argument; Determinism.

# CONTENTS

# Introduction

A perennial question in the metaphysics of moral responsibility is whether causal determinism precludes moral responsibility. The two main positions are compatibilism and incompatibilism. Compatibilists think it does not; determinism and responsibility can co-exist according to compatibilists. Incompatibilists think that causal determinism does preclude moral responsibility. One popular argument for incompatibilism is the Four-Case Argument, put forward by Derk Pereboom (2001; 2007; 2014). I will argue against the Four-Case Argument in this dissertation.

In my attack against the Four-Case Argument, I will focus on a specific idea central to it. Pereboom presents four actions, each with its specific causal history. He says that these histories do not differ in a way that matters for attributions of moral responsibility. I will scrutinize this claim. For that, I will assume some specific theories about causation. Hence, I will also present these theories and the problems that motivate them.

## Overview

The first chapter sets the stage. I will introduce the notion of moral responsibility and the problem of determinism. After that, I will present the Four-Case Argument, the main focus of this dissertation. After a brief discussion of the argument's structure and some compatibilist replies, I will explain how causation, the other focus of this dissertation, relates to the Four-Case Argument.

The rest of the dissertation is divided into two parts. In each, I will go from a discussion about causation to an argument against the Four-Case Argument. In the first part, I will start with the debate about the extensionality of causation (chapter 2). One of the positions in this debate is contrastivism. With this theory in hand, I will argue that the first case of the Four-Case Argument has a problem with luck (chapter 3). This problem cannot be solved with a reformulation of the case. Hence, the incompatibilist has to abandon the first case. I will defend that this degrades the incompatibilist's argument.

The second part follows the same structure. In chapter 4, I will discuss causal selection, a phenomenon we find in our ordinary causal ascription. I will present Peter Menzies's (2007; 2009) theory of causation. This theory tries to account for causal selection. In chapter 5, I will propose a reply against the Four-Case Argument based on his theory of causation.

In the final chapter, I will try to paint the bigger picture by discussing what my arguments show about the relation between causation and the problem of determinism. I will also point out future work to be done.

# 1 The Four-Case Argument

In this chapter, I will present the main argument I will discuss during the dissertation: the Four-Case Argument. It is an argument in favor of incompatibilism, the view that moral responsibility and determinism are incompatible. Hence, I will start with an introduction to moral responsibility and the problem of determinism (section 1). This will help to locate the Four-Case Argument in the debate. After presenting the argument (section 2), I will discuss its structure (section 3). I will then discuss how compatibilists should reply to it (section 4). I will present some compatibilist replies and briefly argue for the kind of strategy I will be following throughout the dissertation. Last, I will discuss how the Four-Case Argument is connected to causation (section 5).

## 1.1 Moral Responsibility and determinism

The expression "is responsible for" has many meanings. Sometimes, it attributes a causal role to an entity: "The traffic jam was responsible for our delay". In other cases, we use it to attribute some duties to an individual: "João was responsible for getting us there in time". Besides these uses, we say things like "João is responsible for our delay". This sentence seems to have the same causal role attribution as the first, but it goes beyond that. If we believe that João is responsible for our delay, we have a whole range of attitudes about him — we demand an apology or feel resentment, for example — that we do not have about the traffic jam, despite both bringing about similar delays. In other words, we can hold João accountable for our delay. João is responsible for the delay in a way only agents, i.e., beings capable of acting, can be. This kind of responsibility is also harder to achieve then the others. One may be causally responsible without being morally responsible. If João has an unexpected nervous breakdown while driving, he may be the cause of our delay, but he is not morally responsible for it. We call this sense of responsibility "moral responsibility."[1] This is the sense of responsibility I will discuss in this dissertation.

Moral responsibility pervades our lives. We take pride in our achievements so far we feel responsible for them. We believe we should condemn criminals only if they are responsible for their crimes. We feel sad or angry when bad things happen to us, but there is a myriad of emotions and attitudes that we tend to have only when someone is responsible for these bad things. The actions of people dear to us make us feel loved insofar as we believe these actions are not mere accidents but expressions of their feelings. When someone fails us but is not responsible,

[1] In this introduction to the concept of moral responsibility, I assumed that there is only one sense for the term. However, a popular position in the literature is that there is more than one kind of moral responsibility. We find this position, for example, in McKenna (2012), Shoemaker (2015) and Mason (2019). If there is more than one kind of moral responsibility, the term "moral responsibility" — in this dissertation — refers to the kind of responsibility discussed in the literature about the problem of determinism.

we usually excuse the person; if someone is to blame for causing us harm, however, we adapt our relationship with them accordingly. It was no exaggeration when Peter Strawson affirmed that " in the absence of any forms of these attitudes it is doubtful whether we should have anything that we could find intelligible as a system of human relationships, as human society." (Strawson, 1974, 26).

Strawson's essay *Freedom and Resentment* reshaped the contemporary debate about moral responsibility. He brought attention to the feelings and attitudes that constitute our responsibility practices, which he labeled "reactive attitudes". He listed as examples of these "gratitude, resentment, forgiveness, love, and hurt feelings" (Strawson, 1974, 5). Even philosophers who do not accept his more controversial theses accept the general framework of moral responsibility put forward by him.

For this dissertation, we do not need a fleshed-out theory of moral responsibility. Taking the Strawsian understanding of our practices as a background, we can assume the following characterization of moral responsibility.

**Moral Responsibility** An agent $S$ is morally responsible for $X$ if, and only if, someone (possibly $S$ herself) is justified in responding to $S$ with a subset of react attitudes about $X$.

I will assume this because something like it is commonly accepted in the literature. For instance, we find this biconditional, or something similar, in McKenna (2012), Shoemaker (2015), Mason (2019), Pereboom (2014), Fischer & Ravizza (1998).

Responsibility is a cornerstone of our daily lives, but it is not always present. Sometimes we are not responsible. In these cases, we have a defense, an explanation for our lack of responsibility. Usually that can come in one of two ways. We may have an excuse or an exemption. An excuse shows that the agent, despite being accountable for her actions in general, did not satisfy the conditions for responsibility in this particular case. An exemption points out a problem with the person's agency that makes her unable to be accountable for her actions in general.

Despite their frequency, both excuses and exemptions are, in a sense, exceptional in our daily practices. We use them to deal with cases that deviate from what we take to be usual. Usually, people have the agency required to be responsible and most of their actions express that. Nobody has an excuse for all her actions, and only a few people are exempted.

The possibility of a universal defense threatens this *status quo*. If every action is excusable or every person is exempted, then our ordinary responsibility practices vanish. Nobody is responsible for her actions. Strawson argued that universal defenses are impossible; the exceptionality of defenses is not by chance (Strawson, 1974). Despite Strawson broad framework being widely accepted, his arguments for the impossibility of universal defenses are not equally popular. Philosophers discussing the metaphysics of moral responsibility still discuss if there is a universal defense.

Causal determinism is a strong candidate for feature of the world that could justify a universal defense. Many philosophers throughout history thought that determinism is incompatible

with the level of control needed for us to be responsible for our actions. We find this worry about determinism in Western philosophy since the stoics (O'Keefe, 2016). The problem of determinism is the problem of establishing whether these things are incompatible after all. Before discussing it, I need to make clear what is determinism.

Determinism is a trick concept to define because we may have two competing desiderata when trying to do it. We may want to better understand what is for a scientific theory to be deterministic. Or we may want to better understand what is the thesis that makes us worry about moral responsibility. These two things may come apart. We may have a good understanding of determinism in philosophy of science that does not do justice to the problem of determinism. Another possibility is that the concept of determinism relevant to the problem of determinism is distinct from our scientific view of determinism. Amid this problem of how to understand determinism, I will take the easiest route. I will assume the view of determinism accepted by Derk Pereboom. I want to discuss his argument, so I should discuss it in his terms wherever I can. Pereboom focuses on the idea of an action being causally determined by factors beyond the agent's control. He says:

> An action will be causally determined in this way if causally relevant factors occurring or active prior to the agent's coming to be, and thus beyond her control, together with facts about the laws of nature, also beyond her control, ensure the occurrence of the action by a causal process that begins with those preceding causal factors and ends in the occurrence of the action. (Pereboom, 2014, 1)

With this definition, he seeks to avoid counterexamples put forward by Joseph Campbell (2007, 2010) against the more traditional understanding of determinism in the literature about the problem of determinism. Campbell pointed out that an agent without past (or that had control over all the past) does not have his responsibility threatened by determinism. It is then usual to specify now that the agents being considered in the problem of determinism have a past that they are not responsible for. Pereboom is not giving a definition of determinism, but only of causally determined action. For our purposes here, this will be enough. We will work with the following definition of causally determined action:

**Causally determined action** *S* doing *A* is a causally determined action if, and only if, there are factors *F* outside *S*'s control such that (i) there is a causal chain starting with *F* that ensures the occurrence of *A*, and (ii) *S* is not responsible for the laws of nature that link *F* to *A*.

When talking about "determinism" or "deterministic world", I will just mean a world where every action is causally determined.[2]

Now that we have a notion of determinism, let us see why it can make us worry about moral responsibility. There are two ways to develop the apparent incompatibility of determinism and moral responsibility. These two ways track two distinct understandings that we may have of responsible agency (Timpe, 2016).

On the one hand, we can focus on the idea of *alternatives*. Intuitively, someone is responsible for acting a certain way only if she could have acted some other way. If our actions

[2] For a standard definition of of determinism, see van Inwagen (1983, 65) and Earman (1986, 13). For a criticism of using such definitions when discussing the problem of determinism, see Steward (2021).

are consequences of the past and the laws of nature – things outside our control –, we cannot do otherwise, and therefore, we are not responsible. Loosely speaking, the idea here is that we do not have alternatives if determinism is true because the past and the laws already "decided" our choices.

On the other hand, we can focus on the idea of *source*. It seems that, to be responsible for our actions, we need to be their sources. If determinism is true, however, the past and the laws of nature are the source of our actions, and hence, we are not responsible. When we ponder the truth of determinism, we get the impression that our actions would be just the unfolding of the past and the laws to which we contribute nothing; we would not be the source of our actions.

I presented two ways of developing the *prima facie* incompatibility of determinism and moral responsibility: one focused on the idea of alternative, and the other focused on the idea of source. These views about responsible agency may seem complementary. For us to really choose between alternatives, we must be the source of this choice; for us to be the source of an action, we need to select it among alternatives. If that is the case, compatibilists must show that these two aspects of responsible agency are safe from the apparent threat of determinism.

However, many members of the debate are convinced that alternatives are not necessary for moral responsibility. The main reason to reject the relevance of alternatives is the Frankfurt Cases. First proposed by Henry Frankfurt (1969), these cases feature an agent that intuitively does not have an alternative but still is responsible for her action. The following is a Frankfurt Case:

**Brain Chip** Assassin is planning to kill Victim. Scientist also wants the Victim's death but would prefer not to get her hands dirty; if Assassin kills Victim after all, the better for Scientist. Because of that, Scientist installs a chip in Assassin's brain so that, if she hesitates in killing Victim, Scientist will press a button that activates the chip and makes Assassin commit the murder. In the end, however, Scientist does not need to interfere; Assassin kills Victim for her own reasons.[3]

In this scenario, Assassin does not have a choice besides killing Victim. Nonetheless, intuitively, Assassin is morally responsible for the murder. This scenario, therefore, seems to show that alternatives are not necessary for being morally responsible. The Frankfurt Cases spawned a huge literature.[4] The details of this discussion are not relevant to this dissertation. What is important for us is that Frankfurt split the debate around the problem of determinism with his cases.

Those unconvinced by the Frankfurt Cases keep debating whether determinism precludes alternatives. The two positions in this discussion are leeway incompatibilism –

[3] One discussion I am side-stepping here is whether a successful Frankfurt Case needs to describe an indeterministic world. Some have argued that, if it does not happen in a indeterminsitic world, the argument begs the question against incompatibilism by assuming that one can be morally responsible in a deterministic world (Widerker, 1995). For a defense of Frankfurt Cases in deterministic worlds, see Fischer (2010).

[4] See Sartorio (2017) for an introduction to this debate.

alternatives are relevant to responsibility and incompatible with determinism – and leeway compatibilism – alternatives are relevant to responsibility and compatible with determinism –. Among leeway incompatibilits, we find Peter van Inwagen (1983); among leeway compatibilits, David Lewis (1981; 2020) and Kadri Vihvelin (2013).

Those convinced by the Frankfurt Cases put aside questions about alternatives and focus on debating whether determinism precludes the kind of sourcehood of actions necessary for moral responsibility. *Source incompatibilists* think it does; *Source compatibilists* think we can be the source of our actions even if determinism is true. Robert Kane (1996a; 1996b) and Derk Pereboom (2001; 2007; 2014) are examples of source incompatibilists; John Martin Fischer & Mark Ravizza (1998), Michael McKenna (2001; 2008a; 2008b), and Carolina Sartorio (2013; 2016) are examples of source compatibilists.

The Four-Case Argument enters the scene in this debate about sourcehood. Pereboom puts forward this argument to defend that the kind of sourcehood needed for moral responsibility is incompatible with determinism. The next section shows how he does it.

## 1.2 The Four-Case Argument

In a manipulation argument for incompatibilism, we start with a scenario where an agent is manipulated in such a way that she is not responsible for her action. The incompatibilist then argues that there is no difference between this case and an ordinary agent performing an action in a deterministic world. Hence, no ordinary agent in a deterministic world is responsible for her actions, i.e., incompatibilism is true. Manipulation arguments are the main problem source compatibilists have to handle. Do not take my word for it. Here is Fischer, one of the leading proponents of source compatibilism, talking about manipulation: "I think that manipulation cases are compatibilism's dirty little secret. Compatibilists don't like to admit that this is a problem. [...] We compatibilists have to deal with this." (Fischer, 2000, 390). We can trace the use of manipulation arguments at least back to Richard Taylor (1974, 46-47). Manipulation arguments (and compatibilist theories) have come a long way since then. A prominent manipulation argument now is the Four-Case Argument, put forward by Derk Pereboom (2001; 2007; 2014). This section is dedicated to this argument.

As the name makes clear, the Four-Case Argument consists of four cases. In each, we have an agent, Plum, who is causally determined to make the decision – and to act accordingly – to kill White in order to obtain a personal advantage. The first case features a manipulation case that elicits the intuition that Plum is not responsible. The other cases are modifications of the first, being the fourth case one of an ordinary agent in a deterministic world. The idea here is that our non-responsibility judgment about the first Plum generalizes to the other cases. Given that the common element in the four cases is determinism, we can infer by the best explanation that moral responsibility is incompatible with determinism. I will follow the presentation of the argument given in Pereboom (2014, Chapter 4).

In all cases, Plum satisfies conditions that compatibilists have proposed as necessary for moral responsibility. Pereboom points out that it is not enough for compatibilists to propose necessary conditions; they also need to argue that some conditions are jointly sufficient for responsibility (74). Therefore, Pereboom is entitled to assume that compatibilists are committed to the thesis that an agent satisfying all necessary conditions proposed by main compatibilist theories is morally responsible.

For the argument to succeed, Plum must satisfy all proposed compatibilist conditions for responsibility. Suppose that manipulated Plum does not fulfill a condition for responsibility that can be satisfied by an agent in a deterministic world. In this case, compatibilists have a straightforward reply to the argument: manipulated Plum is not responsible because he does not satisfy this condition. Let us get to the conditions (75). His action is not out of character; he does not act from an irresistible desire; Plum's desire to kill is in accordance with his second-order desires, i.e., he wills his desire to murder. He is reason-responsive and has the general ability to grasp, apply, and regulate his actions by moral reasons. As we will see, however, all these conditions still leave room for manipulation.

Let us start with the first case, a case of direct manipulation:

> **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. (76-77)

As said, Plum-1 (i.e., Plum of the first case) satisfies all compatibilist conditions presented. Nevertheless, the intuition here is that Plum is not morally responsible. These conditions, hence, are not sufficient for being morally responsible. Case 1 in itself is a counterexample to main compatibilist theories. There is a further condition for responsibility.

Pereboom seeks an incompatibilist conclusion: Plum-1 is not responsible because his decision is causally determined by factors beyond his control. The missing condition for responsibility is the absence of determination. Pereboom tries to get this conclusion through an inference to the best explanation: the best explanation for Plum's non-responsibility is that one is not responsible if one's action is determined by factors beyond one's control. The other cases aid in reaching this conclusion. We will gradually eliminate aspects of the scenario that are allegedly

irrelevant to responsibility ascription until we have only determinism playing a role. Let us see how this unfolds. Here is the second case, a case of more indirect manipulation:

> **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). (77)

According to Pereboom, intuitively, Plum-2 is also not morally responsible. We have yet another counterexample to the compatibilist theories. Not only that, but Pereboom argues that our judgment of the first case should carry on to the second case. The difference between the two cases is the temporal distance between the manipulator's action and Plum's decision. This difference, however, seems irrelevant to moral responsibility. Hence, if Plum-1 is not responsible, neither is Plum-2.

Now, the third case, a case of cultural indoctrination:

> **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. (78)

Case 3 differs from the last two cases in two important ways. First, it is more realistic. We have no history of neurological manipulations, but the kind of indoctrination happening in Case 3 is not unheard of. Second, here, the intuition about the case gets muddier. Pereboom does not say that the intuition is that Plum-3 is not responsible. Instead, he argues that there is no difference relevant to attributions of responsibility between Case 2 and Case 3. In Case 2, we had neurological manipulation; in Case 3, social manipulation. How the manipulation occurs does not affect its responsibility-undermining characteristic. Hence, Pereboom argues, our judgment that Plum-2 is not responsible should make us conclude that Plum-3 is also not responsible.

Finally, we get to the fourth case, a case of an ordinary agent in a deterministic world:

> **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. (79)

Pereboom says that, if Plum is not responsible in Case 3, he is also not responsible in Case 4. The difference between these cases is that, in the former, other agents play a role in determining Plum's action; that does not happen in the latter. However, according to Pereboom, this difference is irrelevant to attributions of responsibility. Hence, we should conclude that Plum-4 is not morally responsible for killing White. Why is Plum, in all these cases, not responsible for his actions? The best explanation for it, concludes Pereboom, is that, in each case, there are factors outside Plum's control that causally determine his actions. If this argument is sound, source incompatibilism is correct; determinism precludes agents from being the source, in the sense relevant to moral responsibility, of their actions.

## 1.3 The structure of the argument

So far, I have presented Pereboom's Four-Case argument. I will now discuss the structure of the argument. A usual way of schematizing the argument – and manipulation arguments in general – is the following:[5]

**P1** Plum-1 is not responsible for his actions in Case 1; (Manipulation Premise)

**P2** There is no difference relevant to attributions of moral responsibility between Plum-1 and an ordinary agent in a deterministic world; (No-Difference Premise)

**C1** Ordinary agents in deterministic worlds are not morally responsible for their actions. (From P1 and P2)

P1 is intuitive, and P2 is motivated by Cases 2-4. We could think of P2 as simplifying steps that transfer non-responsibility judgment from each case to the next and from Case 4 to any ordinary agent in a deterministic world. This form, however, is at odds with Pereboom's claim that the argument features an inference to the best explanation. After presenting the four cases and arguing that Plum is responsible in none of them, Pereboom ends the argument by concluding *why* Plum is not responsible in the cases. This step of the argument is an inference to the best explanation. There are two questions we should try to answer about this inference. First, what does it add to the argument? Second, is it necessary for the argument? I will take these questions in turn. First, regarding what it adds to the argument, we find two views in the literature. Benjamin Matheson (2016) argues that the aim of the inference to the best explanation in Pereboom's argument is "to 'push through' the generalization of the non-responsibility intuition from Cases 1–3 to Case 4"

[5] See, for instance, McKenna (2008a, 143).

(1970). If we were to put Matheson's view of the argument in a schema, we would get something like this:

**P1** Plum-1 is not responsible for his actions in Case 1; (Manipulation Premise)

**P2** There is no difference relevant to attributions of moral responsibility between Plum-1 and Plum-2 and Plum-3; (No-Difference Premise)

**C1** Plum is not morally responsible for his actions in Cases 1-3. (From P1 and P2)

**C2** Causal determination by factors outside one's control precludes moral responsibility; (From C1 by inference to the best explanation)

**C3** Plum is not morally responsible in Case 4. (From C2)

According to Matheson, Pereboom adds the inference to the best explanation to deal with a possible objection to the argument, an objection he raises just before making the inference to the best explanation:

> A distinguishing feature of Case 4 is that the causal determination of Plum's decision is not brought about by other agents (Lycan 1997). But the claim that this is a relevant difference is implausible. Imagine further cases that are exactly the same as Case 1 or Case 2, except that states at issue are instead produced by a spontaneously generated machine—a machine with no intelligent designer (Pereboom 2001: 115) or a force field (Mele 2005). Here also Plum would lack moral responsibility. (Pereboom, 2014, 79)

Pereboom is considering the compatibilist reply that the non-responsibility judgment does not carry to Case 4 because it has a relevant difference to the other cases. In the fourth case, but not the others, Plum is not determined by other agents. By pinpointing this factor as the explanation for our intuition of non-responsibility in the first three cases, the compatibilist can conclude that we should not judge Plum-4 as non-responsible. Matheson sees the inference to the best explanation as a way Pereboom found to avoid this objection. The inference gives us an explanation for the non-responsibility in Cases 1-3 which is not the determination by other agents. With this alternative explanation, the non-responsibility judgment can carry to Case 4.

