

Mind-Body Parallelism and Spinoza's Philosophy of Mind

by
Ruben Noorloos

Submitted to
Central European University
Department of Philosophy

In partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Supervisor:
Michael Griffin

Vienna, Austria
2022

Copyright notice

This dissertation contains no materials accepted for any other degrees in any other institutions. It does not contain any material previously written and/or published by another person, except where appropriate acknowledgment is made in the form of a bibliographical reference.

Ruben Noorloos

Vienna, July 27, 2022

Abstract

Mind-body parallelism is the view that mind and body stand in the same “order and connection,” as Spinoza put it, or that corresponding mental and physical states have corresponding causal explanations in terms of other mental and physical states. This dissertation investigates the nature and role of mind-body parallelism, as well as other forms of parallelism, in Spinoza’s philosophy of mind. In doing so, it also considers how Spinoza’s views relate to current discussions. In present-day philosophy of mind, mind-body parallelism is almost never defended. It is seen as a historical dead-end with insurmountable problems. By contrast, I argue that parallelism powerfully responds to the post-Cartesian mind-body problem (which remains with us today) and that it points a way forward in current debates. The dissertation contains five independent chapters. After an introduction that situates parallelism in relation to both Spinoza’s time and to present discussions, Chapter 1 presents an argument for parallelism aimed at a present-day audience. Chapter 2 discusses Spinoza’s own arguments for parallelism. Both chapters help to clarify what parallelism is, in part by distinguishing between several versions of the view. Chapter 3 discusses what is often considered parallelism’s most problematic feature, its rejection of mind-body interaction. I argue that by distinguishing between the post-Cartesian context in which Spinoza wrote and present-day discussions, we can see that parallelism is compatible with mental causation. Chapters 4 and 5, finally, discuss specific ways in which parallelism is at work in Spinoza’s view of the mind. In Chapter 4, I argue that parallelism is at work in Spinoza’s interesting and distinctive positions on the nature of agency and motivation. In Chapter 5, I show the role of parallelism in his representationalist theory of consciousness. A guiding thread throughout the dissertation is that parallelism presents a distinctive and interesting way to combine realism, non-reductionism and naturalism in relation to those features of human self-understanding that seem difficult to fit into a naturalistic worldview.

Table of contents

<i>Introduction</i>	1
<i>1. An Argument for Mind-Body Parallelism</i>	9
1. Introduction	9
2. What is mind-body parallelism?	10
3. An argument for mind-body parallelism	13
4. Three objections	20
4.1 Why anti-interactionism?	20
4.2 What explains the parallelism?	23
4.3 Is this panpsychism?	25
5. Conclusion	27
<i>2. Spinoza's Arguments for Mind-Body Parallelism</i>	28
1. Introduction	28
2. E2p7e: The general form of a parallelism	30
3. E2p7e: Representational parallelism	34
4. E2p7dem: The role of the causal axiom	36
5. E2p7c: God's powers and three identities	41
6. E2p7s: Metaphysical parallelism	45
7. E2p13: Mind-body parallelism	48
8. Conclusion	50
<i>3. Spinoza's Rejection of Interactionism</i>	52
1. Why did Spinoza reject interactionism?	52
2. Five responses	54
2.1 Accommodating apparent interaction	55
2.2 Dualist responses	56
2.3 Causal contexts as intensional	57
2.4 Distinguishing causation and causal explanation	60

TABLE OF CONTENTS

2.5 Causal pluralism	64
3. The minimal exclusion problem and Spinoza	66
4. Spinoza against interactionism	70
5. Two potential sources of mental causation	74
5.1 Appetites as psychophysical causes	74
5.2 Intentional action	75
6. Conclusion	78
4. <i>Spinoza's Deterministic Theory of Motivation</i>	79
1. Introduction	79
2. Intrinsic power and causation by essences	82
3. Answer to the source problem	88
4. The motivation problem	91
5. The distributive model as a vector model of willing	95
6. A parallelist theory of motivation	97
7. Conclusion	99
5. <i>Spinoza's Representationalist Theory of Consciousness</i>	101
1. Introduction	101
2. Does Spinoza have a theory of sensible quality?	103
3. The modern explanatory gap	106
4. The conceivability gap in Descartes	108
5. Intrinsic connection without conceptual connection: Spinoza's conceptual barrier	111
6. Phenomenal consciousness as what it is like	114
7. A sketch of Spinoza's theory of affection	119
8. Spinoza as a representationalist: Content and confusion	124
9. Can the phenomenal quality of an idea be explained?	130
10. Conclusion: Spinoza's pregnant silence on the sensible qualities	131
<i>Conclusion</i>	133
<i>Acknowledgements</i>	137
<i>Abbreviations and Conventions</i>	139
<i>Bibliography</i>	141

Introduction

Mind-body parallelism is the view that mental and physical states stand in “the same order and connection,” as Spinoza put it (E2p7).¹ As will be explained below, this means that corresponding mental and physical states have corresponding causal explanations: any mental state can be explained in terms of mental causes, and any physical state in terms of physical causes, and these causes correspond. So a complete causal story is available (at least in principle) to explain how any mental state was caused that appeals only to mental causes. This is not to say that the *only* or the *best* explanation will be given entirely in mental causes. Rather, parallelism – unless otherwise indicated, ‘parallelism’ will be short for ‘mind-body parallelism’ in this introduction – should be seen as insisting on there being two parallel causal stories to be told, such that when (if ever) the sciences of the mind and the physical world are completed, it will be possible to freely switch between both. Some of this explanatory promise is realized in Spinoza’s psychophysical theory of the mind and the affects, which allows for a way of connecting someone’s behavior and the way it is influenced by external bodies to their emotions and self-understanding.²

This dissertation is, to my knowledge, the first monograph-length study of mind-body parallelism. It also contains one of the very few defenses of parallelism in contemporary philosophy.³

¹ See the key to abbreviations for details on how I refer to Spinoza’s and other works. Spinoza’s writings are cited from G, in Curley’s translation (Spinoza 1985-2016).

² Spinoza also, of course, has his ‘conceptual barrier’ between the attributes (E1p10), which seemingly prevents there from being genuine psychophysical explanations. I think the conceptual barrier plays a less important role in Spinoza’s psychology than is often thought; see Chapter 3. However, the point I am making here does not really conflict with the conceptual barrier, because the latter can be taken to require only that the *fundamental* causal explanation of any mental state is in mental terms, not that the *only* or *best* explanation is. As pointed out in Chapter 3, Spinoza himself occasionally appeals to psychophysical explanations.

³ Skrbina (2014) defends a form of dual-aspect theory that he calls “hylonoetic parallelism.” Zhong (2017) advocates a theory of mental causation that he calls parallelist (see also Zhong 2014).

It consists of five chapters that each deal with an aspect of parallelism: what it is, the arguments for it, objections against it, and possible applications. Most of these chapters take a historical approach and focus on Spinoza's form(s) of parallelism. However, the first chapter is systematic in its approach, and even many of the historical chapters approach parallelism with systematic concerns in mind. In this introduction, I will first contextualize the chapters by providing some motivations for being interested in parallelism; then I will introduce the chapters; finally I will briefly address the dissertation's combination of historical and systematic approaches.

What motivates parallelism? What, in particular, could motivate you to pursue parallelist solutions to various of the disputed problem cases in the philosophy of mind? This is hardly evident from the description of parallelism just given, because the appeal to parallel causal explanations seems only loosely – if at all – connected to the questions at stake in the metaphysics of mind. These questions do not just address the issue of mind-body interaction – on which parallelism does have a stance, albeit a strong one – but, perhaps even more importantly, they are also questions about ontological status and ontological dependence: is the physical more fundamental than the mental or not; if it is, how does the mental depend on the physical?⁴ Clearly, to say that mind and body are causally parallel does not entail any particular stance on that latter question. But this makes parallelism appear ill-suited to address issues in the metaphysics of mind. It also makes it appear unmotivated, because a metaphysical theory like parallelism presumably requires a metaphysical justification that addresses these issues. As it is sometimes put (see Chapter 1, section 4.2), if there is a parallelism, it requires an explanation.

I'm not going to give that explanation in this dissertation. Yet I still think parallelism has enormous potential in relation to the mind-body problem. I think this because I see parallelism as providing an interesting *structural* response to many issues in the philosophy of mind. This response is compatible with different ways of understanding the ontological status of mind and body. So, conceivably, dualists, physicalists, neutral monists and others could all make use of it. As these views all have their own ways of 'explaining' the parallelism, I take that task to be

⁴ For a brief characterization of the mind-body problem as revolving around both questions about interaction and about ontological dependence, see Crane and Patterson (2000).

irrelevant to my purposes, as I am arguing for and exploring parallelism and not any of these metaphysical theories.

It is, however, incumbent on me to say something about this structure. My characterization of it here is familiar and underlies the argument for parallelism that I set out in Chapter 1. This story starts from a bird's-eye overview of the discussions over the contemporary mind-body problem. Many of these debates show a rough shared structure, which can be described as a stand-off between *reductionists* and *dualists*. (These names are used here to name tendencies, not precise positions.) On the one hand, the reductionists claim to be able to explain some contested mental phenomenon in physical terms. However, these explanations appear to lose something that is essential to what it is they are explaining – for example, they are unable to account for the phenomenal character of experience, or for its representational content. This is pointed out by the dualists, who in turn, however, are at risk of removing the contested mental phenomena from the physical causal order, thereby assigning them super-natural powers to interfere in this order. The debate remains stuck in a dilemma: either conform the mental to the physical by lopping off something that seems essential to it, or accept the essential feature at the cost of seemingly removing it from the natural world.

This not-novel diagnosis of the debate suggests a possible way forward: what is needed is a theory that is *non-reductive* (does not accept that all mental phenomena have complete physical explanations) yet *naturalistic* (does not conflict with the causal closure of the physical). Mind-body parallelism is a good candidate for this role. Both features are built into the theory: it is non-reductive because it requires that every mental state has a mental causal explanation that does not reduce to a physical explanation, and naturalistic because it takes the mental to be completely causally parallel to the physical, without a power to interfere in it. The view therefore responds in an interesting way to these recent discussions. As I said, Chapter 1 argues for parallelism along these lines.

As a distinctive and unfamiliar form of nonreductive naturalism, parallelism would be interesting enough. However, I believe its appeal goes deeper than this. Parallelism can also be used to question the assumptions behind current approaches to the mind-body problem. Here, the point that parallelism does not come with its own metaphysics becomes relevant again. What the brief sketch just given suggests is that the solution to these problems will not come from an account of the metaphysical status of mind and body themselves. It is not clear that the choice between dualism and monism (of whatever kind) is going to have a decisive impact on

the ability to understand how humans can be fully subject to physical law yet possess some of the features that are central to human self-understanding. If that is right, then determining whether the mind is physical or not is probably not sufficient by itself to answer these problems.