Matheson argues, however, that this inference to the best explanation is not necessary. He says Pereboom can avoid the objection by simply putting forward versions of some (or all) of the first three cases where the manipulation is done by an intention-less machine instead of agents:

> But there's no reason to deploy the inference to the best explanation here. If Pereboom's extra case (or cases) that feature no agent-manipulators elicit the nonresponsibility intuition, then Pereboom has undercut the compatibilist resistance to the generalisation of this intuition to Case 4 [...]. And if Plum4 is not morally responsible, then compatibilism is false. (1971)

I think Matheson's interpretation of Pereboom is wrong. There are two textual reasons to think that Matheson's interpretation is incorrect. First, if it were correct, it would make more sense for Pereboom to make the inference to the best explanation after presenting Case 3, and not Case

4. Pereboom, however, does not appeal to the inference to the best explanation in the transition from the third to the fourth case. Second, Pereboom concludes his reply to the compatibilist's objection in question *before* making this inference. Hence, *pace* Matheson, the inference to the best explanation has no role in replying to the objection. Here is the passage in question (that follows directly from the last passage I quoted from Pereboom):

> **From this we can conclude that causal determination by other agents was not essential to what was driving the intuition of non-responsibility in the earlier cases.** I claim instead that Plum isn't responsible in these four cases because he is causally determined by factors beyond his control in each. **Here's the reasoning:** It's highly intuitive that Plum is not morally responsible in Case 1, and there are no differences between Cases 1 and 2, 2 and 3, and 3 and 4 that can explain in a principled way why he would not be responsible in the former of each pair but would be in the latter. We are thus driven to the conclusion that he is not responsible in Case 4. The salient factor that can plausibly explain why Plum is not responsible in all of the cases is that in each he is causally determined by factors beyond his control to decide as he does. This is therefore a sufficient, and I think also the best, explanation for his non-responsibility in all of the cases. (Pereboom, 2014, 79) (Bold not in the original)

In the first highlighted part, we see Pereboom concluding his reply to the compatibilist objection. In the second highlighted part, he starts his inference to the best explanation. Therefore, it is not part of the reply.

After discussing Matheson's interpretation of the role of the inference to the best explanation in the argument, let us examine the second interpretation. Kristin Mickelson (2016) sees the inference as supplementing the argument's conclusion.

Before explaining her view, however, we must digress about the taxonomy of the positions in the metaphysics of moral responsibility.

### 1.3.1 A digression about taxonomy

The standard way of classifying the positions is the following:

**Compatibilism** There is a possible world where determinism is true and an agent is morally responsible for her action;

**Incompatibilism** There is no possible world where determinism is true and an agent is morally responsible for her action;

A problem with this classification is that it does not do justice to *impossibilism*, another position out there. Impossibilists, like Galen Strawson (1994), believe that moral responsibility has a necessary condition that is metaphysically impossible. The impossibility of this condition, however, may be completely independent of the truth of determinism. Given that, impossibilism does not seem to be a brand of incompatibilism despite being classified as incompatibilism by the standard taxonomy. Kadri Vihvelin (2008; 2013; 2022) proposes to amend the standard classification to accommodate impossibilism. Her proposal is the following taxonomy:

**Compatibilism** There is a possible world where determinism is true and an agent is morally responsible for her action;

**Incompatibilism** There is a possible where an agent is morally responsible for her action, and there is no possible world where determinism is true and an agent is morally responsible for her action;

**Impossibilism** There is no possible world where an agent is morally responsible for her actions.

Vihvelin uses this three-fold classification to defend a specific view of the dialectics of the debate. The traditional view of the dialectics is that the compatibilist has the burden of proof. In Vihvelin's taxonomy, however, the compatibilist and incompatibilist accept a thesis in common, the possibility of free will; the incompatibilist also accepts the additional thesis that no deterministic world is a world with free will. The burden of proof, hence, is with the incompatibilist, given that she accepts an additional thesis.[6] A consequence of this is that a defense of compatibilism does not need a positive argument, but just a rebuttal of incompatibilist arguments.

Vihvelin's taxonomy and understanding of the dialectics have come under criticism. My focus is on taxonomy, so I will not discuss dialectics. McKenna (2010) argued that incompatibilism and impossibilism are not mutually exclusive positions. According to him, someone can believe that determinism precludes responsibility and also believe that *in*determinism precludes responsibility. That person seems to be both an incompatibilist, given her acceptance that determinism and responsibility are incompatible, and an impossibilist, given her acceptance that responsibility is metaphysically impossible.

Kristin Mickelson (2015a) agrees with McKenna and adds two criticisms against Vihvelin's taxonomy. First, she argues that Vihvelin is mixing two different questions we can ask about the relation between determinism and responsibility. We can ask if determinism and responsibility can co-exist. That question is distinct from whether determinism *undermines* responsibility, i.e., if responsibility can be precluded because of determinism. As Vihvelin formulates the positions, they are replies to the former question, not the latter. As McKenna's example makes clear, however, we tend to understand incompatibilism as a reply to the second question. We see it as the position that, if determinism is true, there is no moral responsibility *because* of determinism. Mickelson's second criticism is that Vihvelin herself has this vision of incompatibilism. When discussing incompatibilism, Vihvelin also formulates it with an explanatory component.

Besides that, Mickelson points out that Vihvelin's definition of compatibilism is unsatisfactory for the same reason. She brings the example of Neil Levy (2011). He identifies as a compatibilist, given that he believes that determinism does not undermine responsibility, but he

[6] An interesting aspect of Vihvelin's argument that I will not have time to discuss is the implicit appeal to a thesis in the epistemology of modality, that we should regard something as metaphysically possible until we have evidence of the contrary. We get a reverse of the argument, i.e., the burden of proof is with the compatibilists, if we accept that we should regard something as metaphysically impossible until we have evidence of the contrary.

is, however, an impossibilist. He thinks moral luck makes responsibility impossible. This case shows that compatibilism and impossibilism should not be mutually exclusive positions.

Mickelson does not explicitly define the positions of the debate, but we can use her insights to formulate the following taxonomy:

**Compossibilism** There is a possible world where determinism is true and an agent is morally responsible for her action;

**Incompossibilism** There is no possible world where determinism is true and an agent is morally responsible for her action;

**Impossibilism** There is no possible world where an agent is morally responsible for her actions.

**Compatibilism** No agent's moral responsibility is undermined by determinism;

**Incompatibilism** Agents in deterministic worlds are not morally responsible because of determinism;

The first three positions answer the question of the coexistence of determinism and responsibility. The last two positions are independent of the first three and answer the question of whether determinism undermines moral responsibility.

### 1.3.2 Back to the Four-Case Argument

After this digression, we can go back to the Four-Case Argument. Mickelson sees the inference to the best explanation as supplementing its conclusion. We go from concluding incompossibilism to concluding incompatibilism. In other words, the inference adds to the conclusion an explanation for the non-responsibility of Plum and agents in deterministic worlds. (Exegesis is not our focus here, but I think this understanding of the argument does more justice to Pereboom's exposition.)

With this understanding of the argument, we can formulate the following schema:

**P1** Plum-1 is not morally responsible for his actions in Case 1; (Manipulation Premise)

**P2** There is no difference relevant to attributions of moral responsibility between Plum-1 and Plum in the other cases; (No-Difference Premise)

**C1** In each of the cases, Plum is not morally responsible; (From P1 and P2)

**C2** Causal determination by factors outside one's control precludes moral responsibility. (From C1 by inference to the best explanation)

We can now ponder the second question about this inference: is it necessary for the argument? Without it, we have an argument for incompossibilism; with it, an argument for incompatibilism. Which conclusion should the proponent of the argument aim for?

The debate about what should the aim of a manipulation argument be is more developed around Alfred Mele's Zygote Argument (1995, 190; 2006, 188-195; 2019, 83-86). Mele explicitly defends that his argument does not need to conclude an explanation for the manipulated agent's non-responsibility. This feature of his argument has come under criticism (Mickelson, 2015b, 2021).

Against this kind of criticism of manipulation arguments, some have replied that it would be enough for the argument to conclude incompossibilism (or, as they put in the traditional taxonomy, that compatibilism is false) (De Marco (2016); Matheson (2016, 1971)).

What the argument should conclude depends on what we take the debate to be about. To conclude incompossibilism would undoubtedly be a significant contribution to the discussion. Even so, enough people, including Pereboom, seem to be also interested in the additional question: Does determinism undermine responsibility? Given that, the inference to the best explanation is not an *un*necessary step in the argument. Concluding incompatibilism would also be a major contribution to the debate. Therefore, I will assume that the Four-Case Argument has an inference to the best explanation that gets us from incompossibilism to incompatibilism.

In this section, I discussed the structure of the Four-Case Argument. According to Pereboom, the argument is an inference to the best explanation. I considered two ways to understand this inference and concluded that it is better to understand it as supplementing the conclusion of the argument; the argument goes from concluding incompossibilism to concluding incompatibilism.

## 1.4 How should the compatibilist reply?

In this section, I will briefly discuss some compatibilist replies to the Four-Case Argument. I will not criticize these replies, but I will point out two alternative strategies that compatibilists can employ against manipulation arguments. I will then briefly explain how my arguments in this dissertation fall under these strategies.

Michael McKenna (2008a) popularized a categorization of responses to manipulation arguments. Replies that deny P1, i.e., the Manipulation Premise, are *hard-line*; Replies that deny P2, i.e., the No-Difference Premise, are *soft-line*. Hard-line replies will try to put doubt on the reliability of our intuition about the manipulation case or will try to generate an opposite intuition. Soft-line replies will try to point out a difference relevant to attributions of moral responsibility between the manipulated agent and ordinary agents in deterministic worlds. In the case of the Four-Case Argument, soft-line replies also need to point out the pair of adjacent cases that go from a case of non-responsibility to a case of responsibility.

McKenna recommends hard-line replies for compatibilists based on the following reasoning. Suppose a soft-line reply that identifies condition $C$ as a necessary condition for responsibility that is not satisfied in the proposed manipulation cases but can be satisfied in a deterministic world. The incompatibilist has a simple way to deal with this. She can reformulate

the manipulation case so that condition *C* is satisfied in this new case. Behind this reasoning, we see McKenna making a concession to incompatibilists. He accepts that compatibilists are committed to attributing responsibility to a manipulated agent if the manipulation's contribution is the same as the determinism's contribution in a deterministic world.[7] To avoid this cycle of new conditions and new cases, McKenna affirms that the best compatibilist strategy is to accept a hard-line reply.

McKenna's own hard-line reply is best summarized as "one's modus ponens is another's modus tollens". Accepting the No-Difference Premise, we could start in any case and transfer our judgment from there. So, McKenna asks: What if we begin with the ordinary deterministic case, stressing all satisfied compatibilist conditions, and go from there to the manipulation case? According to him, an agnostic about (in)compatibilism starting with the deterministic case would remain agnostic all the way to Case 1. If that is the case, that is enough to put a credible doubt on our initial intuition about Case 1. So, if McKenna's prediction is correct, incompatibilists cannot support the Manipulation Premise only with our initial intuition; an additional argument is needed. In the absence of such an argument, the Four-Case Argument is not a problem for compatibilism. (Fischer (2011) accepts the same reasoning against the Zygote Argument.)

I will not debate the compatibilists' replies at length, but I will briefly present Pereboom's counter-argument to McKenna's hard-line reply because I believe it shows why it does not work (Pereboom, 2008). Pereboom says that we should not suppose a *confirmed* agnostic when assessing the merits of the Four-Case Argument. We should suppose someone who does not hold any specific belief about the problem of determinism but is open to re-evaluating her position based on new evidence. According to Pereboom, even if this person goes from Case 4 to Case 1, she will not remain an agnostic. When presented with cases like 2 and 1, she will get the intuition that Plum is not responsible in these cases and then re-evaluate her position about Case 4. In later work, McKenna concedes that the Four-Case Argument presents a reason in favor of incompatibilism but argues that it is not a decisive reason (McKenna, 2014).

I think Pereboom's response is enough to show the problem with McKenna's hard-line reply, but there are other ways to be a hard-liner. McKenna tried to put doubt on our initial intuition about Case 1, but other hard-line replies try to motivate the opposite intuition, that Plum-1 is indeed responsible. The hard-line reply ends with an argument for favoring the latter intuition over the former.

Such is the case with Sofia Jeppsson's hard-line reply (2019). Jeppsson concedes that she has the intuition that Plum-1 and Plum-2 are not responsible and that going from Case

[7] It is interesting to note that, in this point, McKenna agrees with Robert Kane, a prominent libertarian:

> What I am suggesting then [...] is that one important reason why libertarians require the indeterminist condition is to ensure against the success of any potential covert non-constraining (CNC) control of the agent's will by another agent. It should be noted that no existing compatibilist account of freedom [...] defines freedom so that this result is ensured. (Kane, 1985, 37)

4 to 1 does not reduce the appeal of this intuition. She says, however, that we find in fiction cases where an agent is manipulated in a way similar to them and, yet, we have the intuition that the agent *is* responsible for her actions. The example she gives is the Hologram Doctor from Star Trek Voyage. The way Doctor acts is a result of the interaction of the environment with the program a scientist installed in him. In this way, he is similar to Plum-2, except that this scientist did not have the foreknowledge and intentionally that the scientists in Case 2 had. In one episode, Doctor saves a man's life and is uncertain if he deserves praise for it, given his origin. Jeppsson says that, in this case, she has the intuition that Doctor is responsible even if he has a backstory similar to Plum-2.

The difference in intuitions, according to Jeppsson, occurs because the Doctor's story is presented in such a way that we are invited to see things from his point of view. We set aside the fact that he was programmed and we accompany his process of decision-making. When we place ourselves in the shoes of these manipulated agents and see their deliberation from their point of view, we get the feeling they are morally responsible. Jeppsson calls this way of seeing the agent's action "*agential perspective*", in opposition to the "causal perspective" we have when paying attention to the causes of the agent's action in a way not focused on how the deliberation was like for the agent.

Each perspective gives us a different intuition about manipulated agents. Which perspective should we trust when trying to ascribe moral responsibility? According to Jeppsson, we ought to take the agential perspective in our normal interactions with other people. That follows from the more general ethical idea that we should see other people as fully-fledged agents and not just "as sophisticated vending machines" (1945). For a criticism of Jeppesson's hard-line reply, see Seybold (Forthcoming).

We have seen two hard-line replies. Despite their merits, I think it is too early for a hard-line reply. Our non-responsibility intuition about Case 1 (and maybe Case 2) seems to be the kind of datum about how moral responsibility works that we should leave unaccounted for in our theory only as a last measure. Two routes are preferable to a hard-line reply. First, the compatibilist can dwell on the details of Case 1. She can force the incompatibilist to reformulate the case by finding elements of it that should not be there. If the reformulated case produces a weaker non-responsibility intuition, that makes a hard-line reply easier. We have at least one successful instance of this strategy. Kristin Mickelson (2010) has argued that, in a previous formulation of the first case (Pereboom, 2001), Plum did not even count as an agent. That has led Pereboom to reformulate the first case into the version I presented. This represents a net gain for compatibilists. She weakened the Four-Case Argument without having to accept an additional condition for moral responsibility. In chapter 3, I will follow this route.

The other route is to find a soft-line reply that can break free of the circle of soft-line replies and new cases described by McKenna. That is what I seek to do in chapter 5.

## 1.5 Causation in the Four-Case Argument

In this chapter, I have presented the problem of determinism and the Four-Case Argument. I discussed its structure and two hard-line replies against it. I then laid out the strategy I will follow in arguing against the Four-Case Argument. Now, the other main theme in this dissertation besides the Four-Case Argument is causation. I will conclude this chapter by briefly explaining the role of causation in the Four-Case Argument.

A standard view about moral responsibility is that it has two necessary conditions that are jointly sufficient. One condition is epistemic. We need to know (or believe) certain things about our actions (or the consequences of our actions) to be responsible for them. Another condition is metaphysical. We need to have a certain kind of control over our actions to be responsible for it. If the agent meets both conditions, she is responsible for her action.

Determinism is a threat to responsibility by being a threat to our control, the metaphysical condition for responsibility. We can see this in the Four-Case Argument. The argument tries to show that we are not responsible if determinism is true because we would have the same level of control over our actions as a manipulated agent.

The opponent of the Four-Case argument is a sourcehood compatibilist. She believes that the actual causal sequence leading to the action is the only ground of the metaphysical condition for responsibility.[8] Hence, if she wants to argue against the Four-Case argument without accepting a hard-line reply, she will need to say something about the causal sequence leading to Plum's action in Case 1.

If the compatibilist has to say something about the causal sequence, the metaphysics of causation may help. A better understanding of causation can elucidate the causal difference between manipulation and determinism. That is the idea that I put to work in the rest of this dissertation. This idea has been applied by other philosophers (Deery and Nahmias, 2017; Buckareff, Forthcoming), but not with the same views about causation that I will discuss.

[8] She can also accept that whatever grounds the actual causal sequence also grounds the metaphysical condition for responsibility. This consideration does not impact the discussion presented in this section.

# Part I: Extensionality

# 2 Causal extensionality

In the previous chapter, I presented the Four-Case argument and briefly discussed two strategies I think fare well against it. One of them was to find a flaw in the formulation of the manipulation cases that leads up to the incompatibilist needing to reformulate the cases into weaker ones. In the next chapter, I will pursue this strategy. A specific theory of causation will come to hand when arguing against the Four-Case Argument in the next chapter: contrastive causation. In this chapter, I will discuss the problem of causal extensionality to motivate contrastive causation.

The chapter is divided into three sections. In the first section, I will present the problem of extensionality. In the second, I will discuss some possible solutions to the problem with a focus on contrastivism. The last section is a brief conclusion where I will explain how I will link this discussion with the Four-Case Argument.

## 2.1 The Problem of extensionality

In this section, I will present the problem of causal extensionality. I start by presenting the notion of extensionality more broadly. Then, I narrow it down to the causal case. After discussing the reasons to think that causation is extensional, I will present some cases that put doubt on this thesis.

In a more broad sense, we can talk about an extensional context. In an extensional context, we can substitute terms for co-referent terms without changing the truth-value of the sentences in question. A non-extensional context is one where substituting terms for co-referent terms *may* change the truth-value of the sentences. We can extend the notion of extensionality to terms. An extensional term generates an extensional context; a non-extensional term generates a non-extensional context. We can then apply this concept to specific classes of terms, like predicates and operators.For instance, "is in Asia" is an extensional predicate. If "Mt. Everest is in Asia" is true, "the highest mountain on Earth is in Asia" is also true. On the other hand, "Mary believes that" is a non-extensional operator: "Mary believes she is seeing the highest mountain on Earth" may be true, and, at the same time, "Mary believes she is seeing Mt. Everest" may be false.

With this notion in hand, we can ask: Is causation extensional? Or, more precisely, are causal terms extensional terms? In other words, the question is whether the following is true:

**Extensionality of Causation** a causal sentence "*a* causes *b*" does not change in truth-value with substitutions of "*a*"or "*b*" by co-referent terms.

We have two reasons to think **Extensionality of Causation** is true. First, it seems intuitive; if "Mary seeing Mt. Everest caused her to gasp" is true, it seems that "Mary seeing

the highest mountain on Earth caused her to gasp" is also true. Second, one can argue that the extensionality of causation follows from the idea that causation is a natural relation, one that is "in the world" and, therefore, independent of how we talk about it.[1] Strawson adopts this reason for extensionality:

> Suppose a particular happening or a particular condition of things, *A*, is the cause or part-cause of another particular happening or condition of things, *B*. Then if causality is a natural relation, a relation which holds in nature between *A* and *B*, that relation holds however *A* and *B* may be described. (Strawson, 1985, 118)

It is not clear, however, how causation being a natural relation leads to it being extensional. Metaphysical necessity and grounding are, arguably, natural relations and, yet, both are non-extensional. It would be interesting to better develop the link between being a natural relation and being an extensional relation, but that is not the focus of this chapter. I will focus on the putative counterexamples to **Extensionality of Causation**.

I will now present three putative counterexamples to **Extensionality of Causation**. First, a case adapted from Anscombe (1969)[2]:

1. Joan being Rhesus-negative caused the child to die;

2. The tallest girl in town being Rhesus-negative caused the child to die.

1 seems true and 2 seems false. Nevertheless, if causation is extensional, both have the same truth-value. **Extensionality of Causation** clashes with our intuitions about this case.

The second will be an adaptation of a case discussed in Goldman (1970, 3-4) and Lewis (1986a, 255-256):

**John's Greeting** John is very shy but is working on it. He sees Mary walking in the hallway and decides to greet her. He uses all his willpower to say "hello" but ends up saying it much louder than necessary. Mary jumps when she hears the loud "hello". Had John said "hello" at a normal volume, Mary would not have jumped.

With this scenario in mind, let's consider these two sentences:

3. John saying "hello" loudly caused Mary's jumping;

4. John saying "hello" caused Mary's jumping.

---

[1] It may seem, then, that philosophers that are not realists about causal powers do not have a reason, besides maybe intuition, to accept that causation is extensional. We may ask, for instance, what reasons do philosophers that accept the Humean mosaic have in favor of **Extensionality of Causation** (Lewis, 1986b, Introduction). I think the major reason they have is that they usually accept that events have a well-established *causal history* (Lewis, 1986a; Beebee, 2004), and this idea also seems to be at odds with the possible description-dependence of causal relations.

[2] Anscombes formulate the case with "because" instead of "caused" (155). She takes these sentences to express causal relations, but I prefer the term "caused" given the importance of the distinction between causation and causal explanation in this debate.

Apparently, 3 is true, and 4 is false. However, if causation is extensional and *John saying "hello" loudly* and *John saying "hello"* are the same event, we can infer 4 from 3. This gives us a reason to reject **Extensionality of Causation**.

The third counterexample is generated by the use of emphasis and was first presented by Peter Achinstein (1975). The pair of sentences is as follows:

5. Socrates drinking **hemlock** at dawn caused Socrates' death;

6. Socrates drinking hemlock **at dawn** caused Socrates' death;

The intuition here is that 5 is true and 6 is not. However, it seems obvious that *Socrates drinking **hemlock** at dawn* and *Socrates drinking hemlock **at dawn*** refer to the same event. We have, therefore, another argument against **Extensionality of Causation**. In the next section, we will see some ways to deal with these counterexamples.

## 2.2 The positions

In this section, I will present the positions one can have about the extensionality of causation. Philosophers disagree a lot about what is the lesson to be learned from the putative counterexamples to extensionality.

The most conservative position takes our intuitions about the counterexamples to be errors that occur because we do not distinguish clearly between causation and causal explanation (Davidson, 1967; Strawson, 1985). We have a true sentence about causation that seems false because the equivalent sentence about causal explanation is false. For instance, 4 is true, but seems false because'Mary jumped because John said hello" is false. In our daily lives, we do not need to have a clear boundary between causation and causal explanation, so we can have misleading intuitions when these phenomena come apart.

On the other extreme, one can just take those intuitions at face value and reject **Extensionality of Causation**. Causation would be, after all, non-extensional. This position has the problem of explaining how exactly different descriptions of the same event affect the truth-value of causal sentences. Defenders of this position include Peter Achinstein (1975), G. E. M. Anscombe (1969); Ryam Wasserman (Manuscript)).

Most philosophers, however, try to vindicate our intuitions about these cases to some extent while preserving causal extensionality. One way to do that is by rejecting that the different descriptions in the counterexamples have the same referent. In other words, we will *refine* our causal *relata*. We have many alternative views about the causal *relata* that avoid counterexamples to some extent. One can, for instance, reject that the causal *relata* are events and accept facts as the *relata*. We preserve the extensionality in **John's Greeting** because *John said "hello"* and *John said "hello" loudly* are two different facts. This move, however, does not handle the emphasis cases. Therefore, proponents of fact causation must account for the difference between

5 and 6 in another way. See Bennet (1988, 32-35) for a discussion of this problem in a fact causation theory.

Similarly, one can reject events as *relata* and instead accept *aspects* as *relata* (Paul, 2000). Aspects are property instances. Different descriptions of events shift the aspects in question. For example, the difference in emphasis between 5 and 6 changes the aspect referred to as the cause. In the former, it is *drinking hemlock* being instantiated by Socrates; in the latter, it is the *being at dawn* being instantiated by the event of Socrates drinking hemlock. This difference explains why the former is true and the latter false.

One can also keep events as *relata* and accept a fine-grained theory of events. This is the position accepted by David Lewis (1986a), Alvin Goldman (1970), and Jaegwon Kim (1973). In a fine-grained theory of events, *John saying "hello"* and *John saying "hello" loudly* are different events. One challenge here, however, is to do that without seemingly "overcounting" events. How many actions did John do in *John's Greeting*? It may seem that just one – saying hello –, but the proponent of this theory needs to say that John did at least two actions: saying hello and saying hello loudly. As we see other impacts of descriptions in causal sentences, we will need to grow the list of John's actions. That may seem like a point against the theory. The situation gets worse if we try to account for the emphasis case. Lewis's theory of events can accommodate *Socrates drinking **hemlock** at dawn* and *Socrates drinking hemlock **at dawn*** being different events. We have a property that is essentially a drinking-hemlock and one that is essentially a drinking-at-dawn. But now it may seem that Socrates drank hemlock twice. Against this kind of objection, one way to go is by making a distinction between *different* events and *distinct* events (Lewis, 1986a, 256). These events are different, in the sense of being non-identical, but are not distinct. They are not distinct in the same way a part and a whole are not distinct. Lewis says that, as a general rule, if one event implies another, they are not distinct.

Last, we can preserve the extensionality of causation but reject that causation is a binary relation. The putative counterexamples to the extensionality are cases with a difference in omitted *relata*. That is the solution that Jonathan Schaffer (2005; 2012) proposes to the problem of extensionality. In his contrastive theory of causation, causation is a quaternary relation with form "*a* rather than *a** causes *b* rather than *b**".[3] With this view, for example, 5 and 6 are correctly formulated as:

7. Socrates drinking **hemlock** at dawn, rather than drinking something else, caused Socrates' death rather than him staying alive;

8. Socrates drinking hemlock **at dawn**, rather than at other time, caused Socrates' death, rather than him staying alive;

Now we can see why one of these sentences is true and the other false. The change in emphasis was, after all, changing the omitted *relata*. Let us call this position *Contrastivism*.

[3] We can also accept contrastivism only in the cause or effect. In this case, causation would be a ternary relation. Given that the problems presented occur both in the cause and the effect, contrastivism with a quaternary relation seems better motivated.

In this section, I presented some positions in the debate about causal extensionality. One can propose an error theory for our intuitions against **Extensionality of Causation** by appealing to a distinction between causation and causal explanation. One can just reject **Extensionality of Causation** however. One can also change one's view about the causal *relata*. Last, one can accept contrastivism, the view that causation is a quaternary relation.

## 2.3 Conclusion

In this chapter, I discussed the problem of causal extensionality. I presented the problem and some positions about it. One of the positions was contrastivism. In the next chapter, contrastivism will help us get a clearer understanding of C-Luck, a kind of moral luck. With the notion of C-Luck in hand, I will argue against the Four-Case Argument.

# 3 Contrastive causation and the Four-Case Argument

In this chapter, I will argue that the first case of the Four-Case Argument does not support its incompatibilist conclusion. I will show that, as Pereboom formulates the case, Plum is subject to a specific kind of moral luck, C-Luck (Levy, 2011; de Calleja, 2014; Mele, 2015). This renders the usage of this case problematic in the argument. The incompatibilist will need either to modify the case or to abandon it. I will consider some possible modifications of the case and defend that they generate a worse argument than the original Four-Case Argument. Hence, the incompatibilist should abandon the first case.

Without its first case, the Four-Case argument puts less pressure against compatibilism. The argument's persuasiveness is directly related to how strong our non-responsibility intuition about the manipulation scenario is. If the Four-Case loses the first case, its strongest manipulation scenario, it becomes a weaker argument. This weaker version of the Four-Case Argument is also more vulnerable to hard-line replies. Without the first case, a hard-line reply needs to focus only on the second case. If I am right that the incompatibilist needs to abandon the first case, a hard-line reply becomes an easier route for compatibilists to take.

I will employ in my argumentation the contrastive theory of causation (Schaffer, 2005) presented in the previous chapter. It is not a necessary commitment of my argumentation, but assuming it has an advantage. Given a contrastive theory of causation, instead of understanding C-Luck as being grounded on some modal facts, we can see it as grounded on the actual causal sequence that leads to the action. As I will argue later, this difference is important. It makes explicit the threat that C-Luck poses to our responsibility.

This chapter is divided into five sections. In the first section, I will discuss C-Luck and define it in terms of contrastive causation. In the second, I will show that there is C-Luck in the first case of the argument. I will argue that this is a problem for the argument. The first case needs to be modified or abandoned. In the third section, I will consider some possible modifications for the first case and conclude that they all fail to create a plausible version of the Four-Case Argument. The overall conclusion is that the first case needs to be abandoned. In the fourth section, I will discuss how well the Four-Case Argument fares without its first case. I will defend that it puts less pressure on the compatibilist position than before. I will conclude in the last section by discussing how my argument fits in the broader discussion about the Four-Case Argument.

## 3.1 C-Luck

This section is about C-Luck, also named "cross-world luck". I will start by presenting Alfred Mele's present luck objection to libertarian accounts of responsibility.[1] Compatibilists face a similar problem, the problem of C-Luck. I will present it and defend a definition of C-Luck in terms of contrastive causation.

A classical objection to libertarian accounts of responsibility is that indeterministic actions are subject to luck. A libertarian requirement for responsibility is that there are no antecedent causes that determine the agent's decision.[2] However, if the agent's decision is indeterminate, it seems that deciding to *A* instead of deciding otherwise is not up to the agent. It is just a matter of luck. If this kind of luck is responsibility-undermining, no agent is responsible for an action satisfying the proposed libertarian conditions for responsibility.

Alfred Mele gives a more fleshed-out luck objection against libertarianism.[3] Mele argues that standard libertarians[4] have a problem with *present luck*, i.e., luck at the time of the decision (Mele, 2006, 66). Suppose that an agent in the actual world decides to perform action *A*. According to standard libertarians, for the agent to be responsible for her decision, there needs to be a world *w* with the same past and laws of nature of the actual world, but where she decides not to perform the action. Now, compare the actual world with *w*. Given that the actual world and *w* diverge only at the time of the agent's decision, there is nothing to account for her deciding instead of not. We have a difference between these worlds, the agent deciding to *A* instead of deciding otherwise, which is left unexplained. This difference seems to be just a matter of luck (58-60). If luck entails a lack of control and control is necessary for responsibility (49-50), present luck undermines responsibility. If that is the case, standard libertarianism is false because it proposes a necessary condition for responsibility that is incompatible with it.

Present luck is not a worry for compatibilists.[5] If determinism is true, there is no possible world with the same past and laws of nature where an agent decides differently.[6] Hence, no one can be subject to present luck in a deterministic world. Nevertheless, some have argued that compatibilists face a problem similar to the problem of present luck (de Calleja, 2014;

[1] Much of this discussion is centered on free will. However, given that free will is usually equated to *responsibility-level freedom* (Mele, 2015, 1-2), I will frame the discussion around moral responsibility.

[2] This is a requirement for *direct* responsibility. Libertarians can accept that we are (indirectly) responsible for causally determined actions insofar as they bear a specific relation to actions for which we are directly responsible. For a more detailed discussion about the distinction between direct and indirect responsibility, see Mele (2021).

[3] Mele discusses this problem for libertarians but does not think it shows that libertarianism is false. Instead, he proposes a solution to the problem himself (Mele, 2006, Chapter 5).

[4] By "standard libertarian", I mean someone that accepts as a necessary condition for responsibility that there is indeterminacy at the time of decision (or some instants before). Other versions of libertarianism also have a luck problem, but not with present luck.

[5] A compatibilist may also defend that an agent can be responsible for an indeterministic action. In this case, she has a problem with present luck. Nevertheless, it does not come from her commitment to compatibilism.

[6] This follows from more traditional definitions of determinism straightforwardly. Nevertheless, it is also a consequence of the definition of causally determined action that I provided in chapter 1. By keeping the past and the laws of nature fixed, we keep fixed the causally relevant factors that, together with the laws of nature, ensure the occurrence of the agent's action. Hence, there is no possible world with the same past and laws of nature where the agent decides differently.

Levy, 2011). A world where the agent decides differently cannot have the same laws and past as the actual (deterministic) world, but it may differ only in aspects that are not *significant* for decision-making and attributions of responsibility. If we have such a world, the cross-world difference that accounts for the agent deciding to $A$ instead of deciding not to $A$ seems to be just a matter of luck. This kind of luck is called "*C-Luck*". We can define C-Luck as a form of crossworld-difference:

**C-Luck** Agent $S$ doing action $A$ at world $w$ is subjected to C-Luck if, and only if, there is a possible world $w'$ such that (i) $S$ does not do $A$ at $w'$, (ii) $w$ and $w'$ do not differ in any significant factor.

The notion of C-Luck is directly linked to the idea of a significant factor. We need, then, to say something about significant factors to be able to identify which cross-world differences count as C-Luck, and which do not. De Calleja lists as significant factors: "his reasons, his character traits and even his way of deliberating" and "attitudes, tendencies, skills and features of the circumstance which are relevant to what he is motivated and disposed to do in the circumstance" (2014, 115). Levy says that what considerations come to mind, our mood, and the wanderers of our attention are all chancy factors that can result in an action being subject to C-Luck (2011, 90-91). A more general account of what can vary when searching for C-Luck would be good, but their generalizations will be enough for my objectives here.

Let me give an example of C-Luck. Following Mele (2015), I will take an example from the situationist literature. Robert A. Baron (1997) found that pleasure odors, such as roasting coffee, increased the chance of people helping a stranger. Let us use this result to elaborate a plausible scenario:

**Coffee Aroma** John is walking downtown when a homeless person stops him and asks for money. Luckily for her, a nearby coffee shop is exhaling a delicious coffee aroma. John feels generous and decides to give her some money. He would have decided otherwise if it were not for the smell of coffee.

John's action here is subject to C-Luck. To see this, consider the possible world where no coffee is being roasted. In this world, John does not help the person in need. But the difference between this world and the actual world is not something significant about the situation. The presence or absence of the aroma of coffee is not something that enters (or should enter) John's deliberation about helping someone. It is also not part of John's agency or what he values. Hence, this cross-world difference counts as luck. John's action is subject to C-Luck.

Following Mele's presentation of present luck, I have presented C-Luck as a certain cross-world difference. We also found characterizations of both kinds of luck in terms of contrastive explanation. Here is Mele talking about the problem for libertarians: "If there is nothing about the agents' powers, capacities, states of mind, moral character, and the like that explains this difference in outcome, then the difference is just a matter of luck." (1998, 583). And here is de Calleja talking about C-Luck:

> But I cannot see what good it does to Bob* and Bob**, regarding the freedom or rational control they have, that there is a contrastive explanation of their decisions, since these contrastive explanations do not feature causes which distinctively rationalize, or even saliently cause, their deciding as they do rather than otherwise. (de Calleja, 2014, 116)

The idea is that being subject to present luck or C-Luck is a problem because the agent does not ground a contrastive explanation for her deciding to act instead of deciding otherwise. Some philosophers have downplayed the importance of contrast in analyzing these kinds of luck. Randolph Clarke says the following when discussing the luck problem for libertarians:

> Requests for explanation can introduce pragmatic considerations that are extraneous to the causal story, and thus extraneous to the question of a decision's freedom. The debate over free will [...] is not hostage to the debate over contrastive explanation. (Clarke, 2005, 416)

Mele agrees with him (2006, 73). According to him, contrastive explanation is not a significant part of his presentation of present luck (2015, 19). The core of the problem, in his view, is the cross-world difference (2006, 73). I agree with them about contrastive *explanation*, but I think, *pace* Mele, that the cross-world difference is not enough to capture what is worrisome about these kinds of luck. I will argue now that we better capture our worry about these kinds of luck if we define them using contrastive causation.

We worry about luck insofar as it indicates an absence of control.[7] Hence, if present luck and C-Luck are *prima facie* troublesome for responsibility, their presence should *prima facie* indicate an absence of control. However, control is plausibly grounded on causation (see Sartorio 2016, Chapter 1). Therefore, if these kinds of luck are *prima facie* troublesome, they should reveal something about the causal history of the decision. This is Clarke's view about present luck in the quoted passage (see also Clarke 2005, 415), but he does the modus tollens: given that present luck does not tell us something about the causal history, we do not need to worry about it. We should, however, accept the modus ponens: given that there are worrisome phenomena that we are trying to capture under the labels "present luck" and "C-Luck", we need to characterize them in terms of causation.

It may seem, however, that we cannot locate these two kinds of luck in the causal chain leading to the decision. As Clarke points out, there is a cause for the agent's decision in the case of present luck, and hence, present luck cannot be equated to an absence of cause. Likewise, in the case of C-Luck. In these scenarios, we have a deterministic cause of the agent's decision that is beyond her control. That, however, should not be a problem for compatibilists. For instance, in **Coffee Smell**, the coffee aroma is a cause of John's decision. To argue that C-Luck is a threat to responsibility, we need to point out how it differs from other causes of John's action that are not responsibility-undermining according to compatibilists.

This reasoning is onto something, but the conclusion should not be that we are misguided in trying to ground these kinds of luck in the causal chain. The problem here is that

[7] See Levy (2011, 42-43) for a discussion of the relation between luck and control.

we stumbled on the limitations of viewing causation as a binary relation. We can easily define them in terms of causation if we accept a *contrastive* theory of causation.

To illustrate my point, I will use Schaffer's contrastive theory of causation, presented in chapter 2. I will define C-Luck in terms of contrastive causation. The definition is straightforward. This shows that we can easily locate it in the causal chain if we understand causation contrastively. Let us call the total of significant factors a *significant context*. Then, we can define C-Luck as follows:

**C-Luck** An agent $S$ deciding to perform an action $A$ in a significant context $C$ is subject to C-Luck if, and only if, there are event $e$ and set of events $E^*$ such that (i) $e$ rather than $E^*$ caused $S$ to decide to $A$ rather than deciding not to $A$ and (ii) every member of $E^*$ is an event compatible with the significant context $C$.[8]

To illustrate this definition, let us return to **Coffee Smell**. John's decision is subject to C-Luck because there is a pair of events — *the air smelling like coffee* and *the air not smelling like coffee* — such that (i) the air smelling like coffee rather than not smelling like coffee caused John deciding to help rather than deciding not to help and (ii) the *the air not smelling like coffee* is compatible with the significant context of John's decision.

Taylor W. Cyr (2019) and de Calleja (2014) defend that C-Luck does not undermine responsibility. Neil Levy (2011) defends it does undermine responsibility. I will not take a stance on this question because my arguments in this chapter are independent of it.

C-Luck may seem widespread in our daily actions, but it is not a necessary consequence of determinism. We may have an agent $S$ doing an action $A$ such that every world where $S$ does not do $A$ differs in a significant factor from the world where $S$ $A$-ed. For that to be the case, the significant factors must guarantee the performance of the action. That can occur when an action is the obvious choice given the agent's values and reasons to act present in the situation. For instance, we can imagine a doctor on her way to the supermarket who sees someone having an epileptic attack in the street. The doctor takes very seriously her oath to help people and it will not be a huge problem arriving later at the supermarket. So, she decides to help the person. This decision may follow so strongly from her values that she would decide otherwise only if some significant factor were different. Given that it is possible to act without C-Luck in a deterministic world, compatibilists are not committed to a specific solution regarding C-Luck.

In this section, I presented the notion of C-Luck. I also showed how we can use contrastive causation to better characterize it. In the next section, I will show that, as Pereboom presents it, Plum in Case 1 is subject to C-Luck.

[8] For the sake of simplicity, I will assume in the rest of the text that $E^*$ has only one member.

## 3.2 C-Luck in Case 1

So far, I discussed C-Luck and defined it using contrastive causation. With these notions in hand, we can now discuss the Four-Case Argument. In this section, I will first show that Plum-1's decision is subject to C-Luck. Then, I will argue that this poses a problem for the argument.

For Plum-1's decision to be subject to C-Luck, given our definition, the decision has to have a cause such that its contrast is compatible with the significant context. The salient cause of his decision in the case's formulation is the neuroscientists' intervention. Formulating with contrastive causation, the neuroscientists pushing the button rather than not pushing caused Plum-1's decision to kill rather than his decision not to kill. For this to be a case of C-Luck, we need only *the neuroscientists not pushing the button* to be compatible with the significant context of Plum's decision. In other words, we need the intervention to not change some significant aspect of the situation.