Instead, the solution appears to lie at the level of concepts and explanation. What generates the problem is a sense, first, that a complete account of physical reality must be capable of being given in scientific terms. Second, it is the seeming fact that some features of human self-understanding elude this kind of scientific treatment; even though, third, humans are to a significant extent physical beings. Because the scientific story is supposed to be complete, purported facts about human self-understanding cannot influence it to the point where it needs to be modified. At the same time, an account of human self-understanding that takes no notice of the scientific facts remains incomplete, and accounts that combine both exhibit persistent gaps and discontinuities.

Parallelism here exemplifies an original and little-followed tendency, different from the dualist and reductionist tendencies: it takes the disputed features of human self-understanding and attempts to *construe* them *in parallel* to the physical features that are supposed to correspond to them. In doing this, it intentionally avoids trying to build intelligible connections across explanatory gaps or to understand how human agency, as manifested in mental causation, intersects with physical causality. Instead, it seeks a shared “order and connection” between mental and physical features, and then tries to find corresponding causal explanations of them. Insofar as this is possible without antecedently settling the ontological status of the parallel items, it is motivated by issues about the *explanation* of mental and physical states and not their ontological status.

Chapters 4 and 5 present two examples of construing in parallel (or, perhaps, one example, applied to two cases). Chapter 4 shows how Spinoza can situate human agency in a naturalistic worldview by ascribing causal powers to things’ essences and then construing the will in parallel to its affections by external causes. This results in a plausible picture of agency that neither requires an account of how a purely mental will intervenes in a purely physical body nor reduces agency to a purely physical phenomenon. Chapter 5 applies a similar model to the relation between the phenomenal quality of experiences and their representational content. It also contains more discussion of how Spinoza avoids establishing conceptual connections between mind and body.

Admittedly, the dissertation only begins to work out the issues raised in the previous few paragraphs. I do not present a general defense of this parallelist tendency, apart from its applications in the final two chapters and the general remarks just given. However, if I am right, there are at least three reasons why parallelism is interesting: it is interesting for its eschewal of attempts to find conceptual connections between mind and body; for the way it approaches the mind-body problem without first assuming a view about the ontological status of mind and body; and because it is an original and underexplored form of nonreductive naturalism.

As for the structure of the dissertation itself, it takes the shape of five independent papers that all either look into what parallelism is and what arguments there are for the view, deal with an objection to parallelism, or explore its application to specific questions. The choice to make the chapters independent had several reasons. The main one is that the dissertation aims to rehabilitate an underexplored and unfamiliar theory. There is no readily available framework from which I could depart (or insofar as it does exist, it centers on a version of the view that in my eyes does not speak to its strengths). My goal therefore was to develop this framework; but the best way to do that, and also the best way to show parallelism's promise, was by first working out concretely a parallelist approach to certain issues. This broadly bottom-up approach allowed me to work backwards from these parallelist theories to a more comprehensive characterization of the view. As this introduction perhaps suggests, however, this framework is still in development. Making the chapters independent has the advantage of foregrounding the specific contributions of these chapters over any overarching story that is supposed to connect them. In the end, the chapters interconnect and support each other to a greater extent than I initially anticipated; but it is also true that there are various shifts of emphasis between them (on which more in a minute).

Here is a brief overview of the chapters. Chapter 1 presents a defense of parallelism aimed at present-day philosophers of mind. After clarifying the view, it develops an argument for it based on widely accepted premises. It also responds to the most common objections to the view, showing that they are not conclusive. It makes a case for parallelism as a view that should be interesting to anyone who is moved by concerns over the explanatory gap but wants to maintain a strong commitment to the explainability of mental states.

Chapter 2 also considers arguments for parallelism, but this time they are Spinoza's own arguments. In a step-by-step reconstruction of E2p7, it argues that Spinoza's arguments are valid

even if they are often highly condensed. It also defends Deleuze's and Melamed's view that Spinoza has two distinct ways of arguing for parallelism, one based on considerations about representation and one based in his metaphysics. In doing so the chapter also presents a somewhat broader view of Spinoza's parallelism, one that is able to answer recent skepticism about the accurateness of this term as a description of Spinoza's views.

While the first two chapters deal with what parallelism is and how it can be argued for, Chapter 3 considers what is probably the major objection against the view, which is that its rejection of mental causation is unacceptable. Part of my response to this objection is already given in the first chapter, where I argue that parallelism does not rule out all mental causation. In Chapter 3, I discuss Spinoza's own rejection of interactionism. I stress the distinction between the form of mental causation – 'interactionism' – that parallelism *does* reject and the other forms of mental causation that it does not. I also show that Spinoza is aware of the distinction. The aim of the chapter is to account for this apparent tension in Spinoza's view.

Chapters 4 and 5 concern applications of parallelism to specific issues. First, Chapter 4 considers agency and motivation. It sets out Spinoza's distinctive theory of freedom and what I call his 'distributive' theory of the will. I argue that these views allow Spinoza to accommodate a core assumption behind libertarian views of free will within a strictly deterministic framework. I also show how Spinoza's overarching theory of agency exhibits a form of parallelism.

Finally, Chapter 5 discusses Spinoza's views on phenomenal consciousness. Despite having received a significant amount of attention, these views remain controversial, to the point where it is questioned if Spinoza has a theory at all. Paying attention in particular to the sensible qualities, I explain why Spinoza seemingly has so little to say about phenomenal quality: this is because parallelism prevents the purely qualitative aspect of experience from playing a significant role in the explanation of behavior, except insofar as it is 'affective' (i.e., pleasant or unpleasant). Moreover, Spinoza does not think that the purely qualitative can be rationally explained. Parallelism, as I said above, here works to accommodate phenomenal qualities in his theory of the mind without requiring the explanatory gap between these qualities and the representations that determine them to be closed.

The choice to make the chapters independently readable has had the inevitable result that there is some overlap between chapters. This is especially true for the first chapter, which borrows and repeats arguments set out in more detail in other chapters. And each chapter briefly

introduces the main features of parallelism and Spinoza's philosophy of mind and metaphysics that are relevant to that chapter.

More importantly, the independence of the chapters also results in some slight changes in emphasis, in different formulations of parallelism, and in unexplored connections between and across chapters. Most notably, the form of parallelism at issue in Chapters 4 and 5 is an *intra-mental* parallelism, a parallelism between the orders of the affections and the intellect. The other chapters discuss mind-body parallelism instead. To resolve these discrepancies, the conclusion of the dissertation presents an overview of the various forms of parallelism that have been discussed. It also returns to parallelism as a method and shows some ways in which the chapters have exemplified this method.

A final word on the dissertation's method more broadly. In the history of philosophy, a debate is currently going on over the correct way to interpret historical authors. At stake is the role of historical context in interpreting past authors: should commentators try first of all to understand philosophers in a way that is sensitive to the philosophical issues, concepts and distinctions that were relevant in their time, or, on the contrary, is the main role of such interpretation to make historical insights relevant to contemporary concerns, even if that means ignoring historical context or, stronger, reconstructing a philosopher's views in ways they might not have recognized? As the field of (Anglophone) history of philosophy has grown in size and become more specialized over the last few decades, there has been a shift away from the latter, 'rational reconstructionist' method towards the former, 'contextualist' one. Spinoza has become a focal point of this debate, either as someone whose ideas are said to have been fit into unhistorical molds in order to make them acceptable to modern audiences, or whose ideas remain so strange yet compelling that engaging with them can be a good way to overcome historical biases.⁵

In the dissertation, I focus on reconstructing Spinoza's arguments for various views and spend relatively little time on historical context. I also make a serious effort to interpret Spinoza's views in terms that are understandable to an audience of present-day philosophers, in particular by using terms and distinctions from the current discussions over mental causation, free will and consciousness. However, this is not because I am committed to a rational reconstructionist

⁵ Melamed (2013a) argues against the former way of reading Spinoza and defends the latter. See also the debate between Daniel Garber and Michael Della Rocca (Garber 2015a,b; Della Rocca 2015), as well as Lærke, Smith, and Schliesser (2013) (which contains Melamed's essay) and Mercer (2019).

method and reject contextualism. Instead, as I hope this introduction has made clear, my choice to use modern terminology and concepts is informed by my overarching systematic aim of introducing parallelism as a theory that has not just historical but also contemporary interest. Moreover, even when I try to make Spinoza intelligible to present-day readers, I do intend my concepts and distinctions to be answerable to Spinoza's texts; so to that extent the dissertation remains contextualist. On the wider issue of reconstructionism versus contextualism I am officially agnostic, although it seems to me that the two methods do not exclude each other and that the best work in the history of philosophy combines both.

*An Argument for Mind-Body Parallelism**1. Introduction*

Mind-body parallelism is the view that mind and body are causally and explanatorily parallel to each other – in other words, that mental and physical phenomena have corresponding causal explanations. To use Spinoza’s famous expression, mind and body “follow one and the same order and connection”.⁶ Mental and physical phenomena stand in causal chains that correspond point for point. (This may be because the mental and physical phenomena are identical.) Mind-body parallelism – or parallelism, for short⁷ – therefore differs from interactionist forms of dualism in denying that the fundamental mind-body relation is a causal one. And it differs from standard forms of physicalism and idealism in denying that the mind must be understood from the way it is grounded in matter (or vice versa). A parallelist conception of the mind-body problem focuses attention on the nature of the order that is shared between the mental and the physical, and on the need to coordinate mental and physical explanations of corresponding mental and physical processes. Apart from elaborating these causal explanations – seeking the shared “order and connection” – it is often tasked with explaining why this order and connection is there.

Besides Spinoza, parallelism is associated with Leibniz and a group of nineteenth-century authors who, following Gustav Fechner’s success in developing the first rigorous psychophysical laws, called themselves “psychophysical parallelists.”⁸ At present, however, the view is out of

⁶ This paraphrases E2p7s.

⁷ Unless otherwise indicated, all uses of “parallelism” in this chapter refer to mind-body parallelism.

⁸ Besides Fechner himself, the group included many of the more famous writers on psychology in the second half of the 19th century, including Clifford (1879), Mach (1897), and (at one point) Wundt (1894). See

fashion and tends to be seen as a historical curiosity at best. It is thought to be abstruse, speculative, and to suffer from several devastating objections. But this widespread attitude is mistaken. Parallelism is a powerful, subtle and interesting view. It promises a way forward in debates on the mind-body problem that have on the one hand been powered – rightly – by high explanatory ambitions, but that have also run into various conceptual problems, with the “explanatory gap” (Levine 1983) being perhaps the most important one.

This chapter clarifies what parallelism is, presents a new argument in its favor and defends it from some of the most common objections. In a nutshell, the argument I will present is this. The mind is a part of nature. It is tightly connected to the body. But we cannot explain it just in physical terms. Still, because we have no reason to think that it cannot be explained, our best option is to think of the mind as parallel to the body. That is, we should become mind-body parallelists.