Let us go back to the description of Case 1 to see exactly what the neuroscientists' intervention does: "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" (Pereboom, 2014, 76). So the manipulation causes Plum to reason more egotistically, but not *so much more* that his decision is out of character. This intervention results in a decision that would not occur otherwise. If this difference in reasoning is not a difference in significant aspects of the situation, then *the neuroscientists not pushing the button* is compatible with the significant context and, therefore, Plum-1's decision is subject to C-Luck.

For Plum-1's action to be subject to C-Luck, this difference in reasoning needs not be a change of the significant context. We can establish that it is not a change of the significant context even without a better account of what is part of the significant context. That is because Case 1 is similar to paradigmatic cases of C-Luck, like the cases we encounter in the situationist literature. Compare Plum-1's action to John's action in **Coffee Aroma**. John could have reasoned more egotistically if it were not for the pleasant smell; Plum-1 could have reasoned less egotistically if it were not for the neuroscientists. In both cases, we have a contrastive cause for the agent's decision that is not grounded on elements such as character traits or reasons to act present in the situation.

I have shown that Plum-1's decision is subject to C-Luck. I will now argue that this poses a problem for the argument. We have two possibilities to consider. First, C-Luck may undermine responsibility. The compatibilist has an easy reply against the Four-Case Argument in this case: Plum-1's non-responsibility is explained by the presence of C-Luck. Hence, the Four-Case Argument cannot get its incompatibilist conclusion, which requires that Plum-1's non-responsibility be explained by factors outside his control determining his actions.

The second possibility is that C-Luck does not undermine responsibility. In this case, C-Luck still is a *prima faccie* threat to responsibility. There are cases of C-Luck that propel the

intuition that the agent is not responsible because of C-Luck. This intuition may be wrong after all, but it does occur. If the presence of C-Luck can produce non-responsibility intuitions, it undermines Case 1's role in the Four-Case Argument. The success of the Four-Case Argument hinges not only on our intuition that Plum-1 is not responsible, but also on a specific interpretation of that intuition: that it tracks responsibility-undermining features of determinism. If the presence of C-Luck, despite not undermining responsibility, can produce a non-responsibility intuition, we have reasons to doubt our intuition that Plum-1 is not responsible. Even if our intuition about Plum-1 is correct, the incompatibilist cannot show that it is produced by determinism instead of C-Luck. The compatibilist does not need to accept that Plum-1 is responsible. She needs only to argue that, given the presence of C-Luck in Case 1, it cannot be used to elucidate the supposed responsibility-undermining nature of determinism.

I will now briefly consider two objections. First, the incompatibilist may argue that C-Luck is not even *prima faccie* a threat to responsibility. In other words, she may argue that the presence of C-Luck does not affect our intuition about Plum-1. If she is right, it should not be a problem for her to remove it from the scenario. After all, if C-Luck is irrelevant in Case 1 and does not affect our intuitions, the incompatibilist should be able to propose a modified Case 1 that does not feature C-Luck and still produces a non-responsibility judgment. As I will argue in the next section, however, no modification in Case 1 that removes C-Luck produces a case that works well in the Four-Case Argument. Therefore, we cannot save the argument by arguing that C-Luck is irrelevant in Case 1.

The second objection is that Plum's action is subject to many kinds of moral luck, like constitutive and circumstantial luck, and that the argument's proponent does not need to get rid of all these kinds of luck in her formulation of the case. After all, agents in deterministic worlds are subject to them, so it is appropriate for them to be present in manipulation cases. Indeed, the incompatibilist does not need to formulate a manipulation case without moral luck. But not all kinds of moral luck are equal. Some kinds of moral luck, like constitutive and circumstantial luck, are unavoidable for agents like us. Hence, it would be dialectically inappropriate to ask for a manipulation case where these kinds of moral luck do not occur. As discussed in the first section, however, we can act without C-Luck.[9] Therefore, if the incompatibilist wants her cases to be relevantly similar to *all* deterministic actions agents like us can perform, she needs to formulate cases without C-Luck.

In this section, I have argued that there is C-Luck in Case 1, which undermines its role in the argument. The next section is about how the incompatibilist may try to reformulate the first case.

[9] See Cyr (2019) for a discussion.

## 3.3 No way out

If my arguments in the previous section are sound, the Four-Case Argument cannot start with Case 1. In this section, I will consider how one can reformulate it to avoid C-Luck. I will argue that no reformulation renders a plausible version of the Four-Case Argument.

Given our definition of C-Luck, there are two ways to avoid it in Case 1: Modify the contrast for the decision – making the neuroscientists cause Plum-1's decision, but not in contrast to him deciding not to kill – or modify the contrast for the intervention – making it not compatible with the significant context. The first way is a non-starter. We would have the neuroscientists causing an aspect of Plum's decision – like the time – but not causing the decision to happen rather than not happen. I cannot see how someone would think this kind of intervention precludes responsibility for the decision.

Given that, the only plausible way to modify the case is to make the contrast to the intervention incompatible with the significant context. In other words, the neuroscientist's intervention must change something in the significant context. Many aspects of the significant context could be changed, but I think we can categorize the context elements into two groups: those that are part of the situation Plum is in and those that are part of Plum's agency. This distinction may not be clear in some cases, but I think it bears some consequences to how responsibility would work in these scenarios. To further this point, I will present two cases where the intervention affects a significant factor and discuss the different replies compatibilists can give.

The first variation of Case 1 is based on a case presented by Maria Sekatskaya (2019, 1290) when proposing her reply against the Four-Case Argument. We have an intervention that changes the reasons to act that are presented in Plum's context. The case is as follows:

**Case 1.1** Plum and White are co-workers. A group of neuroscientists wants Plum to kill White. After some investigation, they found that the easiest way to get him to do it was to let him know about an upcoming promotion that White would receive that would otherwise go to Plum. They know that if Plum gets this information, a deterministic causal chain will start which will result in him killing White. By pressing a button, the neuroscientists open the door of Plum's office when two other co-workers are discussing White's future promotion. Eavesdropping on the conversation, Plum learns that White will be promoted instead of him. After reflecting on this information, Plum decides to kill White. If the neuroscientists had not pressed the button, the door would not have opened, and Plum would not have eavesdropped and decided to kill White.

This causal chain does not feature C-Luck. After all, knowledge about the promotion is relevant to deliberations; hence, it is part of the significant context. Therefore, Case 1.1 avoids my argument centered on the presence of C-Luck. However, the case has a bigger problem. It does not elicit a non-responsibility judgment as strong as Case 1. This kind of manipulation –

where we present someone with an additional reason in the hope she will act as we want – is not unusual in real life and seems to not (completely) mitigate responsibility. If the Four Case Argument's proponent needs to abandon Case 1 in favor of Case 1.1, that is a huge blow against her position.

Let us then consider another version of Case 1. A version where the significant element of the context changed is something in Plum's agency:

**Case 1.2** Plum and White are co-workers. A group of neuroscientists wants Plum to kill White. They know that Plum knows about an upcoming promotion that White will receive that would otherwise go to him. The neuroscientists decide, then, to increase the value Plum puts on his career. They know that if Plum starts valuing his career more, a deterministic causal chain will start which will result in him killing White. By pressing a button, the neuroscientists activate a chip in Plum's brain that increases the value he puts on work achievements. After reflecting on the upcoming promotion, Plum concludes that it is so important for him that it is worth murdering someone. So, he decides to kill White.

Like Case 1.1, Case 1.2 does not feature C-Luck. How much Plum values his career is part of the significant context. Furthermore, unlike Case 1.1, I think we have a strong non-responsibility intuition about Plum-1.2.

Nevertheless, starting the Four-Case Argument with Case 1.2 will create a problem for incompatibilists. Case 1.2 gets us close to cases of radical reverses. That opens Case 1.2 to some objections that did not apply to Case 1. In cases of radical reverses, we have an agent that suffers a radical change of values.[10] Let us consider, for instance, a case from Mele (2006, 164-165). Mele tells us the story of Ann and Beth. Both are very talented philosophers working in the same department. Philosophy, however, is not a priority for Beth. She values many other things more than philosophy. She does that as a result of careful reflections she had over many years. Their dean decides to make Beth more like Ann by brainwashing her so that she has the same values as Ann. After the brainwashing, Beth reflects on her new values and fully embraces them. Beth now is as dedicated to philosophy as Ann.

Ann seems to be responsible for her dedication to philosophy, but Beth does not seem to be responsible for her later dedication. Both, however, have the same internal structure, i.e., they have the same values and deliberate about their decisions in a similar fashion. If our judgment about this case is correct, responsibility has another condition besides internal conditions, like reason-responsiveness. Motivated by that consideration, many compatibilists have proposed historical conditions for responsibility (McKenna, 2016; Fischer, 2021).

The manipulation that Plum-1.2 suffers seems similar to the manipulation that Beth suffers. The change in value is not as severe but equally intrusive. That leaves open to compatibilists the possibility of explaining our judgment about Case 1.2 using lessons learned with radical reverses. That could be an appeal to historical conditions, but it also could work

[10] See Mele (2019) for a discussion of radical reverses. Especially Chapters 2 and 3.

with other analyses of radical reverses (Matheson, 2014; Björnsson, 2020; Cyr, 2020, 2023). We could reformulate Case 1.2 so Plum satisfies historical conditions concerning his value about his career. The intervention would influence Plum much earlier than the time for his decision. They would start a causal chain that makes Plum change his values through time and with some degree of deliberation. This reformulation puts this case closer to Case 2. It would, then, have no special role in the Four-Case argument besides making the argument open to additional objections. Hence, Case 1.2 is a bad start for the Four-Case Argument because it leaves the argument open to criticisms that it was immune to with Case 1.

I have argued against using Case 1.2 at the start of the Four-Case Argument. Doing so leaves the first case of the argument open to counter-arguments coming from the literature about radical reverses. Given that neither Case 1.1 nor Case 1.2 wield a strong version of the Four-Case Argument, modifying Case 1 will not help the proponent of the argument against my objection related to C-Luck.

## 3.4 The Three-Case Argument

I have argued that Case 1 needs to be abandoned and that no modification can save it. In this section, I will consider another possible incompatibilist reply to my argument. The proponent of the Four-Case Argument can accept that she has a problem with Case 1, but simply argue that this problem does not occur in Case 2, and that our non-responsibility judgment about this case is strong enough to start the argument with it. We get, then, the Three-Case Argument. I concede that this new argument avoids my main argument against the Four-Case Argument, but I will argue that it is a considerably weaker argument.

The Three Case Argument runs as follows. We start with Case 2, a case where Plum was manipulated from birth so he would decide to kill White. The proponent of the argument argues that this case elicits a strong non-responsibility intuition. She, then, presents two further cases (Cases 3 and 4), being the last a case where Plum is just an ordinary agent in a deterministic world. She argues that these three cases do not differ in any aspect relevant to attributions of responsibility. Hence, our non-responsibility intuitions about Case 2 should carry on to the other two cases. The conclusion is that ordinary agents in deterministic worlds are not morally responsible for their actions.

The major difference between Case 1 and Case 2 is that, in Case 2, the manipulation occurs at the beginning of Plum's life. Because of that, Case 2 is compatible with Plum acting without C-Luck. We only need to add to the scenario that the neuroscientists make him develop his values in such a way that his action is akin to the doctor's action in the example I gave in the first section, i.e., without C-Luck. Therefore, Case 2 avoids my objection centered on C-Luck.

There is a downside in abandoning Case 1 in favor of Case 2 however. Some philosophers embrace a distinction between cases of manipulation and cases of *initial design* (Mele, 2008; Fischer, 2016). In the former, we have an intervention in the action of an already-

existent agent; in the latter, we have an agent designed prior to her existence in a way that guarantees she will perform a specific action. Case 1 was a case of manipulation, but Case 2 is a case of initial design. Some compatibilists do not treat these two kinds of scenarios equally. That some compatibilists may change their response to the Four-Case Argument when we exclude Case 1 is not on its own a problem to the Three-Case Argument. The worry for incompatibilists appears when we pay attention to *how* the response may change. For instance, consider what Fischer has to say about these kinds of scenarios:

> On my semicompatibilism, we distinguish between initial design and manipulation. Initial design is compatible with full moral responsibility; the mechanism that issues in the relevant behavior can be the agent's own, appropriately reasons-responsive mechanism. On the other hand, manipulation would generally rule out moral responsibility insofar as the mechanism in question would either not be the agent's own, or not be reasons-responsive (or both). (Fischer, 2021, 266)

We do not need to dwell on the details of Fischer's semi-compatibilism, but we can see that, while he thinks manipulation will tend to undermine responsibility, he believes that an agent designed from birth can be fully responsible for her actions. In other words, while he tends to a soft-line reply to manipulation cases, he accepts a hard-line reply to initial design cases. That seems to reveal that he considers the latter cases to be less troublesome for compatibilists. Even being a major compatibilist, Fischer is just one case, but this brings out something that the Three-Case Argument's proponent should worry about. Our intuition about Case 2 seems weaker than the one about Case 1, especially if we focus on all compatibilist conditions for responsibility that Plum-2 is satisfying. I am not saying that compatibilists should accept a hard-line reply against the Three-Case Argument; I am only saying that it is a softer bullet to bite than a hard-line reply against the Four-Case Argument. Therefore, the Three-Case Argument is a weaker argument for incompatibilism when compared to the Four-Case Argument.

In this section, I considered a strategy the incompatibilist could use to reply to my objection. She could abandon Case 1 and start the argument with Case 2. I defended that despite avoiding my previous objection, this new argument is considerably weaker than the original one.

## 3.5 Conclusion

My goal in this chapter was to argue that the first case of the Four-Case Argument does not support its incompatibilist conclusion. For that, I appealed to a contrastive theory of causation and C-Luck. I showed that Plum in Case 1 is subject to C-Luck, and, because of that, it does not motivate an incompatibilist conclusion. I defended that no modification in Case 1 yields a good start for the Four-Case Argument. Therefore, the proponent of the argument needs to abandon Case 1.

I, then, considered how well the Four-Case Argument would fare without its first case. I argued that, without it, the Four-Case Argument loses its manipulation case that produces

the strongest non-responsibility intuition. Compatibilists, then, have to worry less about the counter-intuitiveness of accepting a hard-line reply. Losing the first case is not a fatal problem for the Four-Case Argument, but it is a step in easing the pressure that manipulation cases put on compatibilism.

# Part II: Causal Selection

# 4 Causal selection

In the previous chapter, I argued against the Four-Case Argument using contrastive causation and the notion of C-Luck. One shortcoming of my argument is that it does not provide a fatal blow against the Four-Case Argument; it only weakens the argument. As discussed in chapter 1, one way to get a definitive counter-argument to the Four-Case Argument would be to propose a soft-line reply that is not vulnerable to new manipulation cases. I will propose such a reply in the next chapter as part of a compatibilist theory I will put forward.

The compatibilist theory I will present in the next chapter is based on Peter Menzies's theory of causation (Menzies, 2007, 2009) . His theory is motivated by the phenomenon of causal selection. Therefore, to motivate this theory of causation, I will present and discuss the phenomenon of causal selection in this chapter. Causal selection is our practice of not electing as causes all the events that satisfy traditional metaphysical conditions for causation (Bebb and Beebee, 2024). Causal selection occurs when norms, be they statistical or prescriptive norms, affect our judgments of causation. In the first section, I will present this phenomenon in more detail.

The standard philosophical position regarding causal selection is to dismiss it as a pragmatic phenomenon or an error in the layperson's application of the concept of causation. In the second section, I will discuss these strategies and argue that they are not the easy route that the orthodoxy makes it look like.

In the third section, I will present Menzies's theory of causation. He intends to accommodate the phenomenon of causal selection in his theory through the idea of *defaults*. In the next chapter, I will use his theory to propose a compatibilist theory. In the fourth section, I will conclude with some final remarks regarding the debate around causal selection.

## 4.1 The phenomenon

This section is about causal selection. Causal selection is a phenomenon found in our ordinary attributions of causation. When we search for the cause of an event, we do not limit ourselves to the metaphysical conditions that philosophers traditionally associated with causation, like counterfactual dependence and physical processes. Statistical normality and moral norms also affect our causal judgments. In daily life, people do not call "cause" many events that philosophers have traditionally considered causes. Causal selection is this practice of not electing as causes all the events that satisfy metaphysical conditions for causation. It is experimentally well-established that our ordinary attributions of causation are subjected to causal selection (Willemsen and Kirfel, 2019).

I will now introduce two cases debated in the literature about causal selection that I will discuss for the rest of the text. The first case shows how statistical norms, i.e., norms about

how frequent the occurrences are, affect causal judgments. The case is the following:

> **Venusians:** A lightning strike has occurred in a remote area on Earth, resulting in a forest fire. The forest fire counterfactually depends on (at least two) things: the lightning strike, and the presence of oxygen in the forest. Some (Earthling) forest rangers are trying to get to the bottom of what caused the fire. As it happens, some Venusian scientists are observing the event through their high-powered telescope, and they are also trying to get to the bottom of what caused the fire. Lightning strikes (we are supposing) are a very common occurrence on Venus; the presence of oxygen, not so much. (Bebb and Beebee, Forthcoming)

This case was first discussed by Hilary Putnam (1982, 150). With this case in mind, let us consider two sentences:

9. The lightning strike caused the fire;

10. The presence of oxygen caused the fire.

Intuitively, when the forest rangers assert these sentences, 9 is true, and 10 is false. On the other hand, in the discussion that the Venusians are having, the opposite is the case: 9 is a false assertion and 10 a true one. This case seems to show how our causal judgments can be affected by our expectations regarding the statistical frequency of an occurrence. For the forest rangers, oxygen is taken for granted because it is always there. So, it is not eligible as a cause. However, for the Venusians, lightning strikes are the usual. So, they take the oxygen as the cause of the fire.

The other case comes from an experiment conducted by Joshua Knobe and Ben Fraser (2008). They presented the following vignette to 18 students in an introductory philosophy class at the University of North Carolina:

> **Pen:** The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own. The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly e-mailed them reminders that only administrative assistants are allowed to take the pens. On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message . . . but she has a problem. There are no pens left on her desk. (Knobe and Fraser, 2008, 443)

After reading this, they were asked if they agreed or not with the following statements:

11. The administrative assistant caused the problem.

12. Professor Smith caused the problem.

The overall result was that people tended to deny 11 and accept 12. This shows another kind of norm playing a role in our causal judgments. It seems that the fact that Professor Smith should not take a pen makes us conclude that he was the cause of the problem, despite the problem also counterfactually depending on the administrative assistant taking a pen.

That our ordinary judgments of causation are prone to causal selection is well-established experimentally. What is still an open question is how to explain the role causal selection plays in our concept of causation. There are four main broad hypotheses regarding the possible selective nature of our ordinary concept of (actual) causation:

**I. Selective Descriptive Concept** Our ordinary concept of causation is selective. Despite norms affecting our judgments of causation, it is a purely descriptive concept. (Hitchcock and Knobe, 2009; Knobe, 2009; Phillips et al., 2015)

**II. Selective Normative Concept** Our ordinary concept of causation is selective. Norms affect our judgments of causation because causation is a partly normative concept. (Sytsma et al., 2012; Sytsma and Livengood, 2016; Livengood et al., 2017; Sytsma et al., 2019; Sytsma, 2020, 2022)

**III. Pragmatic Selection** Our ordinary concept of causation is non-selective. The phenomenon of causal selection is explained by pragmatic mechanisms. (Samland and Waldmann, 2016)

**IV. Defective Mechanism** Our ordinary concept of causation is non-selective. The phenomenon of causal selection is explained by a defective psychological mechanism we have. (Alicke, 2008; Alicke et al., 2011; Sauer and Bates, 2013; Sauer, 2014; Güver and Kneer, 2023)

I will not discuss the plausibility of these hypotheses in detail. What I will do in the next sections is discuss some philosophical positions about causal selection and mention how they relate to these empirical hypotheses. I will use the cases **Venusians** and **Pen** to illustrate these philosophical positions.

## 4.2 The orthodox position and its problems

In the previous section, I presented the phenomenon of causal selection. It occurs in our ordinary concept of causation. So far we are interested in capturing at least part of our ordinary view of causation when doing metaphysics of causation, causal selection is of philosophical interest. In this section, I will discuss *egalitarianism*, the standard philosophical view regarding causal selection. Egalitarianism is the view that causal selection does not correspond to any aspect of how causation really works. For egalitarians, both the lightning strike and the presence of oxygen are causes in **Venusians**. I will argue that egalitarianism depends on some particular views of how causal selection works and, hence, its plausibility is an open question and cannot be taken for granted.

Before discussing egalitarianism, however, I shall consider another possible position: that "cause" is ambiguous; in one sense, it is selective, and in another, it is non-selective. This position could preserve egalitarianism for one sense of "cause" while accommodating the phenomenon of causal selection in another sense. The apparent problem for egalitarianism originates because we conflate the two senses (McDonnell, 2018). This position is not feasible,

but before exposing the problem I see with it, I will discuss a bad reason to reject it. Peter Menzies also rejects that "cause" is ambiguous, saying: "it seems to me to be remarkably implausible that the verb 'cause' should be a homonym like the verb 'bank"' (2007, 198-199). However, the charitable interpretation of the view that "cause" is ambiguous is not that "cause" works like "bank", but that "cause" is *polysemic*, the different senses are related. An example of polysemy would be "book", which can refer to a physical object – "the book on my table" or an abstract object – "Lord of The Rings is Tolkien's most famous book". Therefore, the comparison with "bank" is misplaced. Even so, the position that "cause" is ambiguous in the way proposed still has some hard questions to answer. We need to pay attention to the nature of the proposed thesis. That a term is ambiguous in not a strictly philosophical claim but a linguistic one. There are linguistic tests for ambiguity. Without a confirmation that "cause" passes these tests, we should be skeptical of philosophers who argue that "cause" is ambiguous. Phil Corkum (2022), for instance, argues against Ned Hall's proposal that "cause" is ambiguous (2004) by showing that it does not pass ambiguity tests.[1] In the absence of linguistic evidence for ambiguity in "cause", we should not endorse this view. We are, then, operating under the assumption that there is only one sense of "cause".[2]

Now, let us talk about egalitarianism. The egalitarian rejects the counterexamples proposed. 9, 10, 11, and 12 are all true according to her. This view is well described by David Lewis:

> We sometimes single out one among all the causes of some event and call it "the" cause, as if there were no others. Or we single out a few as the "causes," calling the rest mere "causal factors" or "causal conditions." Or we speak of the "decisive" or "real" or "principal" cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation. (Lewis, 1973, 558-559)

What the egalitarian needs to explain, then, is why we have these intuitions. She has two ways to go (Bebb and Beebee, 2024): to defend **III. Pragmatic Selection** or to defend **IV. Defective Mechanism**. As the quote shows, Lewis accepts the first option; David Rose (2017), on the other hand, is an example of an egalitarian who accepts the second option.

Given that both **III. Pragmatic Selection** and **IV. Defective Mechanism** are empirical hypotheses, it seems that the prospect of these egalitarian views is dependent on the plausibility of these theses. Another route open to egalitarians is to defend that, if our ordinary concept of causation is selective, it is not aligned to the worldly phenomenon that is causation, and it is this phenomenon that we are interested in when doing metaphysics (Beebee, 2004, 293).

[1] Unfortunately, I do not have the space to discuss Corkum's proposal of how to explain the possible multi-interpretability of "cause".

[2] Another possibility is that there is more than one sense of "cause" but they are all selective or all unselective.

This dependence on empirical hypotheses is not a fatal blow against egalitarianism, but it shows that, while we do not get a definitive conclusion about the functioning of causal selection in our ordinary concept of causation, we should be open to embracing the philosophical consequences of accepting causal selection as part of our concept of causation. With this in mind, in the next section, I will present a theory of causation that tries to account for causal selection.

## 4.3 Menzies's theory of causation

If we reject egalitarianism, we need to account for causal selection in our theory of causation. Peter Menzies (2007, 2009) aims to do that by using the idea of *default*. This gives causation a normative aspect and, hence, it is more aligned to **II. Selective Normative Concept**. Nevertheless, there are independent reasons to incorporate defaults in our theories of causation (Fischer, 2021). In this section, I will present Menzies's theory. In the next chapter, I will use his theory to propose a compatibilist theory.

Menzies takes as a starting point for his theory the view that actual causation is difference-making as articulated by Hart and Honoré (1959) (2007; 2009, 201). He identifies three elements in their understanding of causation that he wants to capture in his theory. First, applying causation to a situation requires conceptualizing the situation as a kind of system. Second, we assume that there is a way for this system to behave when left on its own. Third, when the actual state of the system deviates from the expected state given its normal behavior, we search for something that made this difference.

To capture these elements in his theory, Menzies appeals to the structural equations framework. The structural equation framework was made famous by Judea Pearl (2000) and popularized in philosophy by James Woodward's interventionist theory of causation (Woodward, 2003). We satisfy the first element of Hart and Honoré's view right out of the batch with the structural equations framework. In this framework, we assess actual causation in relation to a *causal model*, which we can understand as a conceptualization of the situation. Formally, a causal model is an ordered triple $< U, V, E >$, where $U$ is a set of exogenous variables, $V$ is a set of endogenous variables, and $E$ is a set of structural equations. Exogenous variables are those whose value is determined by factors outside the model; endogenous variables are those whose value is determined by factors inside the model. In $E$, we have a structural equation for each variable. For exogenous variables, the equation just states its value. For instance, supposing an exogenous variable $X$, we may have the structural equation $X = 1$. For endogenous variables, the equation will have this form:

- $Y = f_Y(X_1, ..., X_n)$

Where $X_1, ..., X_n$ are the variables (endogenous and/or exogenous) that determine the value of $Y$.

To exemplify this, let us return to **Venusians**. Let us use the variables $S$ (lightning strike), $O$ (oxygen), and $F$ (fire), and use 1 for presence and 0 for absence. Then, the causal model for the forest fire is $< O, L; F; F = O \times L >$.

Menzies accepts an objectivist view of causal models (203-204). Here, he sees himself as following Pearl. According to this view, there are objective features of the modeled situation that limit acceptable causal models. The variables need to represent the objective dimensions of change of the situation. The structural equations represent the causal mechanisms that govern the system's behavior.

We get difference-making from a causal model by checking for counterfactual dependence. The most natural way to define counterfactual dependence in the SE framework is as follows:

**Definition 1** A variable $Y$ *counterfactually depends* on a variable $X$ in a causal model if and only if it is actually the case that $X = x$ and $Y = y$ and there exist $x' \neq x$ and $y' \neq y$ such that the result of replacing the equation for $X$ with $X = x'$ yields $Y = y'$. (206)

The problem Menzies has with this definition is that it does not capture the idea that a cause makes a difference in the system's normal behavior. For instance, if we apply Definition 1 to our forest fire model, we see that both the presence of oxygen and the lighting strike made a difference (and are, hence, causes). Definition 1 captures an egalitarian view of causation.

Menzies proposes to deal with this problem by modifying the SE framework to permit the values of exogenous variables to be fixed at non-actual values. In the standard SE framework, we calculate counterfactual dependence using the actual values of the variables. Menzies proposes, instead, that we fix variable values at default values. The idea is that default values represent the *normal* state of the system. We call a causal model with the variable values set at their default values a *default causal model*. We then adapt our definition of making a difference:

**Definition 2** A value of a variable $X$ makes a difference to the value of another variable $Y$ in a default causal model if and only if plugging in the default values of the variables in the structural equations yields $X = x$ and $Y = y$ and there exist actual values $x' \neq x$ and $y' \neq y$ such the result of replacing the equation for $X$ with $X = x'$ yields $Y = y'$. (208)

With this new definition, we can capture the idea of causation as difference-making described by Hart and Honoré. The assumption of a default causal model captures that the system behaves a certain way when left on its own. Definition 2 captures that a difference-maker is a deviation from this normal behavior.

This theory entails Causal Contextualism, the thesis that the truth-conditions of causal assertions are sensitive to the conversational context in virtue of their causal terms. The theory ends up contextualist because of the default values setting. In different contexts, even when modeling the situation with the same variables and structural equations, we can arrive at different truth-values for causal assertions because of differences in the default causal models.

We can see this unfolding in the forest fire case. Suppose first that the forest rangers are talking about the fire. The model for the forest fire is $< O, L; F; F = O \times L >$. But what are the default values of these variables? The presence of oxygen in the forest is very normal; the

occurrence of lightning strikes, not so much. Hence, we have $O = 1$ and $L = 0$. From that, we get $F = 0$, the occurrence of fire is not normal. The actual values are $O = L = F = 1$. Is the presence of oxygen a cause of the fire? No, given that its actual value is equal to its default value. On the other hand, the lighting strike is a cause of the fire. Changing $L$'s default value (0) to its actual value (1) changes $F$'s default value (0) to its actual value (1). Therefore, in the context of the forest rangers, 9 is true, and 10 is false.

Nevertheless, we get the opposite result when considering the Venusians's context. For them, lighting strikes are very normal, but the presence of oxygen in an atmosphere is a novelty. Hence, we get normal values $O = 0$ and $L = 1$. From that, we can proceed like in the last paragraph to conclude that, in their context, 9 is false and 10 is true.[3]

What do we take into consideration when fixing the default values of the variables? Menzies gives this generalization:

> In very general terms, the default values of the exogenous variables represent the initial state of a system of a given kind that is normal or to be expected or is taken for granted because it requires no explanation (220)

The rangers expected the presence of oxygen but not lightning strikes. The Venusians had the opposite expectations. The expectations that define default values can come not only from generalizations but also from normative expectations. That occurs in **Pen**. Professor Smith should not take a pen, and hence, the default value of a variable representing it is a value representing that he does not take a pen.

So far, I have presented Menzies's theory. I will not ponder its merits, but there is a question I need to discuss given that I will apply it to a problem about moral responsibility. How subjective is actual causation in this theory?[4] Menzies does not take a clear stance on this issue. He sees his theory as a form of perspectival-realism (193), but it is not clear exactly what this entails. I think that the answer is that it depends on how realists we are about the normality of the modeled situations. Given that causation is divergence from normality, we are as realists about causation as we are about normality. If the normal behavior of a system is just what one thinks is normal for it, we have an extremely subjective view of causation. Each person can have her own view of normality and end up with different true causal assertions.[5] We can also think that what is normal is set at a social level. Hence, we would have a form of cultural relativist view of causation. We can also think – and that is the view I will assume – that there are objective normality considerations we can appeal to when trying to settle what is normal. For instance, we can assume that the frequency of a type of occurrence sets whether it is

[3] We may think, however, that the Venuasians got to the wrong conclusion. Given that the fire occurred on Earth, they should draw conclusions about normality having the Earth in mind. Hence, they should model the situation with the same default values as the forest rangers. This reasoning points out an interesting variation of Menzies's theory. Instead of the conversational context of the evaluators of causation setting the default values, the modeled situation sets it. We have, therefore, a more objective view of causation and causal selection. I will not discuss the merits of this other theory. For my purposes, I will keep assuming Menzies's theory. I thank Renato Valvassori for bringing my attention to this possible diagnosis of the Venusians' scenario.

[4] I thank Eduarda Calado for questioning me on this matter.

[5] It would still not be a complete relativism, given the assumed objectivist view of causal models.

a statistically normal occurrence. Moral facts or institutional rules can also give us objective normality considerations. We get contextualism because of different normality considerations being appealed to. Nevertheless, not every understanding of what is normal will be acceptable. If there is no normality consideration in favor of a given understanding of the situation, we cannot accept it as the normal.

This question about how realists we are about causation may seem important given its connection to moral responsibility. Moral responsibility is a morally charged phenomenon, so we may worry about grounding it in a subjective phenomenon. However, the link between responsibility and causation seems to be undeniable. If we are to conclude that causation is subjective, we may need to accept that responsibility also is subjective or propose a revisionist theory of responsibility. Hence, the possible subjectiveness of causation in this theory does not pose an impediment to using it to investigate moral responsibility.

In this section, I presented Menzies's Causal Contextualist theory. In the next chapter, I will propose a compatibilist theory based on it.

## 4.4 Final remarks

This chapter was dedicated to causal selection. I first presented it through two examples. After that, I discussed egalitarianism, the standard philosophical position regarding causal selection. I argued that we should not take it for granted and should seriously consider alternative views. I then presented Peter Menzies's theory of causation, which accounts for causal selection using the notion of default. In the next chapter, I will relate his theory to the problem of determinism.

One thing missing in this chapter is a discussion of the possible philosophical consequences of **I. Selective Descriptive Concept**. This view can motivate an interventionist view of causation (Hitchcock and Knobe, 2009), but possible consequences for our understanding of moral responsibility have not been explored. This is left as a possible continuation of the work done in this dissertation.

# 5 Causal Contextualist Compatibilism

In this chapter, I will propose a compatibilist theory based on Menzie's contextualist theory of causation, which I presented in the previous chapter. In the first section of this chapter, I will formulate and motivate the compatibilist theory. In the second section, I show how we can use this theory to give a soft-line reply to the Four-Case Argument that breaks free from the cycle of new cases pointed out by McKenna (2008a). One advantage of my soft-line reply is that it can explain why other soft-line replies partially account for Plum's non-responsibility and why they are vulnerable to new cases. In the third section, I will discuss this advantage and exemplify it with two soft-line replies: Xiaofei Liu's another-agent reply (2022) and Andrew J. Latham and Hannah Tierney's Strawsonian reply (2022). In the fourth section, I discuss how the theory I proposed relates to contextualism about moral responsibility. In the fifth section, I will anticipate and reply to some possible objections to my theory. In the sixth section, I will conclude by discussing the merits and limits of it.

## 5.1 Causal Contextualist Compatibilism

In this section, I will use Menzies's causal contextualist theory presented in the previous chapter to sketch a compatibilist theory. I will name the theory "Causal Contextualist Compatibilism" (shortened CCC).

What does a compatibilist theory need? We can see a compatibilist theory as having two parts. First, the compatibilist needs to give necessary conditions for responsibility that can be satisfied by an ordinary agent in a deterministic world. These conditions should show the control one can exercise over one's actions even if determinism is true. After that, she needs to argue that these conditions are jointly sufficient for responsibility.[1] In other words, she needs to argue that there is no necessary condition for responsibility that cannot be satisfied by an ordinary agent in a deterministic world. Let us call such a necessary condition an *incompatibilist condition*.

I will not discuss the first part. We already have many good proposals about the necessary control conditions we can satisfy in a deterministic world. Causal contextualism will not help the compatibilist in *that* field. The theory I am proposing here is compatible with main compatibilist theories of control. We can adjunct it with a reason-responsiveness theory or a deep-self theory.

Causal Contextualism will be important in order to show that there is no incompatibilist condition. However, before talking about how it will do that, let us discuss what the task amounts to. What must the compatibilist do in order to show the absence of such a condition? One possible answer is that she only needs to show that all proposed incompatibilist conditions are not

[1] She can also identify many different sets of conditions that are jointly sufficient for responsibility, but let us assume for simplicity that there is only one set of jointly sufficient conditions.

necessary for responsibility. In other words, it is enough to show that no proposed incompatibilist argument is sound. As we have seen in chapter 1, that is Kadri Vihvelin's view of the dialectics around the problem of determinism.

I think, however, that this falls short of what a compatibilist theory should accomplish. Incompatibilism has some intuitive appeal and is, for many lay people, their initial position when thinking about the issue. The compatibilist should aim to get whatever grain of truth these intuitions have and propose an error theory for the intuitions that lead some people to embrace incompatibilism.[2] That is what I aim for. I will take some incompatibilist conditions and assume something similar but show that, given causal Contextualism, it does not entail incompatibilism.

The main idea behind source incompatibilism is that the existence of a sufficient cause for an agent's action outside her control undermines responsibility. We see this in Pereboom's conclusion of the Four-Case Argument: "The salient factor that can plausibly explain why Plum is not responsible in all of the cases is that in each he is causally determined by factors beyond his control to decide as he does" (Pereboom, 2014, 79). Robert Kane accepts a condition for responsibility in the same vein: "The basic idea of [Ultimate Responsibility] is this: *To 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, 14). The Causal Contextualist Compatibilist will say that they are basically right about these conditions but are wrong about the conditions entailing incompatibilism. They reached the wrong conclusion because they accepted a false theory of causation.

So, the Causal Contextualist Compatibilist will also accept that an agent is responsible only if there is no sufficient cause for her action beyond her control. Let us call this the *No Other Cause* principle:

**NOC** Agent *S* is responsible for her action *A* only if there is no actual sufficient cause for *A* such that *S* has no control over it.

CCC can motivate NOC by appealing to the same considerations that incompatibilists appeal to when defending their view. The problem for CCC is to explain why NOC does not entail incompatibilism. The answer is that, given its theory of causation, the remote past (usually) is not a cause of one's action.[3] The incompatibilists were right about NOC but reached the wrong conclusion about determinism because they assumed that the remote past would be a cause of our actions.

Remember that, in Menzies's theory, a cause needs to be a disturbance in the normal behavior of the system. But the remote past simply is not such a disturbance. There is no

[2] We could also aim higher: the compatibilist should give a positive argument for compatibilism. There are good positive arguments for compatibilism, like Lewis (2020), but I do not believe it is necessary for a compatibilist theory. Unfortunately, I will need to leave this claim undefended for the sake of brevity.

[3] I focus on the remote past to keep the discussion around the problem of determinism in its usual terms. For CCC, there is no relevant difference between the remote past and the immediate past of one's action so far both are normal and, hence, are not causes.

consideration one can appeal to to understand the remote past's actual state as an abnormality compared to some other possible state. Hence, the remote past is not a cause of our actions.

We can also formulate this point by using the SE framework. When modeling our actions' antecedents, the default value of a variable representing the remote past will be equal to its actual value. Because of that, the remote past will not be a cause of our actions according to Definition 2. Therefore, agents in a deterministic world can satisfy NOC.

Before proceeding, I need to make two clarifications. First, the remote past fails to be a cause of our actions because of its normality, not because of its temporal distance in relation to our actions. There is no restriction on the temporal distance between a cause and its effect. Second, the remote past may be a cause of some events. It can be abnormal when we are seeking a cause for some events. For instance, the distribution of mass in the universe can be caused by an asymmetrical distribution of mass in the early history of the universe. However, when we are searching for a cause of *our actions*, the remote past will be taken as normal. Therefore, it will not be a cause of our actions.

There is a problem that CCC needs to address. If default values are defined contextually, maybe there is a context such that the remote past is a cause of our actions. If this is the case, it seems that or incompatibilism is true, or we need to accept a form of contextualism about moral responsibility.

The objection here is that there is a model according to which the remote past is a cause of our actions. Let us call this model the *Incompatibilist Model*. If we can evaluate causal claims from it, then incompatibilism is true. However, we are assuming contextualism about causation after all. Hence, we can just change our conversational context to evaluate causation from the Incompatibilist Model. Therefore, CCC collapses into incompatibilism or into contextualism about moral responsibility. In the fourth section, I will discuss how CCC relates to contextualism about moral responsibility, but this future discussion is irrelevant to solving the present problem. So I will focus now on the possible collapse into incompatibilism.

This objection gets Causal Contextualism wrong. The default model we use is defined contextually, but it cannot be any model we want. Remember that default models are supposed to map the *normal* behavior of things. We can appeal to many different considerations when trying to set what is normal – and here conversational context matters –, but we cannot define it arbitrarily. We cannot evaluate causation from the Incompatibilist Model because there is no consideration about normality we can use to have the default values of the Incompatibilist Model.

A similar objection is that it is possible to exist a normality consideration that renders the remote past an abnormality in relation to our actions. If so, there are possible scenarios where causal determinism precludes responsibility according to CCC. The objection is correct. Given that Causal Contextualism accepts that norms have a role in setting the default values of variables, we can construct a scenario where the remote past is a cause of an agent's action. We would need a society with certain expectations about the relation between actions and the remote past. Something along these lines: A powerful god ought to have made the universe in a way that precluded evil actions. There are, however, evil actions because the god failed with her duty.

With this cosmological view, the remote past is a cause of every wrong action. This kind of case could pose some trouble for CCC, but I think it is fair for CCC to dismiss these cases. Even if determinism would preclude our responsibility in this case, we have a satisfactory compatibilist solution to the problem of determinism: determinism does not preclude *us* being responsible for our actions.

So far, I have only sketched CCC. The next four sections will help to flesh it out.

## 5.2 The soft-line reply

I have sketched Causal Contextualist Compatibilism in the previous section. Now, I will show how it responds to the Four-Case Argument. As it was mentioned above, in this theory, the remote past does not cause our actions. That will be the key ingredient in the soft-line reply I will propose. One advantage of this reply is that, as I will show, there is no way to reformulate the manipulation cases to get a Four-Case Argument that bypasses it.

Against the Four-Case Argument, CCC has an easy, but compelling, soft-line reply. In cases 1 and 2, Plum's decision does not satisfy NOC. In those cases, there is an actual cause of Plum's action that is beyond his control, the neuroscientists intervening. In Case 4, however, Plum does satisfy NOC. As it was said, the remote past is not abnormal, so it does not count as an actual cause of human actions.[4]

We have two desiderata for a good soft-line reply. First, the necessary condition for moral responsibility defended should be motivated in a principled way. Second, if a manipulated agent satisfies the condition, we do not have a strong intuition of non-responsibility. As said, NOC is very similar to what incompatibilists accept. Therefore, they cannot object to CCC's soft-line reply by arguing that NOC is unmotivated. The second desideratum is where CCC's reply to manipulation arguments really shines. As I will show, there cannot be a manipulation case where the manipulated agent satisfies NOC. Therefore, the incompatibilist cannot reformulate her argument to attack CCC.