The chapter’s structure is as follows. In the next section, I elaborate on the nature of parallelism. Then, in section 3, I give a more detailed version of the above argument. Section 4 responds to three main objections against the view. Section 5 briefly concludes.

2. *What is mind-body parallelism?*

Before we can meaningfully talk about why one should be a mind-body parallelist or not, we have to get clear on the meaning of parallelism itself. So what do I mean by this term?

Not every parallelism is a mind-body parallelism. Things can be parallel even if they are not minds or bodies. Additionally, there are different *ways* or *respects* in which things can be parallel. Depending on how we specify these things, what it means to be parallel, and hence what it means to be a parallelist, changes. This section will explain these claims.

I claim that two (sets of) things are parallel if and only if they correspond somehow and they stand in the same relations. Naturally, to be a parallelist about two (sets of) things in respect of some relation is to hold that they are parallel with respect to that relation. In terms of a formula, the relation ‘being parallel to’ can be stated in terms of two conditions:

Wegener (2009) for a general survey of the movement. For Fechner, see Heidelberger (2004, esp. ch. 5), and Pauen (2011, 94-8). For two of Fechner’s own statements of psychophysical parallelism, see Fechner (1851, 2:373-86; 1860, 1:1-20).

a 's are parallel to b 's iff there is a relation R such that, for any a_n and b_n , $Ra_n a_{n+1}$ iff $Rb_n b_{n+1}$.

In this formula, the a 's are members of one of the two sets and the b 's of the other. Furthermore, I use a_n to indicate the a -item that corresponds to b -item b_n .

While elaborate, the formula is also noncommittal. It does not specify the nature of the a 's and b 's, apart from suggesting that the a 's on the one hand and the b 's on the other belong to a shared category (for example, a 's are mental phenomena, and b 's are physical phenomena). In fact, the parallel items may all belong to the *same* metaphysical category. This is true in the case of Leibniz's famous example of the two parallel clocks. Moreover, while Leibniz's clocks are numerically distinct, this too is not required by the formula. The parallel items may turn out to be "one and the same thing," as is the case for mind and body in Spinoza (E2p7s).

The formula also does not specify the nature of the correspondence between the parallel items. If the parallel items really are "one and the same thing," then the correspondence is based in identity. Note, however, that the formula also allows for specifying the relation in terms of some relation of ontological dependence (such as supervenience). That is, I do not think that it is part of what it means to be parallel that the parallel items are equally ontologically fundamental. This will be important in what follows. Finally, the correspondence could be specified in yet other ways. For example, Spinoza holds a doctrine of *representational* parallelism between ideas and the things they represent. In this case, the correspondence is determined by the fact that the ideas represent the things they are parallel to.⁹

Third and finally, the formula is neutral as to what the relation R is with respect to which the things are parallel. The most interesting candidates for this relation are *causation*, *conceptual dependence* and *explanation*. Leibniz's clocks stand in a relation of *causal* parallelism: their successive states correspond causally. Conceptual dependence can be understood as a relation of a priori entailment. This relation was important to Spinoza and Leibniz, who in different ways took causal relations to correspond to such relations.

Perhaps the most interesting form of parallelism, however, is a parallelism of explanatory relations. By an explanatory relation I mean simply the relation that connects an explanandum to an explanans. It connects, say, the shattering of a window to the window's being hit by a ball, explaining *why* the window shattered. I will often speak of *things* explaining each other, instead

⁹ For discussion of representational parallelism and its relation to mind-body parallelism, see Deleuze (1990, ch. 7), Melamed (2013, chs. 5-6), and Chapter 2.

of the more common practice of restricting the relata of explanations to propositions, facts or sentences. So I will say, for example, that one mental event m_1 explains the occurrence of some other mental event m_2 . If you think that explanatory relations hold only between propositions or linguistic entities, then you can translate such statements into their propositional form – e.g., “*that* m_1 occurred explains *that* m_2 occurred.” Additionally, an explanation of m_2 in terms of m_1 will presumably presuppose the occurrence of various background conditions, laws of nature, etc. That laws might be involved in explanations raises an interesting question for an explanatory parallelism: given that this kind of parallelism may pair explanations between radically different kinds of things, what relations do the laws involved in one kind of explanation have to the laws involved in the other? Absent a satisfactory explanation, the correlation of the laws will seem arbitrary. This problem will need to be addressed later on (see section 4.2).

Depending on the correspondence involved and the relation relative to which things are parallel are explained, then, we will end up with different kinds of parallelism. What kind of parallelism is mind-body parallelism? Clearly, the items involved in it are mental and physical phenomena – typically, mental states and states of the body or brain. (I will remain neutral on whether the phenomena in question are events, properties, tropes, or something else; my use of the term ‘states’ is also intended to be neutral between these terms.) It is less clear what the relation is with respect to which these mental and physical states are parallel. It is common to describe mind-body parallelism in causal terms, as the view that mental and physical states stand in corresponding causal relations.¹⁰ Evidently, this is a core part of the notion of a mind-body parallelism. However, one contention of this chapter is that there is an important sense in which mind-body parallelism is also an *explanatory* parallelism. The view is to a large extent motivated by concerns over what it means to give a satisfactory naturalistic explanation of the mind. This will become clear in a moment, when I spell out my argument for mind-body parallelism.

The upshot, then, is that mind-body parallelism is a *causal-explanatory* parallelism between mental and physical phenomena. It is the view that mind and body have corresponding causal explanations. It can be defined as follows (where m_n is the mental state that corresponds to physical state p_n):

¹⁰ Two examples from introductory texts are Heil (2020, 31-3) and Robinson (2020, §3.3).

Causal-explanatory mind-body parallelism. For any m_n and p_n , m_n causes and explains m_{n+1} iff p_n causes and explains p_{n+1} .

Before continuing, it is worth stressing once again that parallelism, so defined, is neutral on the ontological status of the mental and physical phenomena it relates. Parallelism asserts there to be a *structure* in one's ontology and epistemology, so to speak, without determining the nature of the elements of this structure (apart from their being 'mental' and 'physical' in some sense) and without settling the question why this structure is there. Parallelism is compatible with both a dualistic and a monistic metaphysics. This is both a strength and a weakness of the view. It is a weakness because the view needs to be supplemented by a more fleshed-out metaphysics before it can be said to fully address important issues – the mind-body problem first among them. To say that mental and physical phenomena are parallel does not in itself answer the question *why* they are parallel, and parallelism itself cannot, it seems to me, answer this question. (This also means, however, that faulting it for not answering this question is misguided, because it asks something of the view that it was not designed to do.) At the same time, its lack of an inherent metaphysics also makes parallelism adaptable. It means that parallelism can be used in many different contexts, even when these contexts involve different ontologies. I'll return to the questions raised in this paragraph in section 4.2.

3. *An argument for mind-body parallelism*

Parallelism is motivated by the thought that mental phenomena must be explainable, but cannot be given a complete physical explanation. It is a form of nonreductive naturalism (Perler 2018) that stresses the tight connection between mind and body and rejects the idea that the mind could be super- or non-natural, but that nevertheless wants to avoid committing to strong psychophysical dependence relations. This section sets out an argument for mind-body parallelism. I would like this argument to do a few different things. In presenting it, I'm guided by Thomas Nagel's classic argument for another often-dismissed view, panpsychism (Nagel 1979; see also Coleman 2018; see below for the relation between parallelism and panpsychism). Like Nagel did for panpsychism, I hope to show first of all how parallelism follows from a number of premises that, while strong, are each defensible and seem more plausible than their negations. Additionally, the argument presented will elucidate the logic and motivations behind parallelism. And thirdly, I will use the exposition of the argument to draw some connections

between current discussions and their historical predecessors in Descartes, Spinoza and Leibniz.

Let's now consider this argument. Its first premise is as follows:

- (1) *Non-reductionism*: Mental states cannot be explained entirely in terms of physical states.

“Explained” here has to be understood the way it was introduced in the previous section. This premise therefore states that no physical description alone entails any mental description. The issue may be phrased as asking *why* any physical state is associated with (or just *is*) a mental state. A non-reductionist states that no amount of physical information is sufficient to answer this question. The issue is known as the “hard problem.” There are at least two such hard problems in the philosophy of mind, depending on which feature of the mind is thought to give rise to it. First, and most famously, the hard problem of *consciousness* asks why any physical state would give rise to *conscious experience*, a state that has a kind of *phenomenal character*, “something it is like” to experience it (Nagel 1974; Chalmers 1996). But it has also been argued that there is a hard problem of *content*: namely, the problem of explaining how any amount of syntactic information can be sufficient for instantiating *semantic content* (Hutto and Myin 2013; their characterization of the problem is indebted to the “symbol grounding problem” of Harnad 1990; see also Floridi 2011).¹¹

It is important to realize two things about *Non-reductionism* as stated. First, it is a form of what can be called *epistemic* non-reductionism. It says that mental states cannot be entirely explained in physical terms. This does not prevent them, nevertheless, from possibly *being* physical states, and so (1) is neutral on the issue of *ontological* reduction. Second, (1) does not amount to the claim that *no* physical information is ever relevant to the explanation of a mental state; it just claims that physical information by itself is not *sufficient* for such an explanation.

- (2) *Rationalism about the mind*: Mental states have an explanation.

¹¹ This division between two approaches to the mind-body problem, depending on whether consciousness or (contentful) representation is taken to be the fundamental feature of the mind, is also mirrored in the early modern period. Simmons (2001) argues that this choice divides Descartes, who took consciousness to be fundamental, from Spinoza and Leibniz, who stressed the role of representation. Similarly, there is important precedent for explanatory gap worries in Descartes's claim, fundamental to his dualism, that there are no conceptual connections between mind and body. Because of this, it is difficult for Descartes to answer the question *why* a certain physical state is connected to a certain mental state except by having “recourse to the cause of the whole universe, i.e., to God,” as Spinoza objected (E5pref). See Garrett (2009) and Chapter 5, sections 4-5, for more on Spinoza's reaction to Descartes's dualism.

The explanation involved may be in principle only: maybe we will never be able to actually formulate it. But it exists nonetheless. If God existed, he would know it.

The idea that *everything* has an explanation is known as the Principle of Sufficient Reason, and someone who accepts it subscribes to metaphysical rationalism. Metaphysical rationalism is a controversial position. Against it, it is claimed that there appear to be many brute facts and that it seems to imply necessitarianism – the view, on one formulation, that every truth is necessarily true – which is held to be a deeply unattractive position.¹² Recently, several authors have come to the defense of metaphysical rationalism.¹³ Here, however, there is no need to go into this debate, because although Spinoza and Leibniz were both parallelists and rationalists, the defense of parallelism offered here is not based on metaphysical rationalism, but only on the weaker claim that *mental* phenomena have explanations. Clearly, accepting (2) does not commit one either to necessitarianism or to the rejection of all brute facts.