The incompatibilist cannot generate a scenario where a manipulated agent satisfies NOC because manipulation is the kind of intervention in the normal course of things that Causal Contextualism classifies as a cause. In all manipulation cases, the manipulator makes a difference in the agent's action that differs it from normal actions. To propose a manipulation argument against CCC, we would need a case where the manipulation is not an actual cause of the manipulated agent's decision. Hence, the manipulation should be something normal. Not only statistically but also normatively. After all, the lesson from **Pen** is that moral norms affect what counts as a cause. The manipulation could not be magical or physically extraordinary, given that this substantial divergence from the natural order would also make it a cause. When we consider all these things, it seems that this is not manipulation at all. How could a banal

[4] Whether Plum satisfies NOC in Case 3 will depend on how we flesh-out the scenario. If his upbringing is not sufficiently abnormal, it will not count as a cause and he will satisfy NOC. Otherwise, it is a cause and he does not satisfy NOC.

event, statistically normal and without moral valence, generate the non-responsibility intuition necessary for a manipulation argument? Therefore, the Causal Contextualist Compatibilist does not need to worry that the incompatibilist will create a manipulation case that endangers her position.

In this section, I proposed a soft-line reply against the Four-Case Argument using CCC. This soft-line reply has the advantage of not being open to new manipulation scenarios.

## 5.3 What other soft-replies got right

CCC's soft-line reply has the benefit of being able to explain what other soft-line replies got right. A soft-line reply may point out a difference between Case 1 and Case 4 that partially explains why the manipulation is a cause, but determinism is not. CCC can account for this partial success while going the extra mile of preventing new manipulation scenarios. In this section, I will exemplify this with two soft-line replies. Each subsequent subsection deals with one of them. Respectively, Xiaofei Liu's another-agent reply (2022) and Andrew J. Latham and Hannah Tierney's Strawsonian reply (2022).

### 5.3.1 Xiaofei Liu's another-agent reply

Xiaofei Liu (2022) argues that the presence of other agents controlling Plum's action is the factor undermining his responsibility in Cases 1 and 2. He calls this an "another-agent reply". As we have seen, Pereboom's reply to this kind of soft-line reply is to propose a reformulation of the cases using a machine instead of an agent. According to him, the non-responsibility judgment remains. Liu considers such a case (a version of Case 2):

> **Machine Induction** Plum grows up in an environment that is saturated with radio signals randomly and spontaneously sent out by a machine, which is not designed or controlled by any agent. These signals work directly on Plum's sensory organs to cause him not only to have the character, the reasoning patterns and the value system that he has, but also to think in particular ways in various circumstances by, for instance, presenting certain stimuli to arouse particular reactions. Despite all this, the signals happen to work in such a way that Plum thinks just like an ordinary person and he satisfies all the compatibilist conditions for moral responsibility (for example, he is moderately reasons-responsive and his effective first-order desires conform to his second-order volitions). Under the causal influence of these radio signals, Plum designs a plan and kills White (536)

Liu defends a hard-line reply to *Machine Induction*. He does that by following McKenna's strategy. He asks us to consider this other case:

> **Multi-Machine Induction** $\text{Plum}_2$ grows up in an environment that is saturated with radio signals randomly and spontaneously sent out by millions of machines, which are not designed or controlled by any agent. These signals work directly on $\text{Plum}_2$'s sensory organs to cause him not only to have the character, the reasoning patterns and the value system that he has, but also to think in particular ways in various circumstances by, for instance, presenting certain stimuli to

> arouse particular reactions. Despite all this, the signals happen to work in such a way that Plum2 thinks just like an ordinary person and he satisfies all the compatibilist conditions for moral responsibility (for example, he is moderately reasonsresponsive and his effective first-order desires conform to his second-order volitions). Under the direct influence of these radio signals, $\text{Plum}_2$ designs a plan and kills White. (Liu, 2022, 537)

The number of machines manipulating Plum seems utterly irrelevant to moral responsibility. Hence, we should have the same judgment about *Machine Induction* and *Multi-Machine Induction*. The latter case, however, is very similar to an ordinary deterministic world. The difference is that we get these stimuli from our environment instead of machines. Given that, Liu says "since compatibilists believe that an actor is morally responsible for an ordinarily causally determined action in the actual world, they have good reason to believe that Plum2 is morally responsible for his action in the world of *Multi-machine Induction*" (538). If compatibilists should judge $\text{Plum}_2$ responsible, they should also judge Plum responsible in *Machine Induction*. If they accept this but deny that Plum-1 is responsible, they accept an another-agent reply. This is the first argument Liu presents for his soft-line reply.

The second argument is a dilemma against soft-line replies that are not another-agent replies (543-547): Plum is responsible in *Machine Induction* or not. If he is, but the presence of other agents (instead of machines) is not relevant for attributions of responsibility, then Plum is responsible in cases 1 and 2; the soft-line reply ended up being a hard-line reply. If Plum is not responsible in *Machine Induction*, then $\text{Plum}_2$ is not responsible in *Multi-Machine Induction*. However, if $\text{Plum}_2$ is not responsible, compatibilism is false, given *Multi-Machine Induction*'s similarity to the actual world. Therefore, soft-line replies that are not another-agent replies are inconsistent.

Does Liu's soft-line reply satisfy the desiderata for soft-line replies that we discussed in the previous section? No, but these were not desiderata for him. Let me start by explaining how it fails the desiderata. Liu explicitly does not give a principled reason for the presence of other agents mattering for attributions of moral responsibility. What he does offer is reasons for a convict compatibilist to accept that this presence matters. Likewise, he accepts that Plum is responsible in *Machine Induction* because, otherwise, he would have to reject compatibilism. Hence, Liu's reply satisfies neither of our desiderata. That is not a shortcoming of his reply, though. Liu's proposal is better understood as a conditional: if compatibilists are to accept a soft-line reply, they should accept an another-agent reply. As he concludes his paper: "my arguments give soft-liners a reason for taking the another-agent reply more seriously." (547). For a discussion of Liu's arguments, see Marco (2022).

CCC can account for the impact of other agents in our judgments of responsibility. Moreover, it can accept that Plum is not responsible in **Machine Induction** while avoiding Liu's arguments. Given that (i) usually our actions are not intentionally determined by other agents, and (ii) manipulating someone in order to produce a morally bad outcome is wrong, a person manipulating someone is a considerable deviation from the normal history leading up to a human action. So it would usually count as a cause according to Menzies's theory. Hence, agents

manipulated in this way would not satisfy NOC. CCC can, therefore, explain why another agent's presence in the history leading up to an action can undermine one's responsibility.

Despite agreeing with Liu about the importance of another agent in the scenario, CCC can avoid the unpalatable conclusion that Plum is responsible in **Machine Induction**. Liu's arguments depend on our non-responsibility judgment about **Machine Induction** carrying on through **Multi-Machine Induction** all the way to an agent acting in an ordinary deterministic world. This last step, from a bizarre scenario to normality, is where we go from an agent not satisfying NOC to one that does satisfy. Therefore, CCC can explain why Plum is not responsible in **Machine Induction** while one can be responsible in deterministic worlds.

### 5.3.2 Andrew J. Latham and Hannah Tierney's Strawsonian reply

I will now pass to Andrew J. Latham and Hannah Tierney's (2022) soft-line reply. They distinguish two kinds of manipulation:

**Existential manipulation** At least one agent, but not all, is manipulated;

**Universal manipulation** All agents are manipulated.

We can see that Cases 1-3 are cases of existential manipulation, but Case 4 is a case of universal manipulation. According to them, this difference is relevant for attributions of responsibility if we accept a strawsonian theory of moral responsibility. In other papers, they present empirical evidence that the difference between existential and universal manipulation affects our pre-theoretical attributions of responsibility in conformity to their theory (Latham and Tierney 2021; Latham et al. 2023).

The difference between existential and universal manipulation matters in a strawsonian approach because, in the universal case, and only in the universal case, our non-responsibility judgment would result in a systematic revision of our responsibility practices. In this view of responsibility, one agent can be exculpated, but not all agents.[5] If Plum-1 is not responsible, we need to excuse him; this is common practice in our system of responsibilization. However, "taking the truth of universal determinism to exculpate Plum in Case 4 would require an impossible feat: the elimination of our responsibility practices and the relationships that these practices are built upon." (Latham and Tierney, 2022, 152). Maybe the situation would not be so drastic, we could preserve some practices in a revised form. Nevertheless, there is still enough difference between Case 1-3 and Case 4 to reject the *No-Difference Premise*.

Let us see if their proposal satisfies our desiderata. It does meet the first one; there is a principled reason for the pointed-out difference between the cases. The reason in favor is the Strawsonian approach. That makes their proposal committed to a specific set of theories, but a popular set. However, they have a problem dealing with the second desideratum.

[5] For an exegetical discussion of Strawson's thesis that statistical normality matters for responsibility, see Hieronymi (2020), especially Chapter 2.

Given the distinction between existential and universal manipulations, the proponent of the Four-Case Argument should amend Cases 1-3 to be cases of universal manipulation. This way, she avoids Latham and Tierney's soft-line reply. About this possibility, they say:

> In the universal manipulation formulations of Cases 1–3, every other agent in the universe is the target of the exact same kind of manipulation, including those who manipulated Plum. So, if Plum is not responsible in these cases in virtue of being manipulated in the way that he was, then no one is. And this would mean that the practices of holding others morally responsible, and the relationships that make these practices appropriate and fitting, would be destroyed or significantly altered. But, as Strawson argues, "The human commitment to participation in ordinary interpersonal relationships is... too thoroughgoing and deeply rooted" to take this possibility seriously. So, it cannot be the case that Plum is not morally responsible for his decision in virtue of being manipulated in the way that he was. Thus, the Strawsonian can reject the Manipulation Premise. (155)

As we can see, they are recommending a hard-line stance against these cases of universal manipulation. They also have a theoretical reason in favor of this stance, given the Strawsonian view they accept. However, there is a problem with how they imagine such cases of universal manipulation. We can distinguish two pairs of existential-universal manipulation:

**Worldly-Existential manipulation** At least one agent in the universe, but not all, is manipulated;

**Worldly-Universal manipulation** All agents in the universe are manipulated.

**Socially-Existential manipulation** At least one agent in a specific moral community, but not all, is manipulated;

**Socially-Universal manipulation** All agents in a specific moral community are manipulated.

What will be important for my argument is that Socially-Universal manipulation does not entail Wordly-Universal manipulation. As they make clear in the first sentence of the last quote, Latham and Tierney think the incompatibilist needs to use Worldly-Universal manipulation scenarios against their proposals. I will now argue that (i), given the Strawsonian background we are accepting, the incompatibilist needs to appeal only to Socially-Universal manipulation cases and that (ii) accepting Socially-Universal manipulation cases should not affect our responsibility practices may be a bullet too hard to bite, even for Strawsonians.

Strawson stresses the importance of moral responsibility in our interpersonal relationships. Strawsonians usually develop theories of moral responsibility focusing on our responsibility practices' role in our moral communities. For instance, communication theories of responsibility understand our reactive attitudes as ways to communicate with the person we are holding responsible and other people in our community (Watson, 1987; Fricker, 2014; Macnamara, 2013; McKenna, 2012). Given that, a Strawsonian should not hold that the appropriateness of the totality of our responsibility practices depends on agents outside our moral community.

To illustrate my point, let us imagine the following scenario. We discover that, light-years away from us, exists a humanoid alien species that exercises an extremely high level

of control over their actions and mental life. Akrasia is unknown to them and they can easily get rid of undesired emotions. They do have a complex emotional life, but they can always ponder their decisions with a clean mind. Also, suppose that these aliens – after discovering that we know about them – use their advanced technology to make any more contact between us and them impossible. What should strawsonians say about this scenario? It seems that Latham and Tierney think that strawsonians are committed to saying that the existence of these aliens is a threat to us being morally responsible agents. In their view, if nobody in the universe satisfies condition $C$, then condition $C$ is not necessary for moral responsibility. Prior to discovering the aliens, we believed that no one satisfied the condition of having an impeccable control over their emotional lives. So, we could assume that this kind of control was not necessary for moral responsibility. However, after discovering the aliens, we come to know that there are agents with this level of control. Therefore, it opens the possibility that having this kind of control is necessary for responsibility. Given that we do not have this control, discovering the aliens open the possibility that we are not morally responsible agents.

But that is an odd conclusion, especially for someone who accepts a Strawsonian theory. We are a moral community of agents that struggle to balance our emotions with our rationally desired actions; therefore, this struggle does not threaten our responsibility. It is a part of our agency that we all need to deal with. If the aliens were to become part of our community, that could justify a change in responsibility practices, but no change is warranted if they are not part of our community.

Given that, Latham and Tierney's reasoning in favor of universal manipulation not undermining moral responsibility in a Strawsonian approach works only with Socially-Universal manipulation. Hence, the incompatibilist arguing against their soft-line reply needs to propose a case of Socially-Universal manipulation that seems to undermine responsibility. A case along these lines:

**Socially-Universal Case 1** A very advanced alien civilization likes to toy with humanity. Every human has a chip in their mind that some of these aliens can use to manipulate our actions. These aliens manipulate all our actions: some directly by using the chip, some indirectly by being the foreknowledge consequence of a directly manipulated action. They, however, do not have a grasp on human morality or values. They manipulate us as part of a very fancy board game of them. Part of the rules of this game is that, when directly manipulating someone, they need to do it in such a way that the human satisfies a set of conditions when acting: her action is not out of character; she does not act from an irresistible desire; her desire is in accordance to her second-order desires. She is reason-responsive and has the general ability to grasp, apply, and regulate her actions by moral reasons. One day, one of the aliens presses a button, activating the chip in Plum's head to start a deterministic causal chain leading to his decision to kill White (the alien's objective). Satisfying all the listed conditions, he decides to kill White and proceed to murder him.

I think our non-responsibility judgment about this case is much stronger than if it

were a case of Wordly-Universal Manipulation. If so, the distinction between those two kinds of Universal manipulation matters. We could adapt Cases 2 and 3 to be cases of Socially-Universal manipulation. With that, we have a Four-Case argument that avoids their soft-line reply.

Could they also take a hard-line stance regarding Socially-Universal Case 1? They could appeal to the strawsonian thesis that completely abdicating our responsibility practices is humanly impossible. With that, they could argue that we should not regard everybody in Socially-Universal Case 1 as not responsible. Against that, the argument's proponent can concede that responsibility practices should not end in Socially-Universal Case 1 but should at least be significantly revised. These revisions could go something similar to what Pereboom has proposed we ourselves should do (2014, Chapter 6). Therefore, concluding that we should significantly revise our responsibility practices is good enough for Pereboom's aim. In conclusion, their soft-line reply struggles to accommodate cases like Socially-Universal Case 1.

CCC can explain why the quantity of agents subjected to manipulation affects responsibility attributions. At the same time, its soft-line reply still applies to Socially-Universal Case 1. Hence, it accounts for what Latham and Tierney got right in their soft-line reply while avoiding its problem.

According to Latham and Tierney, how many agents are subjected to the threat to responsibility is relevant to responsibility attributions. CCC agrees with them. The statistical normality of the threat to responsibility is important to responsibility attributions, but not because of strawsonian considerations. It matters because statistical normality is one factor that defines variables' normal values. With this, CCC can explain the empirical data they appeal to.

Nevertheless, Plum is not responsible in Socially-Universal Case 1. Despite his action being statistically normal, it is normatively abnormal. People's actions should not be manipulated as part of an alien's board game. This deviation from normality is enough to make the alien's action a cause of Plum's action. Hence, Plum does not satisfy NOC in Socially-Universal Case 1. CCC's soft-line reply works against this case as well.

In this section, I showed how CCC can account for what other soft-line replies got right while avoiding their problems. That shows that the soft-line reply presented in the previous section has a greater explanatory power in relation to other soft-line replies. Moreover, it can deal with cases that are troublesome to them.

## 5.4 The risks of contextualism

I have presented CCC and shown how it deals with incompatibilist arguments. However, there is a looming problem with it. If we accept Causal Contextualism and ground responsibility on actual causation, are we not committed to contextualism about moral responsibility? Let us call the thesis that responsibility attributions are contextually sensitive "Responsibility Contextualism". The entailment between CCC and Responsibility Contextualism can be weaponized against CCC in many forms. In this section, I will discuss the relation between

CCC and Responsibility Contextualism. This section has two parts. In the first, I will defend that we can have CCC without Responsibility Contextualism. In the second, I will defend the conjunction of CCC and Responsibility Contextualism against some objections.

### 5.4.1 Responsibility-Invariant Causal-Contextualist Compatibilism

In this subsection, I will defend that we can have CCC without Responsibility Contextualism. I will call this version of CCC *Responsibility-Invariant Causal-Contextualist Compatibilism*, or *RICCC* for short. The problem for RICCC is to explain why we do not get Responsibility Contextualism from CCC.

I have already shown how Causal Contextualism has some constraints on which models we can use to evaluate actual causation. We can only use default models that represent the system's normal behavior. There are cases, however, where we have different considerations about normality that produce different views of what is normal for the system. In these cases, we can evaluate causation from multiple models and may come to different conclusions about responsibility. Consider, for instance, the following case:

**Broken Promise** A doctor recommended a diet to Talita because her weight may cause some health problems in the future. Talita is unhappy about entering a diet, but her husband insists. He insists so much that she ends up promising to follow the diet. In a moment of stress, however, she succumbs to temptation and eats what she is not supposed to. When he discovers this, the husband becomes eager to blame her. When he complains to a friend about Talita breaking the promise, the friend brings attention to all the social factors that may play a role in Talita's difficulty in following the diet. She says that, in our society, weight is a complex matter for women and is attached to all sorts of body image issues. Not only that, but that stressful obligations and schedules leave us with almost no time to plan and prepare healthy meals; that many ultra-processed foods have enough fat and sugar to be quite literally addictive, and so forth.

This kind of case may seem troublesome for theories of responsibility in general, but it is especially problematic for RICCC. To see that, let us first model Talita's action in light of the social environment pointed out by the friend. For simplicity's sake, we will have only two variables: $E$ for the social environment and $T$ for Talita. These variables can have value 1 or 0. $E = 1$ represents that the social environment has all the problems highlighted by the friend; $E = 0$ represents the absence of these problems. $T = 1$ represents that Talita breaks the promise; $T = 0$ represents that Talita does not break the promise. We have, then, the following model $< E; T; T = E >$.[6] If the default value of $E$ is 1, Talita satisfies NOC; if it is 0, she does not. Therefore, whether she is morally responsible for breaking the promise depends on the default value for $E$.

[6] Given our objectivist view about models, it is plausible that this model does not actually exist. Correctly modeling the relation between someone's action and social factors that produce it requires way more variable and more complex structural equations. Nevertheless, this example will help to elucidate RICCC.

There is more than one way to think about what is normal in this situation. On the one hand, our social environment has existed for a long time and is usually seen as a given when thinking about the causes of our actions. If we follow this reasoning, we end up with 1 as the default value for $E$. Let us call this default model S-Model, given its focus on statistical normality. On the other hand, the current social environment is abnormal because it should not be as it is.[7] All factors pointed out by the friend are injustices or flaws in our social organization. If we follow this reasoning, we end up with 0 as the default value for $E$. Let us call this default model N-Model, given its focus on normative normality.

If we evaluate causation using the S-Model, Talita satisfies NOC; if we use the N-Model, she does not. This is a problem for RICCC because, according to it, Talita's responsibility does not depend on the context of evaluation.[8] RICC needs to amend NOC to specify how it deals with these cases. There are three routes available:

**One Model NOC** Agent $S$ is responsible for her action $A$ only if there is at least one default model such that there is no actual sufficient cause for $A$ such that $S$ has no control over it.

**All Model NOC** Agent $S$ is responsible for her action $A$ only if, according to all default models available, there is no actual sufficient cause for $A$ such that $S$ has no control over it.

**Specific Model NOC** Agent $S$ is responsible for her action $A$ only if, according to a specific default model available, there is no actual sufficient cause for $A$ such that $S$ has no control over it.

The two first routes accept that all models are equally insightful for attributing responsibility, but differ in the result they arrive at about **Broken Promise**. With **One Model NOC**, we need the agent to satisfy NOC in one model. One evaluation of the situation that renders the agent responsible is enough to hold her responsible. Hence, Talita is responsible according to **One Model NOC**.

**All Model NOC** takes a different approach. The agent must satisfy NOC in all models. One evaluation of the situation that excuses the agent is enough to render the agent not responsible. Hence, Talita is not responsible according to **All Model NOC**.

We may think, however, that when we have more than one model, moral responsibility is tied to only one of them. That is the idea behind **Specific Model NOC**. We can get an answer about **Broken Promise** only if we establish a criterion for deciding which model determines responsibility. For instance, maybe models related to normative considerations are more important than models related to statistical considerations.

I will not argue for any of these options. To show that RICCC is a viable option, I only need to show that these routes are not *ad hoc*. This latter claim is not hard to argue for. Our

[7] I am assuming here that there are objective moral facts about how our society morally ought to be organized. Hence, the sense of "should" I am appealing to here is moral.

[8] RICCC can deal with this problem without amending NOC by defending that each model has a different responsibility related to it. For instance, the S-Model may be relevant for the husband's blame; the N-Model may be appropriate for a policymaker trying to attribute responsibility for people's weight.

three options can be understood as three hypotheses about how responsibility attributions work. A discussion about what is fair in **Broken Promise** or how we should proceed in evaluating someone's responsibility can settle which version of NOC is preferable.[9]

In this subsection, I showed how CCC can reject Responsibility Contextualism. The resulting theory, RICCC, has to deal with some complicated cases, but many options are on the table here. Therefore, RICC is a viable theory for someone attracted to CCC, but who does not want to embrace Responsibility Contextualism.

### 5.4.2 Double Contextualist Compatibilism

Maybe, however, the correct answer to **Broken Promise** is that "Talita is responsible for breaking the promise" is true in a context where we use the S-Model and false in a context where we use the N-Model. Let us suppose, then, that we want to complement CCC with Responsibility Contextualism. Let us call the conjunction of these two theses *Double Contextualist Compatibilism* (DCC). The addition of Responsibility Contextualism to CCC may seem problematic for several reasons. I think there are three pressing objections that DCC needs to respond to. In this subsection, I will list them and offer replies.

The objections are the following:

**Unacceptability** Responsibility Contextualism is an unacceptable position. Responsibility attributions have morally charged consequences about how we should treat people. They cannot change according to flimsy elements of the conversational context, such as the underlying causal model.

**Nothing New** We already have proposals of contextualist views of free will or moral responsibility (Hawthorne, 2001; Rieber, 2006; Willaschek, 2010; Jeppsson, 2019; Whittle, 2021).[10] Moreover, these views can deal with the problem of determinism. DCC is just a less parsimonious view – given its commitment to a specific theory of causation – and adds nothing useful to existing responsibility contextualist theories.

**Not a Solution** DCC, like other contextualist theories, offers, at best, a dissolution of the problem of determinism. It shows how we can preserve incompatibilist intuitions in some cases but still have responsibility in ordinary contexts. This is not really a compatibilist solution to the problem.

Let us start with **Unacceptability**. As usual in philosophy, we must first make a distinction. Let us distinguish between two kinds of contextualism:[11]

[9] I tend to prefer **All Model NOC**. It seems to me to better capture how we make (and should make) responsibility attributions. If there is one understanding of the situation that gives the agent an excuse, appealing to another understanding does not render the excuse unacceptable.

[10] In this section, I will treat all proposed contextualist theories about free will as extendable to moral responsibility.

[11] I got the term "Elusive Contextualism" from Lewis (1996) and the term "Non-Elusive Contextualism" from Jeppsson (2016). The definitions are my own.

**Elusive Contextualism** Elusive Contextualism about predicate "$F$" is true if (i) the truth-value of sentence "$Fx$" is context-dependent and (ii) a speaker **can** easily change the truth-value of "$Fx$" (without changing the facts of matter);

**Non-Elusive Contextualism** Non-Elusive Contextualism about predicate "$F$" is true if (i) the truth-value of sentence "$Fx$" is context-dependent and (ii) speaker **can not** easily change the truth-value of "$Fx$" (without changing the facts of matter).

Let me exemplify these two forms of contextualism. Suppose that contextualism about knowledge is true. According to Lewis (1996), "I know I have a hand" is usually a true assertion. Even so, a skeptic can make it false simply by bringing attention to skeptical possibilities[12] – "But can you be sure you are not just a brain in a vat?" –. This makes Lewis's Contextualism elusive. Bringing attention to previously ignored[13] facts can change the truth-value of an assertion about knowledge. On the other hand, Michael Williams's brand of Contextualism is Non-Elusive (Williams, 2001, 2007). In his view, not every challenge against knowledge is legitimate. The current concerns in a given situation put constraints on what are acceptable challenges. Bringing attention to a skeptical possibility when doing science, for instance, does not produce a context-shift that precludes us from asserting our scientific knowledge. We can just disregard the skeptical possibility in this context.

At its core, the **Unacceptability** objection is against *Elusive* Contextualism about responsibility. If bringing attention to previously ignored facts can change one's responsibility for something, it seems we do not have a strong foundation for our responsibility practices. Remember that moral responsibility is linked to the appropriateness of reactive attitudes. It seems an absurd conclusion that a reactive attitude – such as moral anger – can become inappropriate because an agent said something that shifted the conversational context.

Some proposed contextualist theories about free will and responsibility can be classified as elusive (Hawthorne, 2001; Rieber, 2006; Whittle, 2021). Bringing attention to some ignored facts can change the truth-value of assertions about responsibility. But not all of them are (Willaschek, 2010; Jeppsson, 2019). Hence, these non-elusive theories avoid the **Unacceptability** objection.

Is DCC an Elusive Contextualism? Given DCC, we can easily change the truth-value of responsibility attributions by changing the underlying causal model. It may seem, then, that it is an elusive theory. However, we can only evaluate causation from existing default models, which need to have their default values supported by an understanding of what is normal for the situation in question. So, we get elusiveness only in very specific cases, when we have more than one model available. Therefore, I think that DCC is non-elusive enough to deal with the

[12] The true utterance of "I know I have a hand" and the false one do not express the same proposition. Hence, paying attention to the skeptical possibility changes the true-value of the assertion by changing the expressed proposition.

[13] "Ignored" in this debate does not mean *unknown*, it means *unnoticed*.

**Unacceptability** objection.[14]

My reply to the **Nothing New** objection is just to point out how DCC differs in interesting ways from existing contextualist theories. I will not discuss these theories in detail; I will just discuss some aspects of them and show how DCC is different.

DCC offers a unique solution to the **Unacceptability** objection. It is not an Elusive Contextualism, so it already differs from these theories. Moreover, it avoids elusiveness in a way unlike existing Non-Elusive theories. In Willaschek's theory, we avoid elusiveness because the context is defined by social practices and not (only) conversational movements. In Jeppsson's theory, we have a meta-ethical reason for why we should assess responsibility attributions from a specific perspective. With this, we avoid elusiveness. There are contexts where the agent is not responsible because of the distant causes of her action, but we should disregard these contexts when attributing responsibility. (It is the same idea we find in her hard-line reply to the Four Case Argument, discussed in chapter 1.) In DCC, however, we do not need to present an additional argument to avoid elusiveness. We get non-elusiveness because of constraints on available default models. In this sense, DCC is more parsimonious. Causal Contextualism provides all the resources for a satisfactory contextualist theory about responsibility; we do not need to amend it with meta-ethical arguments or claims about what ground contexts.

Existing contextualist theories give too much credit to incompatibilist intuitions. All of them concede that there are some contexts where it is true to say that an agent does not have the control necessary for responsibility because of determinism (if determinism is true) (Hawthorne 2001, 69; Rieber 2006, 235; Willaschek 2010, 577; Jeppsson 2019, 806; Whittle 2021, 212-213). In the Non-Elusive theories, this incompatibilist conclusion does not hold in contexts where we are evaluating if we should hold someone morally responsible (Willaschek 2010, 582; Jeppsson 2019, 806). With this, they avoid the **Unacceptability** objection, but I think they still are conceding too much. One may argue it is a theoretical virtue to grant that the incompatibilist is correct in some contexts. Even so, we should not make such a concession if we aim for a compatibilist theory. According to DCC, there is no context where determinism precludes our responsibility (given the reasons exposed in the first section). Hence, this is a difference between DCC and existing contextualist theories.

One advantage of not corroborating incompatibilist intuitions is that, in DCC, reflecting on how elements of the past produced who we are is not at odds with seeing ourselves as free and responsible agents so far there is no actual cause of our actions in these elements. Even in Willaschek's Non-Elusive Contextualism, for instance, when we focus on what produced who we are, we stop being the original cause of our actions, and, in this context, our control over our actions seems threat.[15] This does not affect our reactive attitudes because it is not an

[14] Another option would be to amend DCC with arguments presented in other Non-Elusive Contextualist theories about responsibility.

[15] Willaschek's account is actually more complex than that. We are not the original cause in this context, but the control necessary for responsibility is being the original cause *in the evaluative context*. Therefore, even in this context, we have the control necessary for responsibility. Nevertheless, there is a sense in which our control is at risk: We are seeing that, in some contexts, we are not the original cause of our actions.

evaluative context, but nevertheless, it seems we do not have the opportunity to, at the same time, see ourselves as products of our environment and as responsible agents. I take this to be an unfortunate consequence. Hence, I think it is advantageous for DCC not to lead to this.

Ann Whittle's contextualist theory (2021) is the most similar to DCC. In her theory, we get contextualism about responsibility from contextualism about agential models. Likewise, DCC gets Responsibility Contextualism from Causal Contextualism. We could see both proposals as complementary. Both show that Responsibility Contextualism is true because contextualism is true about the notion that provides metaphysical grounding for responsible agency. In her theory, abilities; in mine, causation. One theory gets Responsibility Contextualism from the leeway view of responsibility; the other from the sourcehood view.

Nevertheless, I think DCC has two advantages over Whittle's theory. First, to deal with the **Unacceptability** objection – that she calls "The Problem of Unfairness" –, she recommends that we disentangle moral responsibility and harmful reactive attitudes (Chapter 9). This is a heterodox thesis that DCC does not need to commit itself to. Second, her theory also has the discussed problem of making the incompatibilist right in some contexts. As I said, this does not occur in DCC.

Last, we have the **Not a Solution** objection. I think that this objection holds some water against other contextualist theories. If we say that incompatibilists are right in some contexts, but not in others (or that they are not right in contexts relevant for holding someone responsible), we are not defending compatibilism; we are defending that the problem of determinism is founded on the misconception that responsibility attributions are invariant. That does not occur with DCC. According to DCC, the incompatibilist is wrong in all contexts. Therefore, we have a solution to the problem of determinism: compatibilism is true.

In this section, I considered how CCC relates to Responsibility Contextualism. I first showed that the former thesis does not entail the latter. We can have CCC with invariantism about responsibility. Then, I considered the theory we get from the conjunction of CCC and Responsibility Contextualism. I defended this theory against three objections. Therefore, CCC does not have a troublesome relation to Responsibility Contextualism.

## 5.5 Objections

In this section, I will consider some objections to Causal Contextualist Compatibilism and reply to them.

**Objection 1:** NOC is false because of cases of overdetermination. We can construct a case in which the agent's deliberation causes her action and, at the same time, some other thing also causes it. NOC will conclude that the agent is not responsible. That, however, is the wrong verdict.

**Reply:** I could try to amend NOC to get these cases right, but instead, I will ignore this complication. After all, overdetermination is a problem for difference-making theories of

causation in general. We should try to solve it in our theory of causation, not in our theory about how causation relates to responsibility. Given that I am just assuming a theory of causation, I will not try to defend or improve it.

**Objection 2:** You did not propose a *compatibilist* theory. You gave a reason to think that if Causal Contextualism is true, causal determinism is false. After all, if the remote past is not a cause of the agent's decision, this means that determinism (in the sense relevant to the debate) is false.

**Reply:** I sympathize with this objection and think it is problematic for some compatibilist positions. However, it does not apply to CCC. The Causal Contextualist Compatibilist does not accept that the remote past is an actual cause of our actions, but she can accept a metaphysically robust necessitation relation holding between them. For instance, a governist theory of laws of nature would guarantee this necessitation, even with Causal Contextualism. If there is a relation between the past and our actions that is strong enough to make justice to our initial worry about determinism, we have determinism in the sense relevant to the debate. That is the case even if the relation in question is not actual causation. Therefore, Causal Contextualist Compatibilism is not denying determinism.

**Objection 3:** Peter Menzies already proposed a compatibilist theory using an interventionist theory of causation (Menzies, 2017). His theory does not presuppose Causal Contextualism, so it is preferable to yours.

**Reply:** Menzies proposed a reply to the Consequence Argument using interventionism. This is not at odds with what I defended in this chapter. His proposal focuses on the leeway view of responsibility; mine, on the sourcehood view. Hence, CCC and his interventionist reply to the Consequence Argument can be seen as complementary.

Besides, Menzies's proposal has some severe shortcomings not found in CCC. It denies the sense of "determinism" relevant to the debate. It also gives an implausible interpretation of the necessity operator present in the Consequence Argument.

I will briefly present Menzies's compatibilist proposal before critiquing it. Menzies responds to the Consequence Argument using interventionism. So, I will start by introducing the version of the Consequence Argument discussed by Menzies. Menzies discusses a slightly revised version of the modal Consequence Argument presented by Peter van Inwagen (1983). We have a modal operator for unavoidability:

**N** $NP \equiv$ I am not able to do anything to render $P$ false.

And we have one deductive rule for this operator:[16]

[16] Instead of $\beta$*, van Inwagen's formulation has two deductive rules:

$\alpha$ $\Box P \vDash NP$

$\beta$ $NP, N(P \supset Q) \vDash NQ$

However, $\beta$ is usually taken to be invalid given a counter-example proposed by Thomas J. McKay and David Johnson (1996). For a criticism of their counter-example, see Merlussi (2022).

$\beta$* $NP, \Box(P \supset Q) \vDash NQ$

Being $P_0$ a proposition describing the complete state of the world in a moment of the remote past, $L$ a conjunction of the laws of nature, and $P$ any true proposition, we have the following argument:

**P1** $\Box((P_0 \wedge L) \supset P)$ Determinism

**P2** $N(P_0 \wedge L)$ Fixity of past and laws

**C1** $NP$ From P1 and P2 by $\beta$*

Menzies argues that, given an interventionist theory of causation, we should reject P1. According to the interventionist theory, the laws of nature are the structural equations for the system modeling the whole universe. However, structural equations are *ceteris paribus*. The value of an endogenous variable is as specified by the structural equation only in the absence of an intervention. Given that, an interventionist should understand laws of nature as *ceteris paribus* and, therefore, P1 is false even if all laws are deterministic. In Menzies's words:

> The only logical implication is this: the actual values of the exogenous variables, taken in conjunction with the structural equations, logically imply the corresponding variables of the endogenous variables *only if no interventions take place in the system*. This means that the first premise of the argument — the unqualified thesis of determinism — is false. (322)

Menzies's argument has two critical problems. The first one is very explicit in the cited passage. Menzies understands the box in the Consequence Argument as logical necessity. However, the box should be understood as metaphysical necessity. Van Inwagen says the box represents "broadly logical necessity" (van Inwagen, 1983, 94). As he himself says (230), he gets this terminology from Alvin Plantinga. And Plantinga (1974, 1-2) used it to refer to what currently we would call *metaphysical necessity*.[17]

How we understand the box determines how strong the central thesis is in Menzies's argument. If we are talking about logical necessity, Menzies needs only to defend that, given an interventionist theory, we should accept the logical possibility of an intervention in the laws of nature. However, if what is at stake is metaphysical necessity, Menzies needs to defend that an interventionist theory makes plausible the metaphysical possibility of the laws being violated. The latter thesis is stronger than the first. We can get the first from how we conceptualize causation in an interventionist theory. The latter, however, we can get only with a more controversial thesis about the nature of laws of nature. Given that we need to understand necessity in this debate as metaphysical, Menzies needs to commit himself to the latter thesis. Therefore, his compatibilist view is also committed to a very controversial thesis. I see no initial reason to prefer it rather than Causal Contextualist Compatibilism.

The main problem, however, is that Menzies's argument is not a compatibilist reply but actually a denial of determinism in the relevant sense. Daniel Nolan and Eddy Nahmias have,

[17] For a discussion about the consequences of reading this box as metaphysical or logical necessity, see Merlussi (2016).

each independently, posed this objection against Menzies on personal correspondence. Replying to Nolan, Menzies said:

> I think that in many scientific disciplines, including many areas of physics, scientists use models that appeal to laws that are said to be deterministic even though they are known to be falsifiable by interventions. (The label "determinism" seems to characterize the functional form of the law rather than its status as exceptionless.) (326)

Menzies may well be right about that, but I think that he is wrong about what the thesis of determinism in the debate about free will and moral responsibility should be. To make sense of the discussion on the problem of determinism, "Determinism" in this debate should refer to a thesis that makes sense of our worry that determinism may be incompatible with responsibility. Menzies's understanding of determinism may make justice of the scientific use of the term but does not capture the worrisome thesis we seek to debate when discussing free will and moral responsibility. His discussion about how interventionist theories relate to determinism is fascinating, but it is not *compatibilist*. His arguments are more compelling when presented as a defense that, given our most advanced philosophical and scientific understanding of laws and causation in our world, determinism, in the sense that makes us worry about free will and moral responsibility, is false. Hence, since he does not have a compatibilist theory after all, his view is not directly concurrent with CCC.

**Objection 4:** Interventionism, especially in a contextualist version, cannot be used as our underlying theory of causation if we want to analyze moral responsibility. In these theories, causal attributions work very similarly to our pre-theoretical attributions of responsibility. Therefore, we cannot derive any illuminating result about responsibility from it. Causal Contextualist Compatibilism can, at best, show why our pre-theoretical attributions of responsibility are not overtly incompatibilist. It cannot, however, make any progress on the problem of determinism.

**Reply:** My reply has two parts. First, it is not a problem that our theory of causation vindicates the way we usually do responsibility attributions so far we think that we are not bad at causal or responsibility attributions. If we do not have an independent reason to believe that we need a revision in our pre-theoretical attributions, it is not a problem that a theory corroborates them.

Second, Pereboom is also appealing to how we pre-theoretically make responsibility attributions when putting forward the Four-Case Argument (2007, 98). He argues that given certain features of our responsibility practices, no one is morally responsible if determinism is true. I am arguing that a mechanism in how we attribute responsibility – NOC (assuming Causal Contextualism) – can show the flaw in his argument. Hence, I am meeting his argument on his own terms.

## 5.6 Final remarks

In this chapter, I sketched a compatibilist theory: Causal Contextualist Compatibilism. It is based on Menzies's contextualist theory of causation. With this theory, we can show why incompatibilism is very alluring but wrong. Incompatibilists are right about causal determination by factors beyond one's control precluding responsibility. What they got wrong is that, even if determinism is true, the remote past is not a cause of our actions. To be a cause, the remote past would need to be a disturbance in the normal course of things, which simply is not the case. I showed how this theory provides a strong soft-line reply against the Four-Case Argument. I also showed that it can account for insights found in other soft-line replies. To make the theory plausible, I considered several objections. First, I discussed how it relates to Responsibility Contextualism. I showed that we can have Causal Contextualist Compatibilism without Responsibility Contextualism. Then, I defended the conjunction of these two theses against three objections. I also discussed a series of minor objections against Causal Contextualist Compatibilism. The overall result is a compatibilist theory that presupposes a lot, given its unorthodox underlying theory of causation, but that delivers an exciting view of the relation between causation and responsibility.

I formulated Causal Contextualist Compatibilism using Menzies's theory, but could I have used another theory of causation? The theory needs to be contextualist, but will any contextualist theory do the job? I think we can get Causal Contextualist Compatibilism out of any theory that satisfies Hart's and Honoré's view of causation as difference-making. If a cause is a disturbance of the system's normal behavior, the remote past will not be a cause of our actions. Therefore, we can have Causal Contextualist Compatibilism. A continuation of the work done here can be to formulate this minimal Causal Contextualist Compatibilism.

# Final remarks

My objective in this dissertation was to argue against the Four-Case Argument. Moreover, I intended to do so using some debates in the metaphysics of causation as a starting point.

The bulk of chapter 1 consisted in a presentation of the Four-Case Argument. I presented the argument, discussed its structures and two hard-line replies against it. I argued that it would be best for compatibilists to weaken the Four-Case Argument without accepting a new condition for responsibility or to propose a soft-line reply immune to new manipulation cases. Each one of the main arguments I proposed followed one of these routes.

In chapter 2, I discussed the problem of extensionality. There is a tension between the appeal of causal extensionality and intuitions we have about some cases. Among the theories discussed, we saw that contrastivism could both preserve causal extensionality and vindicate our intuitions regarding the problematic cases.

Causal contrastivism played a role in my argumentation in chapter 3. My main claim was that Case 1 does not support an incompatibilist conclusion. That is because, in Case 1, Plum is subject to a C-Luck, a kind of moral luck that can plague agents in deterministic worlds. C-Luck is usually characterized as a cross-world difference, but I defined it using contrastive causation. I argued that my definition better captures the phenomenon. Given this problem with Case 1, the incompatibilist has to either modify the case or abandon it. No modification that avoids C-Luck is a good start for the Four-Case Argument. Hence, the only way to go is to abandon Case 1 and start the argument with Case 2. I considered this new argument and argued that it poses less of a threat to compatibilism than the original Four-Case Argument. The overall conclusion is that the current version of the Four-Case argument is inadequate and a proper amendment will result in a weaker argument for incompatibilism.