Even so, (2) remains a strong position. This is particularly so in light of the hard problems that were just mentioned. A hard problem is a request for explanation that seemingly cannot be fulfilled. A skeptic may believe that the persistence of these “explanatory gaps” (Levine 1983) reflects the fact that there is no acceptable explanation of consciousness.¹⁴ Less steadfastly, it is reasonable to think that we do not have such an explanation at present, and this in turn may suggest that there may be mental phenomena that have no explanation.

Still, the fact that we don’t have these explanations yet is almost universally treated as a *problem*, to be answered by further research. This shows that the dominant attitude of philosophers is to pragmatically assume that (2) is true. That is, they proceed on the assumption that there is a complete explanation of the mental, we just have not found it yet. Parallelism proceeds in just this manner.

From these first two premises, it seems we can infer the following principle:

- (3) *Explanatory closure of the mental*: Every mental state can be explained in terms of mental states.

¹² For the latter criticism, see Bennett (1984, 115) and van Inwagen (2014, ch. 7). See also Lin (2012).

¹³ See Dasgupta (2016), Della Rocca (2003a, 2008b, 2010) and Pruss (2006). See also Melamed and Lin (2021).

¹⁴ A position somewhat like this is mysterianism, the view that humans are constitutionally incapable of explaining how physical states produce consciousness (McGinn 1989). Note, though, that even if *humans* are incapable of explaining this, it does not follow that there is no explanation. So even a myste-rian could accept (2).

The reasoning here is intuitive: every mental state has an explanation, but it cannot be (fully) explained in physical terms. The obvious alternative is that it is explained in mental terms instead.

However, this deduction needs more defense. There are at least two gaps in the argument. First, does it follow that every mental state has a *complete* mental explanation? Second, while it is intuitive and elegant to infer that the explanation involved must be mental, this does not strictly follow. After all, not everything that is not physical must be mental.

Let's discuss these in turn. First, the obvious alternative to (3) is not that every mental state has an explanation in mental terms, but that it has an explanation that contains both mental *and* physical terms. This is consistent with the observation that was made before, that (1) does not preclude the occurrence – perhaps even the necessary occurrence – of physical information in an explanation of a mental state. What we get is not (3) but (3'):

(3') *Explanatory closure of the mental (weak)*: The explanation of any mental state involves mental terms.

To say that an explanation involves terms of a certain sort is to say that these terms play a role in the explanation. (3') must be read as saying that mental terms play a *necessary* role in the explanation of any mental phenomenon. Understood like this, the principle is still strong enough to play a part in the deduction of a form of mind-body parallelism. To explain why any mental event occurred requires appealing to mental terms, and in that sense there is a parallel mental explanation to the physical explanation of the corresponding physical event.

Still, (3') introduces a disparity in the status of physical and mental explanations which (3) does not: on (3') mental terms are only necessary to explain mental states, while there may be *complete* explanations in only physical terms of physical states. For this reason, (3') does not fit the overall spirit of mind-body parallelism as well as (3). (3) does not just claim that mental terms are *necessary* to explain a mental event, but that there is a *complete* explanation in mental terms. (3) can plausibly be read as saying that there are psychological laws sufficient for explaining any mental phenomenon.

Explanatory closure of the mental is, no doubt, a strong view. Consider psychophysical laws: laws that allow for the explanation of a mental state *m* in terms of a physical state *p* and a law that connects *p* to *m*. On (3'), the explanation of mental states may be in terms of such laws. It is important to note that (3) does not rule out such lawful connections. However, (3) does say

that there has to be another and more fundamental explanation of mental states as well, one that is fully mental. Accepting this principle places a strong burden on the parallelist to supply such explanations. It is not obvious that a complete mental explanation of either phenomenal character or representational content – the two issues, recall, that can motivate *Non-reductionism* – can be given. However, it is also not obvious that an explanation of them can be given that appeals to both mental and physical terms. Such a fundamental use of psychophysical laws amounts to appealing to a seemingly brute connection between some mental and physical states. In line with parallelism’s commitment to *Rationalism about the mind*, it makes some sense to accept (3) and with it, the demand for a psychological explanation of mental states. So while a form of parallelism based on (3’) is possible, in the remainder I will assume the stronger (3).

The second objection to (3) and (3’) is that, from the fact that an explanation is not physical, it does not follow that it must be mental. (1) and (2) together validly imply that there is no complete physical explanation of a mental phenomenon. But how does that entail that the explanation is (or involves the) mental?

The answer is that it does not entail this, but that it is nevertheless sensible to assume it.¹⁵ Suppose that the explanation *wasn’t* mental. Then there would need to be a third class of phenomena, neither mental nor physical, that was capable of explaining the mental. Discussions in the metaphysics of consciousness suggest two main candidates for this role: either a ‘neutral’ class of entities (which may itself explain *both* the physical *and* the mental, while itself not intrinsically being either); or a ‘proto’-mental kind of entity (specifically, the ‘proto-phenomenal properties’ that, themselves not (yet) phenomenal, are capable of generating true phenomenality).¹⁶ However, while I don’t want to rule out that these things exist, it is surely much more intuitive to ascribe the missing explanatory role to mental phenomena. Perhaps unsurprisingly, the existence of both neutral entities and proto-phenomenal properties is debated, as is the question how such things can help to explain mental phenomena. But the existence of mental entities, and the availability of such mental entities to feature in mental explanations, is surely uncontested. (At any rate, I will ignore those who contest them here.) Given that this provides

¹⁵ An objection along these lines, and suggesting a similar response, was posed to Nagel (1979) by Van Cleve (1990). For discussion, see Coleman (2018).

¹⁶ On neutral monism, see esp. Stubenberg (2018). On proto-phenomenal properties, see Chalmers (2015), Goff (2017) and Stoljar (2018).

us with a convenient source for mental explanations, it makes sense to take it, while acknowledging that other sources may also be available.

So far, we have got the view that there are complete mental explanations for mental states (or, on (3'), that explanations of mental states need to involve mental terms). To get to causal-explanatory mind-body parallelism, it still needs to be established that these explanations are parallel to physical explanations of physical phenomena; and it needs to be shown that there is a chain of mental causes parallel to the physical causal chain. To close these gaps, the argument now appeals to two widely accepted claims.

- (4) *Mental supervenience*: Mental states supervene on physical states.
- (5) *Causal closure of the physical*: Every physical effect has a sufficient physical cause.

Especially (5) has occasionally been contested. Nevertheless, it has become a widely accepted thesis and I will take it as given.¹⁷

This puts us in the home stretch. All that is needed is a principle that allows us to connect causal relations to explanations. This principle can be formulated as follows:

- (6) *Causal explanation*: Explanations of mental and physical states refer to causal relations.

This should be read as a biconditional, saying on the one hand that if mental and physical phenomena stand in causal relations to other mental and physical states, then there is (at least in principle) an explanation of the effect in terms of its cause. Conversely, if there is an explanation of why a certain mental or physical state occurred, this explanation will refer to what caused the occurrence.

Like the other premises in the argument, (6) should be given the least demanding reading possible. In particular, it does not extend to explanations or causal relations that do not involve either mental or physical states. So it is neutral on the question of whether *all* explanations track causal relations.

(6) has a precursor in Spinoza's so-called 'causal axiom': "The knowledge of an effect depends on, and involves, the knowledge of its cause" (E1a4). The causal axiom is a statement of a principle that goes back at least to Aristotle, that knowledge is knowledge of causes. Spinoza's own understanding of this principle is radical enough that he takes it to imply a form of

¹⁷ Papineau (2001) presents a defense of causal closure. The principle is questioned by Lowe (2003), Montero (2003) and Gibb (2015).

parallelism all by itself. For, the axiom implies that ‘knowledge’ of causes and effects stands in the same order and connection as the causes and effects themselves, in the sense that the knowledge of a cause will explain the knowledge of the effect, just like the cause itself caused the effect. So the explanatory relation between knowledge of cause and effect is parallel to the actual causal relation.¹⁸ Without taking over all the implications and presuppositions of Spinoza’s argument, (6) is used here similarly to how Spinoza used the causal axiom: to connect causal and explanatory relations.

Using (6) we can now derive the two missing claims:

(7) *Explanatory closure of the physical*: Every physical state can be explained in terms of physical states. (from 5,6)

(8) *Causal closure of the mental*: Every mental effect has a sufficient mental cause. (from 3,6)

(Alternatively, using (3’) and (6) it is possible to deduce:

(8’’) *Causal closure of the mental (weak)*: Every mental effect has a mental cause.

Unlike (8), this does not claim that every mental effect has a *sufficient* mental cause.)

At this point, then, we have explanatory closure claims for both mental and physical phenomena. These two claims by themselves, however, are not sufficient to establish explanatory mind-body parallelism. What is missing is a guarantee that the mental and physical explanations correspond. For this, we appeal to (4), the uncontroversial claim that mental phenomena supervene on physical phenomena. That is to say, there can be no change on the side of mental phenomena without a change on the side of physical phenomena. This includes causal changes, and so ensures that the mental and physical causal chains correspond. Similarly, because explanations, by (6) *Causal explanation*, appeal to these causal changes, it also establishes explanatory correspondence. So at this point, we can derive both causal and explanatory mind-body parallelism.

(9) *Causal mind-body parallelism*: Mental and physical states are causally parallel. (from 4,5,8)

(10) *Explanatory mind-body parallelism*: Explanations of mental states are parallel to explanations of physical states. (from 6,9)

¹⁸ Spinoza uses E1a4 to establish a form of parallelism in E2p7dem: see Chapter 2, section 4. For further discussion of E1a4, see Wilson (1999c), Morrison (2013) and Lin (2020).

Conjoining (9) and (10) gives us causal-explanatory mind-body parallelism in full.

The premises in this argument are (1), (2), (4), (5) and (6), but each of these is plausible. More controversial is the choice to accept (3), although a version of the argument can be run with (3') instead. I conclude that to someone who is committed to non-reductionism but who nevertheless wants to insist that mental states have explanations and are tethered to the causal order of the physical universe, mind-body parallelism should at least be an open position.

4. *Three objections*

There are a few objections that are often raised against parallelism. Introductory discussions tend to take these objections to be decisive.¹⁹ In this section, I argue that parallelists can develop plausible responses to them.

4.1 *Why anti-interactionism?*

A core tenet of mind-body parallelism is its denial of mind-body interactionism. Spinoza appeals to parallelism to argue that it is absurd to believe “that the body now moves, now is at rest, solely from the mind’s command, and that it does a great many things which depend only on the mind’s will and its art of thinking” (E3p2s; G II, 142).²⁰ Instead, bodies’ movements are caused by other bodies and minds’ thoughts are caused by other ideas. Leibniz, similarly, explicitly puts forward his doctrine of pre-established harmony as an alternative to Cartesian interactionism.²¹ Their confidence in the falsity of interactionism is not shared by modern commentators, however. To them, this rejection was perhaps understandable given certain widely-shared 17th-century conceptions of causation and the essential difference between the mind and the body; but these conceptions have now been mostly abandoned. Anti-interactionism seems to entail the rejection of all forms of (psychophysical) mental causation. But this rejection is either seen as inherently absurd (cf. Jerry Fodor’s well-known remark that without a causal connection between desires and bodily actions, it would be “the end of the world” (1989, 77));

¹⁹ For examples, see the already mentioned Heil (2020, 31-3) and Robinson (2020, §3.3).

²⁰ Jonathan Bennett (1984, 131-3; 1993, 14) has claimed that worries about interaction were a primary motivation for Spinoza’s adoption of parallelism. See also Lin (2020, 221).

²¹ To give just three examples, this is how he presents his view in *Primary Truths* (1689), in the correspondence with Arnauld (1686-7), and especially in the *New System of Nature* (1695; see Leibniz 1989, 143).

or, more weakly, acceptance of mental causation is just more plausible than the parallelist's extravagant position (e.g., Broad (1925) seems to hold this view). According to this objection, if parallelism does reject all mental causation, that is a sufficient reason by itself to reject it in favor of a less extreme view.

While I don't think that rejecting all (psychophysical) mental causation is inherently absurd or that it is sufficient by itself to dismiss a view that implies it, I agree that it is an implausible and unattractive position. Thankfully, parallelism does not entail it. Parallelism is, in part, *causal* parallelism: the view that mental effects have mental causes, and physical effects have physical causes. However, this does not entail that physical effects may not have mental causes as well, or vice versa. In particular, there would of course be mental causation if mental causes are physical causes. There is nothing to prevent a parallelist from accepting an argument of this form:

The mental causation argument:

- (1) Physical state p_1 causes physical state p_2 .
- (2) p_1 = mental state m_1 .
- (3) Therefore, m_1 causes p_2 .

(It is easy to construct a similar argument for physical-to-mental causation.)

In the contemporary mental causation debate, identity theorists routinely appeal to arguments of this sort to accommodate mental causation.²² It is important to realize, contrary to an apparent implicit assumption in introductory discussions of parallelism, that nothing prevents parallelists from adopting an identity theory (as Spinoza did, for example).²³ But even if they reject the identity theory, this still does not commit them to rejecting all (psychophysical) mental causation. Instead, they can take a position like those who, while rejecting the identity theory, nevertheless try to find a role for mental causation. Discussions of the so-called "causal exclusion problem" (Malcolm 1968; Kim 1993, 1998) have produced many proposals for how to do this. I don't see why parallelists should be barred from adopting a similar position. (However, because a response along the lines of the mental causation argument strikes me as more

²² Two examples are Papineau (2002, 17-8) and Kim (2011, 111-4). For more discussion, see Chapter 3, section 3. My discussion in this subsection repeats points made at greater length there.

²³ Many commentators appear to associate parallelism with Leibniz but not Spinoza. Both Heil (2020) and Robinson (2020) only mention Leibniz as a proponent of parallelism in their (admittedly brief) discussions. The same is true of Lowe (2000, 26) and Velmans (2009, 29). Kim (2011, 95-6) does distinguish between Leibniz's and Spinoza's forms of parallelism.

intuitive than these other positions, I do think this counts in favor of a monist version of parallelism.)

Admittedly, there is a complication here resulting from parallelism's reliance on (6), which says that if m_1 causes p_2 , then there is an explanation of p_2 in terms of m_1 . Parallelists should insist, given their adherence to causal and explanatory closure principles, that in such cases there is also always an explanation of p_2 in physical terms (and analogously for mental effects). And given *Non-reductionism*, they should deny that mental effects can be given complete explanations by citing physical causes. They could argue that psychophysical causation is less fundamental than, and presupposes, mental-to-mental and physical-to-physical causation, and hence that it is only fully explanatory given an understanding of these corresponding causal relations. This is an option that should be explored further. But if successful, it should allow parallelists to freely appeal to mental causation.

How can we square this with anti-interactionism and with the historical parallelists? We should distinguish – as I have implicitly been doing – between anti-interactionism and a rejection of all mental causation, and deny that the former entails the latter. Anti-interactionism is the view rejected by Spinoza in the passage quoted above: the view that a mental cause can move the body *all by itself* (without an accompanying physical cause), and vice versa that the body can *by itself* (without an accompanying mental cause) cause a mental state. This view is incompatible with *Causal closure of the physical* and hence with parallelism. It is not often under discussion in the contemporary mental causation debate, which – like parallelism – for the most part is committed to the causal closure of the physical and mental supervenience. But it was a prominent view in the early modern period. Clearly, rejecting anti-interactionism does not involve a rejection of all kinds of mental causation.

Admittedly, neither Spinoza nor Leibniz clearly distinguishes anti-interactionism from mental causation, with the result that they reject both. Both of them held that causal relations imply conceptual relations, while also thinking that mind and body were in some sense conceptually isolated. Arguably, their insistence on these points gets both authors into trouble.²⁴

²⁴ In Spinoza's and Leibniz's parallelist and determinist systems, mind and body exemplify robust counterfactual dependence on each other that is easily interpreted as causal. In the case of Leibniz, he in fact recognizes a category of "ideal influence" that holds, among other things, between mind and body and seems to be based on considerations of counterfactuals of this sort (see, e.g., *Monadology* §51). Leibniz holds that this influence is not strictly causal, however. For discussion, see Jorati (2017, 46-52). As for Spinoza, Koistinen (1996), Davidson (2005) and Lin (2019, 2020, 2021) have argued that he does not in fact reject mental

Contemporary parallelists need not – perhaps should not – take over these commitments, however. They can accept that there is mental causation while continuing to insist on anti-interactionism. This is because what drives their view is a commitment to causal and explanatory closure among mental and physical states separately, with which mental causation does not necessarily conflict.

4.2 *What explains the parallelism?*

Another common objection to parallelism is that the view fails to give a plausible reason *why* mind and body should be (and remain) parallel. In the case of Leibniz’s parallel clocks, it is clear what makes it the case that they show the same time: someone intentionally made sure that they did so. The analogous explanation in the case of mind and body would be that God is responsible for the parallelism. But this invites the objection that such an appeal restricts the plausibility of parallelism to a “theistic framework” (Robinson 2020, §3.3). Even worse, it seems that God only functions in this story as an unsatisfactory “*deus ex machina*” (Heil 2020, 31-3). On the other hand, if we remove God from the picture, it appears that the parallelism can only be a brute fact. However, it is too strange for this: it does need explanation. In the absence of a further account, it is often complained that the parallelist has failed to explain why mind does not simply “float free” from matter (Robb and Heil 2021, §6.4).

This objection touches upon the problem noted before, that parallelism cannot by itself account for why mind and body are parallel. As I indicated there, I think there is a substantial grain of truth to the objection. To say that mind and body are parallel no more provides a full answer to the mind-body problem than to say that they causally interact. It only specifies certain causal and explanatory relations that mental and physical states enter into, and rules out others. Still, it would be a mistake to fault a view for failing to achieve what it was never intended to do in the first place.

A fuller response to the objection becomes possible once the metaphysical questions that parallelism itself leaves open – questions about the nature and relative fundamentality of mental and physical phenomena – have been answered by the adoption of one of the traditional positions in the mind-body problem. Because parallelism itself is to a large extent neutral on these

causation. In Chapter 3, I discuss their interpretation and argue that, while Spinoza in my eyes does reject mental causation, there is strong pressure on him to accept some version of it.

wider questions, many of these positions are compatible with it (just as many positions are compatible with mental causation). Parallelism neither identifies nor distinguishes mental and physical states necessarily, and so it is neither intrinsically monist nor dualist. Within the general limits that are set by *Non-reductionism*, it is also compatible with either the mental or the physical being more fundamental than the other, or with them being equally fundamental. So there could be neutral monist, dual-aspect monist, dualist, and even physicalist and idealist versions of parallelism.

It is these metaphysical positions that can then incorporate mind-body parallelism in order to develop a full answer to the mind-body problem. They should also explain why mind and body are parallel. Just like in the case of anti-interactionism, monist positions have an advantage here. Not only does it make sense that mental and physical states correspond if they are identical; monist positions can also appeal to the necessity of identity to explain why mental and physical states must correspond. This will also help with the laws issue from before: the easiest explanation of why laws governing bodies produce the same causal regularities as laws involving minds is that the minds and bodies in question are identical (so that the physical laws in fact apply to minds just as the mental laws apply to bodies). However, again I see no reason why a dualist parallelist should not be able to answer these questions as well.

This may at first seem to render parallelism almost contentless as far as its metaphysics goes. But that cannot be right. Parallelism's causal closure principles and commitment to (epistemic) non-reductionism are hardly empty positions. Meanwhile, the real promise of the view lies in the ability it may have of making progress on entrenched debates in a way that is independent of any particular metaphysics. If the argument presented here is correct – and if it can be backed up by successful concrete applications of parallelism, something beyond the scope of this chapter – then this shows that the standoff between physicalist and non-physicalist positions that has been at the core of many discussions in the philosophy of mind is unnecessary, insofar as parallelism can make progress while being compatible with both. Admittedly, this remains to be seen. But it supports my contention that parallelism's lack of a metaphysical framework is not a weakness of the view.

Having shown the kernel of truth in the objection, it is now possible to also point out some of its mistakes. Parallelism no more than most other positions in the philosophy of mind is intelligible only from within a “theistic framework.” Theological considerations have not played any role in the argument above. Instead, what generates parallelism is whatever justifies *Non-*

reductionism, coupled with the considerations about explanation and causation that were set out above. Once these are positioned in an ontological framework, they explain in a non-arbitrary manner why mind and body are parallel and are prevented from “floating free.”

4.3 Is this panpsychism?

A final worry about parallelism is that it looks like a form of panpsychism, and panpsychism is often seen as an unattractive position.

Why would parallelism be panpsychist? Panpsychism is commonly defined as the view that mentality is *fundamental* and *ubiquitous*.²⁵ In other words, mentality is found at the fundamental level of reality and also extends (much) more widely than is commonly thought – i.e., (far) beyond the higher animals.