The starting point of the second part of the dissertation was causal selection. I started chapter 4 by presenting causal selection. I then discussed why we should not take the orthodox philosophical position about causal selection for granted. The orthodox position is egalitarianism, a position that dismisses causal selection as a pragmatic phenomenon or a misuse of the concept of causation by laypeople. I also presented Peter Menzies's theory of causation. He rejects egalitarianism and tries to account for causal selection in his theory. For that, he uses the idea of default. A cause is a deviation from the default behavior of the system that produces another deviation, the effect.

Menzies's theory was the pillar of the compatibilist theory I proposed in chapter 5. I defended that, assuming his theory, we could avoid incompatibilism while accounting for the incompatibilist intuition that we cannot be morally responsible if our action has a cause beyond our control. The key idea behind Causal Contextualist Compatibilism is that, assuming Causal Contextualism, the remote past is not a cause of our actions. A cause needs to be a disturbance in

the normal behavior of the system, but the remote past is not. Hence, even in a deterministic world, most of our actions do not have a sufficient cause beyond our control. I used this theory to put forward a soft-line reply to the Four-Case Argument. Plum is not responsible in the first two cases because there is a sufficient actual cause for their actions that is beyond their control, the manipulators' action. However, in the last case, there is no such cause and, therefore, Plum can be morally responsible. This soft-line reply has the advantage of avoiding the formulation of new manipulation cases. The manipulation necessary for the formulation of a manipulation argument will always count as a deviation from the default way an action is produced. Hence, this soft-line reply will always be available for compatibilists. I also argued that my soft-line reply can account for what other soft-line replies got right while avoiding their pitfalls. This theory has a lot of problems to address, and I tried solving some of them by considering a series of possible objections. We end up with a theory that, despite being hard to accept because of its underlying theory of causation, occupies an interesting spot in the space of possible theories of moral responsibility.

My main arguments are very narrow: they each presuppose a specific theory of causation and are aimed at a specific incompatibilist argument. Nevertheless, they show a possible strategy for compatibilists that has been explored less than it deserves. We can make a case for compatibilism by bringing attention to how causation works. Causation may work in such a way that our worry about causal determinism stems from a misunderstanding about causation.

I will now discuss some possibilities of investigation that were left unexplored in this dissertation. I will do that both to signalize future research as well as to make explicit limitations of the work done here.

In chapter 3, I used contrastive causation to locate C-Luck in the actual causal sequence leading up to the action. It would be interesting to know if that approach could be extended to moral luck in general. Moreover, to strengthen my case in favor of contrastive causation, it would be necessary to consider whether other solutions to the problem of extensionality could also locate moral luck in the causal history of the action.

My discussion of egalitarianism in chapter 4 was too brief to sufficiently motivate its rejection. A more in-depth discussion would be necessary to better motivate Causal Contextualist Compatibilism. I also did not discuss how other views of causal selection beyond the one operating in Menzies's theory could affect our theories of moral responsibility. That is something I want to explore thoroughly in future work.

The theory proposed in chapter 5 needs to be fleshed-out to be a plausible position in the debate. One of the topics that needs to be expanded on is the idea of normality considerations. This notion is essential to the theory, so it needs to be better explained. It would also be interesting to see how Causal Contextualist Compatibilism deals with other source incompatibilist arguments besides the Four-Case Argument. When discussing the relation of this theory to Responsibility Contextualism, I pointed out that it can work both with, and without, Responsibility Contextualism. These variations of Causal Contextualist Compatibilism would also need to be

discussed in detail.

# Bibliography

Achinstein, P. (1975). Causation, transparency, and emphasis. *Canadian Journal of Philosophy 5*(1), 1–23.

Alicke, M. (2008). Blaming badly. *Journal of Cognition and Culture 8*(1-2), 179–186.

Alicke, M. D., D. Rose, and D. Bloom (2011). Causation, norm violation, and culpable control. *Journal of Philosophy 108*(12), 670–696.

Anscombe, G. E. M. (1969). Causality and extensionality. *Journal of Philosophy 66*(6), 152–159.

Baron, R. A. (1997). The sweet smell of... helping: Effects of pleasant ambient fragrance on prosocial behavior in shopping malls. *Personality and Social Psychology Bulletin 23*(5), 498–503.

Bebb, J. and H. Beebee (2024). Causal selection and egalitarianism. In S. Nichols and J. Knobe (Eds.), *Oxford Studies in Experimental Philosophy, Volume 5*. Oxford University Press.

Bebb, J. and H. Beebee (forthcoming). Causal selection and egalitarianism. In J. Knobe and S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy, Vol. 5*. OUP.

Beebee, H. (2004). Causing and nothingness. In L. A. Paul, E. J. Hall, and J. Collins (Eds.), *Causation and Counterfactuals*, pp. 291–308. Cambridge, MA, USA: MIT Press.

Bennett, J. (1988). *Events and Their Names*. Hackett.

Björnsson, G. (2020). Quality of will and radical value reversals. In *PEA Soup Symposium on Al Mele's Manipulated Agents: A Window to Moral Responsibility*.

Buckareff, A. A. (forthcoming). Direct manipulation undermines *Intentional Agency* (not just *Free* agency). *Ergo: An Open Access Journal of Philosophy*.

Campbell, J. K. (2007). Free will and the necessity of the past. *Analysis 67*(2), 105–111.

Campbell, J. K. (2010). Incompatibilism and fatalism: Reply to loss. *Analysis 70*(1), 71–76.

Clarke, R. (2005). Agent causation and the problem of luck. *Pacific Philosophical Quarterly 86*(3), 408–421.

Corkum, P. (2022). Is 'cause' ambiguous? *Philosophical Studies 179*, 2945–71.

Cyr, T. (2019). Moral responsibility, luck, and compatibilism. *Erkenntnis 84*(1), 193–214.

Cyr, T. W. (2020). Manipulation and constitutive luck. *Philosophical Studies 177*(8), 2381–2394.

Cyr, T. W. (2023). Why history matters for moral responsibility: Evaluating history-sensitive structuralism. *Philosophical Issues 33*(1), 58–69.

Davidson, D. (1967). Causal relations. *Journal of Philosophy 64*(21), 691–703.

de Calleja, M. P. (2014). Cross-world luck at the time of decision is a problem for compatibilists as well. *Philosophical Explorations 17*(2), 112–125.

Deery, O. and E. Nahmias (2017). Defeating manipulation arguments: Interventionist causation and compatibilist sourcehood. *Philosophical Studies 174*(5), 1255–1276.

Earman, J. (1986). *A Primer on Determinism*. D. Reidel.

Fischer, E. (2021). Causation and the problem of disagreement. *Philosophy of Science 88*(5), 773–783.

Fischer, J. M. (2000). Responsibility, history and manipulation. *The Journal of Ethics 4*(4), 385–391.

Fischer, J. M. (2010). The frankfurt cases: The moral of the stories. *Philosophical Review 119*(3), 315–336.

Fischer, J. M. (2011). The zygote argument remixed. *Analysis 71*(2), 267–272.

Fischer, J. M. (2016). How do manipulation arguments work? *The Journal of Ethics 20*(1-3), 47–67.

Fischer, J. M. and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy 66*(23), 829–839.

Fricker, M. (2014). What's the point of blame? a paradigm based explanation. *Noûs 50*(1), 165–183.

Goldman, A. I. (1970). *A Theory of Human Action*. Princeton University Press.

Güver, L. and M. H. Kneer (2023). Causation, foreseeability, and norms. *Proceedings of the Annual Meeting of the Cognitive Science Society 45*, 888–895.

Hall, N. (2004). Two concepts of causation. In J. Collins, N. Hall, and L. Paul (Eds.), *Causation and Counterfactuals*, pp. 225–276. MIT Press.

Hart, H. L. A. and T. Honoré (1959). *Causation in the Law*. Oxford University Press UK.

Hawthorne, J. (2001). Freedom in context. *Philosophical Studies 104*(1), 63–79.

Hieronymi, P. (2020). *Freedom, Resentment, and the Metaphysics of Morals*. Princeton, NJ, USA: Princeton University Press.

Hitchcock, C. and J. Knobe (2009). Cause and norm. *Journal of Philosophy 106*(11), 587–612.

Jeppsson, S. (2016). Non-elusive freedom contextualism. *Philosophia 44*(3), 793–808.

Jeppsson, S. (2019). The agential perspective: A hard line reply to the four case manipulation argument. *Philosophical Studies 177*(7), 1935–1951.

Kane, R. (1985). *Free Will and Values*. State University of New York Press.

Kane, R. (1996a). Free will, responsibility, and will-setting. *Philosophical Topics 24*(2), 67–90.

Kane, R. (1996b). *The Significance of Free Will*. Oxford University Press USA.

Kane, R. (2007). Libertarianism. In J. M. Fischer (Ed.), *Four Views on Free Will*. Blackwell.

Kim, J. (1973). Causation, nomic subsumption, and the concept of event. *Journal of Philosophy 70*(8), 217–236.

Knobe, J. (2009). Folk judgments of causation. *Studies in History and Philosophy of Science Part A 40*(2), 238–242.

Knobe, J. and B. Fraser (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*. MIT Press.

Latham, A. J. and H. Tierney (2021). The four-case argument and the existential/universal effect. *Erkenntnis 88*(6), 2379–2389.

Latham, A. J. and H. Tierney (2022). Defusing existential and universal threats to compatibilism: A strawsonian dilemma for manipulation arguments. *Journal of Philosophy 119*(3), 144–161.

Latham, A. J., S. Varga, and H. Tierney (2023). Freedom, moral responsibility, and the failure of universal defeat. *Philosophical Issues 33*(1), 252–269.

Levy, N. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press UK.

Lewis, D. (1973). Causation. *Journal of Philosophy 70*(17), 556–567.

Lewis, D. (1981). Are we free to break the laws? *Theoria 47*(3), 113–21.

Lewis, D. (1986a). Events. In D. Lewis (Ed.), *Philosophical Papers Vol. II*, pp. 241–269. Oxford University Press.

Lewis, D. (1986b). *Philosophical Papers, Volume II*. Oxford University Press.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy 74*(4), 549–567.

Lewis, D. (2020). Outline of ?nihil obstat: An analysis of ability? *The Monist 103*(3), 241–244.

Liu, X. (2022). Manipulation and machine induction. *Mind 131*(522), 535–548.

Livengood, J., J. Sytsma, and D. Rose (2017). Following the fad: Folk attributions and theories of actual causation. *Review of Philosophy and Psychology 8*(2), 273–294.

Macnamara, C. (2013). Reactive attitudes as communicative entities. *Philosophy and Phenomenological Research 90*(3), 546–569.

Marco, G. D. (2016). Rescuing the zygote argument. *Philosophical Studies 173*(6), 1621–1628.

Marco, G. D. (2022). Manipulation, machine induction, and bypassing. *Philosophical Studies 180*(2), 487–507.

Mason, E. (2019). *Ways to Be Blameworthy: Rightness, Wrongness, and Responsibility*. Oxford University Press.

Matheson, B. (2014). Compatibilism and personal identity. *Philosophical Studies 170*(2), 317–334.

Matheson, B. (2016). In defence of the four case argument. *Philosophical Studies 173*(7), 1963–1982.

McDonnell, N. (2018). Making a contribution and making a difference. *American Philosophical Quarterly 55*(3), 303–312.

McKay, T. J. and D. Johnson (1996). A reconsideration of an argument against compatibilism. *Philosophical Topics 24*(2), 113–122.

McKenna, M. (2001). Source incompatibilism, ultimacy, and the transfer of non-responsibility. *American Philosophical Quarterly 38*(1), 37–51.

McKenna, M. (2008a). A hard-line reply to pereboom's four-case manipulation argument. *Philosophy and Phenomenological Research 77*(1), 142–159.

McKenna, M. (2008b). Saying good-bye to the direct argument the right way. *Philosophical Review 117*(3), 349–383.

McKenna, M. (2010). Whose argumentative burden, which incompatibilist arguments?—getting the dialectic right. *Australasian Journal of Philosophy 88*(3), 429–443.

McKenna, M. (2012). *Conversation & Responsibility*. Oup Usa.

McKenna, M. (2014). Resisting the manipulation argument: A hard-liner takes it on the chin. *Philosophy and Phenomenological Research 89*(2), 467–484.

McKenna, M. (2016). A modest historical theory of moral responsibility. *The Journal of Ethics 20*(1-3), 83–105.

Mele, A. R. (1995). *Autonomous Agents: From Self Control to Autonomy*. New York, US: Oxford University Press.

Mele, A. R. (1998). Review of robert kane's *the Significance of Free Will*. *Journal of Philosophy 95*(11), 581–584.

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

Mele, A. R. (2008). Manipulation, compatibilism, and moral responsibility. *The Journal of Ethics 12*(3-4), 263–286.

Mele, A. R. (2015). Libertarianism, compatibilism, and luck. *The Journal of Ethics 19*(1), 1–21.

Mele, A. R. (2019). *Manipulated Agents: A Window to Moral Responsibility*. New York, NY: Oup Usa.

Mele, A. R. (2021). Direct versus indirect: Control, moral responsibility, and free action. *Philosophy and Phenomenological Research 102*(3), 559–573.

Menzies, P. (2007). Causation in context. In H. Price and R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.

Menzies, P. (2009). Platitudes and counterexamples. In H. Beebee, P. Menzies, and C. Hitchcock (Eds.), *The Oxford Handbook of Causation*, pp. 341–367. Oxford University Press.

Menzies, P. (2017). The consequence argument: Disarmed an interventionist perspective. In H. Beebee, C. Hitchcock, and H. Price (Eds.), *Making a Difference: Essays on the Philosophy of Causation*, pp. 307 – 330. Oxford: Oxford University Press.

Merlussi, P. (2016). Determinism, laws of nature and the consequence argument. *Manuscrito 39*(1), 73–95.

Merlussi, P. (2022). Revisiting mckay and johnson's counterexample to ( *?*). *Philosophical Explorations 25*(2), 189–203.

Mickelson, K. (2010). The soft-line solution to pereboom's four-case argument. *Australasian Journal of Philosophy 88*(4), 595–617.

Mickelson, K. (2015a). A critique of vihvelin's three-fold classification. *Canadian Journal of Philosophy 45*(1), 85–99.

Mickelson, K. (2015b). The zygote argument is invalid: Now what? *Philosophical Studies 172*(11), 2911–2929.

Mickelson, K. (2016). The manipulation argument. In K. Timpe, M. Griffith, and N. Levy (Eds.), *Routledge Companion to Free Will.*, pp. 166–178. Routledge.

Mickelson, K. M. (2021). The zygote argument is still invalid: So what? *Philosophia 49*(2), 705–722.

O'Keefe, T. (2016). The stoics on fate and freedom. In M. Griffith, N. Levy, and K. Timpe (Eds.), *The Routledge Companion to Free Will*, pp. 236–246. Routledge.

Paul, L. A. (2000). Aspect causation. *Journal of Philosophy 97*(4), 235.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.

Pereboom, D. (2001). *Living Without Free Will*. Cambridge University Press.

Pereboom, D. (2007). Hard incompatibilism. In J. M. Fischer (Ed.), *Four Views on Free Will*. Blackwell.

Pereboom, D. (2008). A hard-line reply to the multiple-case manipulation argument. *Philosophy and Phenomenological Research 77*(1), 160–170.

Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford University Press.

Phillips, J., J. B. Luguri, and J. Knobe (2015). Unifying morality's influence on non moral judgments: The relevance of alternative possibilities. *Cognition 145*(C), 30–42.

Plantinga, A. (1974). *The Nature of Necessity*. Oxford, England: Clarendon Press.

Putnam, H. (1982). Why there isn't a ready-made world. *Synthese 51*(2), 205–228.

Rieber, S. (2006). Free will and contextualism. *Philosophical Studies 129*(2), 223–252.

Rose, D. (2017). Folk intuitions of actual causation: A two pronged debunking explanation. *Philosophical Studies 174*(5), 1323–1361.

Samland, J. and M. R. Waldmann (2016). How prescriptive norms influence causal inferences. *Cognition 156*(C), 164–176.

Sartorio, C. (2013). Making a difference in a deterministic world. *Philosophical Review 122*(2), 189–214.

Sartorio, C. (2016). *Causation and Free Will*. Oxford University Press UK.

Sartorio, C. (2017). Frankfurt-style examples. In G. Meghan, T. Kevin, and L. Neil (Eds.), *Routledge Companion to Free Will.*, pp. 179–190. Routledge.

Sauer, H. (2014). It's the knobe effect, stupid!: How to explain the side effect effect. *Review of Philosophy and Psychology 5*(4), 485–503.

Sauer, H. and T. Bates (2013). Chairmen, cocaine, and car crashes: The knobe effect as an attribution error. *The Journal of Ethics 17*(4), 305–330.

Schaffer, J. (2005). Contrastive causation. *Philosophical Review 114*(3), 327–358.

Schaffer, J. (2012). Causal contextualisms. In M. Blaauw (Ed.), *Contrastivism in Philosophy: New Perspectives*. Routledge.

Sekatskaya, M. (2019). Double defence against multiple case manipulation arguments. *Philosophia 47*(4), 1283–1295.

Seybold, S. L. (forthcoming). Why causal facts matter: A critique of jeppsson?s hard-line reply to four-case manipulation arguments. *Inquiry: An Interdisciplinary Journal of Philosophy*.

Shoemaker, D. (2015). *Responsibility From the Margins*. Oxford University Press.

Steward, H. (2021). What is determinism? why we should ditch the entailment definition. In M. Hausmann and J. Noller (Eds.), *Free Will: Historical and Analytic Perspectives*, pp. 17–43. Springer Verlag.

Strawson, G. (1994). The impossibility of moral responsibility. *Philosophical Studies 75*(1-2), 5–24.

Strawson, P. F. (1974). *Freedom and Resentment and Other Essays*. London, England: Routledge.

Strawson, P. F. (1985). Causation and explanation. In B. Vermazen and M. B. Hintikka (Eds.), *Essays on Davidson: Actions and Events*, pp. 115–35. Oxford University Press.

Sytsma, J. (2020). Causation, responsibility, and typicality. *Review of Philosophy and Psychology 12*(4), 699–719.

Sytsma, J. (2022). Crossed wires: Blaming artifacts for bad outcomes. *Journal of Philosophy 119*(9), 489–516.

Sytsma, J., R. Bluhm, P. Willemsen, K. Reuter, E. Fischer, and M. D. Curtis (2019). Causation attributions and corpus analysis. In *Advances in Experimental Philosophy*, pp. 209–238.

Sytsma, J. and J. Livengood (2016). Intervention, bias, responsibility. . . and the trolley problem.

Sytsma, J., J. Livengood, and D. Rose (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 43*(4), 814–820.

Taylor, R. (1974). *Metaphysics (4th edition)*. Englewood Cliffs, N.J.,: Prentice-Hall.

Timpe, K. (2016). Leeway vs. sourcehood conceptions of free will. In K. Timpe, M. Griffith, and N. Levy (Eds.), *Routledge Companion to Free Will.*, pp. 213–224. Routledge.

van Inwagen, P. (1983). *An Essay on Free Will*. Oxford University Press.

Vihvelin, K. (2008). Compatibilism, incompatibilism, and impossibilism. In T. Sider, J. Hawthorne, and D. W. Zimmerman (Eds.), *Contemporary Debates in Metaphysics*, pp. 303–318. Blackwell.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oup Usa.

Vihvelin, K. (2022). Arguments for Incompatibilism. In E. N. Zalta and U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2022 ed.). Metaphysics Research Lab, Stanford University.

Wasserman, R. (Manuscrito). Is causation extensional?

Watson, G. (1987). 4. responsibility and the limits of evil: Variations on a strawsonian theme. In J. M. Fischer and M. Ravizza (Eds.), *Perspectives on Moral Responsibility*, pp. 119–148. Cornell University Press.

Whittle, A. (2021). *Freedom and Responsibility in Context*. Oxford, United Kingdom: Oxford University Press.

Widerker, D. (1995). Libertarianism and frankfurt's attack on the principle of alternative possibilities. *Philosophical Review 104*(2), 247–61.

Willaschek, M. (2010). Non-relativist contextualism about free will. *European Journal of Philosophy 18*(4), 567–587.

Willemsen, P. H. and L. Kirfel (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass 14*(1), e12562.

Williams, M. (2001). Contextualism, externalism and epistemic standards. *Philosophical Studies 103*(1), 1–23.

Williams, M. (2007). Why (wittgensteinian) contextualism is not relativism. *Episteme 4*(1), 93–114.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.