One reason why parallelism looks like it might be panpsychist is that it satisfies the ubiquity criterion. It is easy to see why. By *Causal closure of the mental*, any mental state has a prior mental state that causes it. There is no point at which the mental suddenly pops into existence. (So parallelism is incompatible with emergentism, the view that mental states ‘emerge’ from complex compositions of physical states.) On both phylo- and ontogenetic developmental time-scales, mental states enter in causal chains that extend far past what we ordinarily recognize as mentality. So parallelism, on my argument, is committed to there being something describable as mentality in, for example, just-conceived embryos and the remote evolutionary ancestors of *homo sapiens*. Playing on the double meaning of *animata*, which in his time could mean both ‘alive’ and ‘having a mind,’ Spinoza famously deduces from parallelism the claim that all “individuals [...], though in different degrees, are nevertheless *animata*” (E2p13s).

However, while parallelism meets the ubiquity criterion, it is not obvious that it meets the other one, of the mental being fundamental. Again, parallelism itself is mute on the fundamentality of the mental, both in itself as well as relative to the physical. Any concrete view about this will only follow once parallelism is slotted into a specific metaphysics. What is more, even when supplemented in this manner, many versions of parallelism will not assign a fundamental status to the mental. The previous subsection already pointed out that a (nonreductive) physicalist

²⁵ Brüntrup and Jaskolla (2016) and Seager (2020) are two recent collections centered on panpsychism. For discussion of the currently most popular form of panpsychism, Russellian monism, see additionally Alter and Nagasawa (2015) and Goff (2017).

form of parallelism is possible. This is because *Non-reductionism*, as was explained above, does not go beyond an explanatory claim: it is the claim that mental states cannot be entirely *explained* in physical terms. But that is compatible with there being a difference in fundamentality between them. This in itself already shows that the objection is wrong to suppose that parallelism *must* be panpsychist.

A neutral monist form of parallelism will also not see the mental as fundamental. Neutral monism is normally seen as incompatible with panpsychism, because it holds that both the mental and the physical depend on a fundamental level of reality that is intrinsically neither mental nor physical. It takes the mental and the physical to be *equally* fundamental, but neither to be *absolutely* fundamental.

Only dualist and dual-aspect monist forms of parallelism will be panpsychist, therefore. We have already seen that there are reasons to think that dualist parallelism is a less attractive view than its monist alternatives. Still, it is one possible form of parallelism. Moreover, unlike a dualist version, a dual-aspect monist version of parallelism fits the view naturally and it does assign fundamental status to the mental (as one ‘aspect’ under which reality manifests). Furthermore, it is also independently interesting to say a little more about the relation of parallelism to panpsychism. After all, the views do show an important similarity in accepting the ubiquity of the mental. So, suppose that we adopt a panpsychist form of parallelism. Is this a bad thing?

That remains to be seen. Over the last quarter-century, panpsychists have carved out a central place for themselves in the philosophy of mind. Activity surrounding panpsychism is high and the view looks set to become a central fixture of the field. It is often objected to the view that it is just too counterintuitive to be believable. It is difficult, for example, to imagine the kind of mentality that very simple organisms are capable of, and this is sometimes brought up as an objection to the view. But this is a weak objection. Panpsychists themselves admit that their view is unintuitive. They have also pointed out that it is not necessary for us to be able to imagine such simple forms of mindedness. Nothing about panpsychism requires all forms of mentality to be like ordinary human consciousness. So the fact that we cannot imagine what it is like to be a fundamental entity does not count against panpsychism.²⁶

²⁶ Seager clearly makes this point in his introduction to Seager (2020).

To sum up, the objection is wrong to assume that every form of parallelism is panpsychist. Moreover, panpsychism itself is a view that is here to stay and cannot be dismissed just on the basis that non-human mentality is difficult to imagine.

5. Conclusion

The aim of this chapter was to present a defensible version of mind-body parallelism. In order to do this, I first clarified what I take mind-body parallelism to encompass. Parallelism is not the caricature sometimes encountered in the literature, of two distinct entities placed, seemingly arbitrarily, on parallel causal tracks. Instead, it is the view that mental and physical phenomena, closely connected but separated at least conceptually because of various explanatory gaps, have corresponding causal explanations. My argument aimed to show how this view follows from certain widely shared commitments, especially a commitment to non-reductionism and to the view that mental states have explanations. Given the explanatory gap, a reductive naturalism that explains the mind in physical terms is off the table. But parallelism offers one possible model of the next best thing: a non-reductive naturalism that shows how mind is tethered to the causal order in the physical universe yet has some degree of (at least conceptual) independence from it. I further hope to have shown that common objections to the view are inconclusive.

*Spinoza's Arguments for Mind-Body Parallelism**1. Introduction*

Spinoza is a mind-body parallelist: he thinks that the mind stands in the same causal and explanatory relations as the body. In Spinoza's metaphysics, minds are modes of the attribute of thought and bodies are modes of the attribute of extension. Both kinds of modes depend on and inhere in God, the single substance, who produces "infinitely many things in infinitely many modes" (E1p16), and without whom "nothing can be or be conceived" (E1p15).

Spinoza's mind-body parallelism is not just the claim that the mind is parallel to the body, however. It also incorporates the idea that the mind is the idea of, so represents, the body (E2p13). In Della Rocca's terms, Spinoza's mind-body parallelism is not a "bare" but a "representational" parallelism (Della Rocca 1996, 19). And as will become clear, this makes all the difference. I will argue that Spinoza's mind-body parallelism incorporates this claim about what the mind represents, and cannot be stated without it.

For a long time, commentators tended to hold that E2p7 and its scholium both established basically the same view. Deleuze ([1968] 1990, ch. 7) early on disagreed with this consensus and distinguished two separate parallelist doctrines. Recently, a similar distinction has been defended by Melamed (2013b, chs. 5-6). Both authors convincingly show that the parallelism of the proposition is not the same as that of the scholium.²⁷ And I would add that *neither* is the same as the parallelism Spinoza is most famous for, mind-body parallelism.

²⁷ Nadler (2006, 127-9) does not distinguish these as separate views, but does note that Spinoza presents both an "epistemological" and a "metaphysical" argument for parallelism. Gueroult (1974), making yet other

This chapter has two aims. The first is to reconstruct Spinoza's arguments for mind-body parallelism and the place this doctrine takes in his wider philosophical system. E2p7 is considered the *locus classicus* of Spinoza's parallelism doctrine – although we will see that this traditional estimation has to be qualified somewhat. (Note that when I refer simply to 'E2p7,' I normally intend to refer not just to the proposition itself, but also its demonstration, corollary and scholium. When referring to the proposition itself, I will use 'E2p7e,' where 'e' abbreviates 'enunciation.') But what is that doctrine precisely; how is it established; and are the demonstrations valid? Although these questions have been discussed extensively, there is to my knowledge no recent, detailed discussion that answers all three. (The most important recent treatment, that by Melamed 2013b, is mainly concerned to distinguish between Spinoza's two parallelisms and spends little time on the validity of the arguments.)

The second purpose of the chapter is to vindicate the ascription of parallelism to Spinoza. Recently, a few different authors have questioned this ascription.²⁸ Three major concerns are that the term 'parallelism' is not found in Spinoza and hence that it apparently lacks a textual basis; that traditional construals of mind-body parallelism neglect the role of knowledge and representation in Spinoza's view; and that parallelism is unable to account for the mind-body union. In this chapter, I present a version of mind-body parallelism that does not have these flaws. I show that there is a textual basis for the notion and argue – as already indicated – that representation plays a central role in E2p7 and that it plays a role in explaining the mind-body union.

The chapter is structured as a step-by-step discussion of E2p7. I begin by describing the general form of a parallelism (section 2). In section 3, I describe the representational parallelism that Spinoza sets out in E2p7e and that is the topic of its demonstration and corollary. In section 4, I discuss a question that has recently sprung up about the nature and success of E2p7dem, and reconstruct a valid version of this demonstration. Section 5 discusses E2p7c and the surprising theory of representation it implicitly contains. Section 6 is devoted to the metaphysical parallelism that Spinoza establishes in E2p7s and its relation to the representational parallelism. Finally, section 7 argues that neither of these parallelisms is identical to mind-body parallelism proper, which is not established until E2p13.

distinctions, argues that E2p7 establishes no less than three different parallelisms. See also Aquila (1978) and Bennett (1984, 153).

²⁸ See Renz (2018, 109-10), Yakira (2015, 53-71), and esp. Jaquet (2018, 12-9; 2021).

2. E2p7e: *The general form of a parallelism*

E2p7e: The order and connection of ideas is the same as the order and connection of things.

Before talking about the content of E2p7e, let's consider its form. Several questions arise. Why does Spinoza distinguish between “order” and “connection,” and what do these consist in? What does he mean by “the same as”: does he mean that the order and connection of ideas supervenes on the order and connection of things, but not vice versa, or is he making a symmetrical claim? (If he is only making a supervenience claim, he is only saying that the order of ideas maps onto the order of things, not that the order of things likewise maps onto the order of ideas; see Morrison 2013.) And why is he talking about “ideas” and “things” here, given that parallelism is usually taken to be a claim about modes in general? I will try to briefly answer the first two questions here, leaving the last question for the next section. Answering these questions will elucidate the general structure of a parallelism in Spinoza, which will be helpful going forward.

Let's begin with the question about “order” (*ordo*) and “connection” (*connexio*). Why does Spinoza distinguish these in this way? They appear to end up extensionally coinciding: for any two things, if they stand in the same order, then they have the same connection, and vice versa. In the scholium, Spinoza writes of “one and the same order, *or (sive)* one and the same connection of causes.” *Sive* in Spinoza typically indicates the equivalence of the two terms it connects (most famously in the phrase *Deus, sive Natura*). At the same time, the terms do have a different meaning. As the names suggest, “order” refers to the series in which things are produced, while “connection” refers to the nature of the connection between them.²⁹ The former term already occurs in E1p33:

E1p33: Things could have been produced by God in no other way, and in no other order than they have been produced.

E1p33 states that there is a single causal order that subsumes all things and is absolutely necessary. In its demonstration, this order is identified as the “order of Nature.” It is plausible to read E2p7e as picking up this notion and as asserting that this order manifests itself both among ideas and among things. This reading derives support from the scholium, in which Spinoza

²⁹ See Jaquet (2011).

refers to a parallelist thesis as applying to the “order of the whole of Nature.” However, while this notion is in the background of E2p7, Spinoza does not cite E1p33 and so may be using the term “order” in a more neutral sense, in which it refers just to a certain series in which things stand.

Unlike “*ordo*,” E2p7e contains the first occurrence of the term “*connexio*” in the *Ethics*.³⁰ The term, in the first instance, indicates only that there is a connection between ideas and between things and that this connection is the same in both cases. The nature of the connection is left unspecified. However, given that E1p33 refers to a *causal* order in which things are produced, it is natural to take the connection to be a causal one. E2p7e therefore states that ideas and things stand in the same *causal* order and connection. Given some of Spinoza’s fundamental ideas about the relation between causation and conception however – which will be discussed at greater length in section 4 – I think we should also read E2p7e as stating a *conceptual* parallelism between ideas and things. That is to say, ideas are not just caused in the same order and connection as things, but they are also *understood* in the same order and connection.³¹ The order and connection is both causal and conceptual in nature.

The conceptual order can be taken as the order in which things are understood through each other. If that is right, then E2p7e asserts not just one, but two claims: it says not just that ideas stand in the same causal connections as their objects (and vice versa), but also that they stand in the same ‘understanding through’ – or conceptual – relations (and vice versa). It is now possible to introduce the term ‘parallelism’ to describe situations in which two sets of things stand in the same “order and connection.” I will refer to the items in these sets as ‘*x*’s and ‘*y*’s’ and use x_n to indicate the *x*-item that corresponds to *y*-item y_n . The general form of the two cases under discussion can then be described as follows:

³⁰ There is a reference to things being *connexis*, “connected” in E1p15s (G II, 59). But E2p7e has the first occurrence of the noun *connexio*.

³¹ This notion of “understanding through” raises some technical issues. Spinoza says that some things are “conceived through” others (for example, modes are conceived through their attributes). He also says that “the cognition of an effect depends on, and involves, the cognition of its cause” (E1a4; translation modified; see fn. 42 below). An influential interpretation, associated particularly with Della Rocca (2008b), holds these two relations to be the same. But it is noteworthy that Spinoza does not use the term “conceive through” in E1a4. And he generally never appears to use the term when discussing causal relations between modes, but only in connection to modes’ dependence on attributes or substance. After pointing this out, Melamed (2013b, ch. 3) argues that the two relations are in fact distinct. Even if this is so, however, it is clear that both relations involve a conceptual connection of some sort, making it true to say that in both cases one thing is “understood through” another. I therefore use the terms “understanding through” and “conceptual relation” in this broad sense to cover both relations.

Causal parallelism: For any x_n and y_n , x_n is caused by x_{n+1} iff y_n is caused by y_{n+1} .

Conceptual parallelism: For any x_n and y_n , x_n is understood through x_{n+1} iff y_n is understood through y_{n+1} .

These formulations state the existence of a certain correspondence between x 's and y 's. The ontological status of these things and the source of their correspondence is left open by these general parallelist claims. In E2p7e, they are identified as 'ideas' and 'things.' We'll explore the significance of this choice in the next section. Clearly, however, many other things could be parallel in the sense under discussion here. This includes things that are of the *same* ontological category. In Leibniz's famous example, the successive states of the hands of two clocks – both physical objects – are causally parallel.

These parallelisms take the form of biconditionals, which can be decomposed into left-to-right and right-to-left halves. However, I take E2p7e to assert both halves of both biconditionals in relation to ideas and things, and I think the same is true for the parallelism of E2p7s. Given the central importance of E2p7 for what is referred to as Spinoza's parallelism in the literature, I think it is therefore fair to say that the *general form* of a parallelism, for Spinoza, is one where both causal and conceptual parallelism apply. In other words:

General form of parallelism, first pass: For any x_n and y_n , (i) x_n is caused by x_{n+1} iff y_n is caused by y_{n+1} ; (ii) x_n is understood through x_{n+1} iff y_n is understood through y_{n+1} .

This can in turn be simplified as follows:

General form of parallelism: For any x_n and y_n , x_n is caused by and understood through x_{n+1} iff y_n is caused by and understood through y_{n+1} .

There may be cases where a causal parallelism obtains without a conceptual one, or vice versa. Such cases would not fit the general form of a parallelism, as exemplified in E2p7. Because I think that Spinoza thinks causal and conceptual relations are coextensive, I am skeptical that there are such cases, but this chapter is not the place to defend a definite stance on this issue. In the remainder of this chapter, I will use the term 'parallelism' to refer to cases that exhibit this general, causal-conceptual form.

This characterization of the general form of parallelism differs somewhat from how the view is normally portrayed. My choice to express the view in terms of biconditionals is indebted to Morrison (2013). Most other authors instead say that parallelism consists in an "isomorphism,"

or one-to-one correspondence, between the parallel items.³² My definition, by contrast, does not commit to a full isomorphism, because it only spells out a parallelism given two related things x_n and y_n . Isomorphism only follows if these things stand in causal chains going back infinitely, which the definition itself is mute on. I prefer this definition because it decouples parallelism from the heavy-handed metaphysical commitment to these infinite causal chains, while maintaining what I think is parallelism's most salient feature, its commitment to there being corresponding causal explanations of parallel items. This also makes parallelism adaptable to contexts that do not share that commitment.³³ Furthermore, I do not think a commitment to isomorphism is present in E2p7e and E2p7dem, at least (unlike E2p7s, which does clearly commit to it). My reconstruction of E2p7dem in section 4 only depends on the claim that given an effect and the idea that represents it, these have parallel causal explanations in terms of the cause and its idea. I do not see Spinoza relying on or asserting the claim that there is a complete isomorphism between ideas and things in E2p7dem. To be clear, there is no doubt that Spinoza *does* think this isomorphism holds; he clearly thinks that every finite thing has a finite cause, which when coupled with parallelism is enough to generate isomorphism (see E1p28). However, it is not clear to me that isomorphism is necessarily a part of the *definition* of parallelism.

Another divergence from standard definitions is my choice to explicitly indicate a role for a *conceptual* parallelism. A number of recent commentators take E2p7 to in the first place state only a *causal* parallelism.³⁴ This need not mean that they think Spinoza does not also commit to a conceptual parallelism, because they may think, like I do, that conception is coextensive with causation for him. In that case, the ascription of a causal parallelism to Spinoza would entail also ascribing a conceptual parallelism to him. I still think it is preferable, however, to separate these two doctrines. Not only is this clearer, but we will see that claims about the nature of conception play an important role in E2p7dem.

Understanding parallelism in this way, as referring to what Spinoza calls a shared “order and connection,” shows that there is clear textual evidence for ascribing parallelism to him. That Spinoza does not use the term ‘parallelism’ is unimportant. Its status is the same as that of

³² The term “isomorphism” is used explicitly by Deleuze (1988, 87), Della Rocca (the two orders are “causally isomorphic,” 1996, 18; also his 2008b, 90), E. Marshall (2013, 11), Melamed (2013b, 139), and Lin (“causal isomorphism,” again: 2019, 85; see also his 2020, 220). In fact, Morrison also builds isomorphism into his definition.

³³ So, one advantage of the definition is that it can be fairly easily applied to contemporary discussions, as I try to do in Chapter 1.

³⁴ See the references to Della Rocca and Lin in the previous footnote, as well as Bennett (1984, 127) and Melamed (2013b, 152).

other technical terms – such as ‘substance monism’ and ‘necessitarianism’ – that are also not used by Spinoza but are nevertheless widely used in the secondary literature.³⁵

3. E2p7e: *Representational parallelism*

After having characterized the general form of a parallelism, let us now consider the instance of this form at issue in E2p7e.

E2p7e states a parallelism, not between *mind* and *body*, nor even between *modes of different attributes*, but between *ideas* and the *things* they represent. I will accordingly refer to this view as *representational parallelism*.³⁶ That Spinoza is concerned with representation is clear from the demonstration, which talks about the relation between causes and effects on the one hand and the *cognition* of these causes and effects, which represents them, on the other.

Why did Spinoza use the term “thing” (*res*) for this doctrine? The “ideas” in question are clearly modes of the attribute of thought. But the attribute of the “things” is not specified. I think it is clear that the things in question *are* modes.³⁷ This is not obvious from the term itself, given that God for example (a substance) seems a perfectly fine example of a thing. In fact, Spinoza refers to God as a “thinking thing” and an “extended thing” in E2p1 and E2p2, respectively. But before E2p7, Spinoza more consistently uses the term “thing” to refer to those entities that “follow from the nature of God” (E1p16; cf. E2p3, E2p4, and many occurrences in the second half of Part One). That this is the meaning of “thing” intended in E2p7e is made clear by E2p7c’s restatement of a parallelist doctrine as applying to “whatever follows formally from God’s infinite nature” – a clear callback to E1p16. Only the modes, not substance or the

³⁵ Of course it is possible to doubt whether any of these terms describe Spinoza’s actual views, as Lærke (2012) does for substance monism and Curley and Walski (1999) for (strict) necessitarianism. My point here is just that the fact that a historical author does not use a certain term is not sufficient reason not to use that term to describe their views, provided there is a textual basis for it.

³⁶ In this I am adapting the terminology of Della Rocca (1996, 19). Melamed calls it an “ideas-things” parallelism (but does characterize this as a “representational” parallelism; 2013b, 145). Deleuze (1990, 113), stressing the role that claims about God’s knowledge play in the argument for the view, calls it an “epistemological” parallelism.

³⁷ *Pave* Melamed (2013b, 143-4, 179-80) who includes God in the extension of “things”. Cf. Deleuze: “The epistemological viewpoint amounts to this: that given a mode in some attribute, there corresponds to it in the attribute of Thought an idea that represents it, and it alone” (1990, 114). Gueroult (1974, 66) and Rice (1999, 48) take the things in question to be modes, like I do, but modes of the attributes *other than* thought. I don’t think it’s necessary to restrict the domain of representational parallelism in this way. Ideas themselves are also things (see Della Rocca 2008b, 90-1).

attributes, follow from God's nature. To put the point in Spinozistic terms, the "things" are just what is contained in *Natura naturata*, not *Natura naturans* (see E1p29s).

Spinoza never defines the term "thing" that he uses here. E2def7 defines "singular things" (*res singulares*; also referred to as "individuals," *individua*) as "things that are finite and have a determinate existence." Clearly, by this definition, the "things" of E2p7e need not be individuals, because they may be infinite. The infinite modes (E1p21-3, Ep64) can fall under E2p7e.³⁸ (It also covers things that do not exist, a difficult topic that Spinoza addresses in E2p8. Unfortunately, there is no place to discuss it here.³⁹)

If the "things" in question in E2p7e do truly include just the modes, then this necessitates also imposing a restriction on the "ideas." For it rules out the idea of God as a subject of the representational parallelism. This idea has God for its object, and because God is not a thing in the sense under discussion in E2p7e, his idea seemingly also is not covered by E2p7e.⁴⁰

The standard way of referring to Spinoza's parallelism is as the doctrine that the modes of each attribute stand in the same order and connection. (I call this view 'metaphysical parallelism' and discuss it below.) But E2p7e does not state that view. Nor does it state a view that directly implies it. It says that the order and connection of ideas is the same as the order and connection of the modes these ideas represent. This says both *more* and *less* than the standard, metaphysical reading of parallelism. It says more, because the represented things may themselves be ideas. Representational parallelism therefore entails that ideas are parallel to the second-order ideas that represent them. (Spinoza refers to the latter as "ideas of ideas": see E2p19-22.) The standard reading is mute on such a parallelism internal to thought. Secondly, however, the representational reading also says less than the standard reading, because it does not logically rule out violations of parallelism between attributes other than thought. Suppose that a mode of extension x is parallel to a mode of the third attribute y , which in fact is parallel to x . Let $I(x)$ be the

³⁸ This point, however, is a little more delicate than at first appears, because Spinoza refers to what he identifies as the mediate *infinite* mode of extension as an individual in E2p13s (at G II, 102). Even so, there are also *immediate* infinite modes, and these are never described as individuals by him.

³⁹ See Lærke (2017) for an interesting recent discussion of this topic.

⁴⁰ The idea of God poses a number of challenging questions in relation to Spinoza's metaphysics; unfortunately, these cannot be pursued here. For discussion, see Deleuze (1990, ch. 7), Melamed (2013b, ch. 6) and Koistinen (2018). Spinoza refers to the immediate infinite mode of thought as God's infinite intellect (Ep64). Elsewhere, he identifies the same mode as the idea of God (E1p21dem). By representational parallelism, the object of God's infinite intellect is the immediate infinite mode of each of the other attributes, *in addition* to God. In that case, perhaps God's idea falls within the scope of representational parallelism insofar as it represents these other modes, but not insofar as it represents God. Even if that is right, however, it is clear that many difficult questions remain.

idea that has x for its object. Representational parallelism requires there to be ideas of both x and y , i.e. $I(x)$ and $I(y)$. But it says nothing about whether these ideas are identical or not. That is, if we knew that x is parallel to $I(x)$, y to $I(y)$, and that $I(x) = I(y)$, this would allow us to deduce that x is parallel to y . But representational parallelism does not allow us to identify $I(x)$ and $I(y)$, because they represent only x and y and nothing about their relation. Representational parallelism by itself therefore does not rule out a situation in which the parallel modes of different attributes each have their own, distinct idea in the attribute of thought. And so it is also logically compatible with there being one order and connection that is shared by x and $I(x)$, and a *different* order and connection that exists between y and $I(y)$.⁴¹

As the name implies, representational parallelism is a view about how representations relate to the things they represent. In other words, if one thing x is caused by or understood through another thing y , then $I(x)$ will be caused by or understood through $I(y)$, and vice versa. (Here and throughout the chapter, I am only discussing ideas that adequately represent. I may think that vaccines cause autism, i.e., my idea of vaccines is causally related to, and implies, my idea of autism. In this case there is not a parallel causal relation between the things these ideas represent.) I will have more to say about representation when I discuss E2p7c. First, however, we should investigate how Spinoza demonstrates representational parallelism.

4. E2p7dem: *The role of the causal axiom*

E2p7dem: This is clear from E1a4 [“The cognition of an effect depends on, and involves, the cognition of its cause”]. For the idea of each thing caused depends on the cognition of the cause of which it is the effect.⁴²

The demonstration of E2p7, just quoted in full, is notorious for its brevity. It amounts to a recalling of one of the foundational axioms of the *Ethics* – E1a4, also known as the “causal axiom.” Spinoza then derives an unexpected result from this axiom. The result is representational parallelism. But it is difficult at first to see how it follows, or even whether it does.

⁴¹ This paragraph is indebted to Melamed’s discussion of the differences between the two parallelisms (Melamed 2013b, 142-3).

⁴² I have substituted “cognition” for Curley’s “knowledge” in both the demonstration and the quotation of E1a4. The Latin *cognitio* can mean both. However, because E2p7dem to a large extent rests on identifying a *cognitio* as an idea, my rendering makes the demonstration more intuitive.

Given the central role it plays, whether E2p7dem is seen as successful largely depends on how E1a4 is interpreted. It has often been noted that E1a4 is a statement of the traditional Aristotelian principle that knowledge is knowledge of causes.⁴³ These same commentators then point out that Spinoza's use of this principle in E2p7 constitutes a significant divergence from tradition.⁴⁴

E1a4 says that, given two things that are causally related, the effect is to be understood through the cause. This clearly implies the following principle:

Causation \rightarrow *Conception*: If x is caused by y , then x is understood through y .

However, the causal axiom has often also been taken⁴⁵ to state the converse of this claim:

Conception \rightarrow *Causation*: If x is understood through y , then x is caused by y .

How does this follow from the axiom? By saying that the understanding of an effect depends on an understanding of its causes, it can be taken to say that in order to gain knowledge of the effect, we need to understand how it follows from its causes. For example, Deleuze (1990, 115) glosses part of the meaning of E1a4 as the idea that for Spinoza, "nothing can be known independently of a cause of its being, in essence or in existence."⁴⁶ Additionally, it is claimed that other parts of the *Ethics* support ascribing *Conception* \rightarrow *Causation* to Spinoza. E2p7 itself is among them, as we noticed in section 2.⁴⁷ I will call the view that E1a4 implies both claims the *standard reading* of E1a4. In this section I will argue that E2p7dem supports the standard reading.

In spite of some textual evidence in its favor, *Conception* \rightarrow *Causation* is a strong claim, and recently some authors have expressed hesitation to ascribe it to Spinoza. Morrison (2013) has shown that various texts that are often taken to support the standard reading, including E2p7, can in fact be reconstructed in a way that uses only *Causation* \rightarrow *Conception* and not its converse. Without taking a definite stance on whether he ascribes *Conception* \rightarrow *Causation* to Spinoza, Lin (2019, 87-8; 2020) has suggested a similar reconstruction of E2p7dem. Such a reading can seem to fit the text well. In E2p7dem, Spinoza appears to incorporate E1a4 into the metaphysics of ideas that he has been expounding in the early propositions of Part Two. In the second sentence

⁴³ See e.g. Gueroult (1974, 64), Deleuze (1990, 115) and Lin (2019, 24).

⁴⁴ See Wilson (1999c) and Lin (2020) for further discussion of E1a4.

⁴⁵ Jarrett (1978, 29) was perhaps the first to model E1a4 as this biconditional. See Morrison (2013, 2) for further references.

⁴⁶ See also Hübner (2019, 4).

⁴⁷ Other locations cited in support include E1p3, E1p25, E2p5, E2p6, and E3p6.

of the demonstration, he is drawing attention to the fact that, for him, conceiving of something (or ‘cognizing’ it) is tantamount to forming an idea of it. That gets us the following reconstruction:⁴⁸

- (1) The cognition of an effect depends on, and involves, the cognition of the cause. (E1a4)
- (2) To conceive a thing x is to form the idea $I(x)$ representing x . (premise)
- (3) The idea of an effect depends on, and involves, the idea of its cause. (from 1,2)

(3) states that in case of a cause of an effect, the idea of the effect will ‘depend on, and involve’ the idea of the cause. This suggests that (3) can be reformulated as:

- (4) If one thing x is caused by another thing y , then $I(x)$ is caused by $I(y)$.

To get (4), it is necessary to assume that Spinoza believes that the relation of “dependence and involvement” mentioned in E1a4 entails – or perhaps is even identical to – a causal relation. Let’s grant that assumption. There are two further things to note about this reconstruction. First, note that while (4) only states a causal parallelism between things and ideas, we can, by appealing to *Causation* \rightarrow *Conception*, also derive the corresponding conceptual claim. Second, and more importantly, however, this still only gets us one direction of representational parallelism: the thing-to-idea direction. We are left wanting a demonstration of the converse, idea-to-thing half.

Morrison (2013, 12-4) notices this gap, but believes it is intentional and is due to Spinoza’s desire to avoid a commitment to *Conception* \rightarrow *Causation*. Lin (2020, 221, 226) believes it is unintentional, but that Spinoza gives a better, alternative argument in E2p7s. Morrison additionally claims that Spinoza supplies us with the missing direction immediately before E2p7, in E2p6c. There, Spinoza states that

the objects of ideas follow and are inferred from their attributes in the same way and by the same necessity as that with which we have shown ideas to follow from the attribute of Thought.

It seems that if things “follow and are inferred [...] in the same way,” then they stand in the same causal and conceptual relations. So this states that the objects of ideas are related in just the same way the ideas themselves are related. Although E2p6c only refers to the relation between these ideas and objects on the one hand and their attributes on the other hand, Spinoza

⁴⁸ Neither Morrison nor Lin gives a reconstruction of E2p7dem; I am here supplying one on their behalf.

nevertheless does seem to see it as a statement of the idea-to-thing half of representational parallelism. For as Morrison points out, in E5p1dem Spinoza cites E2p6c as evidence for the idea-to-thing half. Furthermore, Morrison also refers to E5p1dem as a location where Spinoza appears to distinguish the two halves.⁴⁹

I will grant that E2p6c already states the idea-to-thing half of representational parallelism. But it still strikes me as unsatisfactory to read E2p7dem as defending only its complementary half. After all, I have already argued that E2p7e is best read as asserting both halves. I now want to briefly argue that if we do take Spinoza to accept *Conception* \rightarrow *Causation*, we can use E2p7dem to deduce representational parallelism in full. Using the standard reading of the causal axiom, the demonstration can be worked out as follows.

- (1) The cognition of an effect depends on, and involves, the cognition of the cause. (E1a4)
- (2) If x is caused by y , then x is understood through y . (from 1; *Causation* \rightarrow *Conception*)
- (3) If x is understood through y , then x is caused by y . (from 1; *Conception* \rightarrow *Causation*)
- (4) To conceive a thing x is to form the idea $I(x)$ representing x . (premise)
- (5) If x is caused by y , then $I(x)$ is understood through $I(y)$. (from 2,4)
- (6) If $I(x)$ is understood through $I(y)$, then x is caused by y . (from 3,4)
- (7) If $I(x)$ is understood through $I(y)$, then x is understood through y . (from 2,6)
- (8) If x is understood through y , then $I(x)$ is understood through $I(y)$. (from 3,5)
- (9) x is understood through y iff $I(x)$ is understood through $I(y)$. (from 7,8; *Conceptual representational parallelism*)
- (10) If x is caused by y , then $I(x)$ is caused by $I(y)$. (from 1,4)
- (11) If $I(x)$ is caused by $I(y)$, then x is caused by y . (from 2,6)
- (12) x is caused by y iff $I(x)$ is caused by $I(y)$. (from 10,11; *Causal representational parallelism*)
- (13) x is caused by and understood through y iff $I(x)$ is caused by and understood through $I(y)$. (from 9,12; *Representational parallelism*)

I have tried to spell out each separate claim in the derivation in order to show the roles that *Causation* \rightarrow *Conception* and *Conception* \rightarrow *Causation* play in it. It is fair to say, seeing this reconstruction, that E2p7dem presents us with an extremely condensed demonstration. Nevertheless, if my reconstruction is correct, it is a valid demonstration of representational parallelism.

⁴⁹ Melamed (2013b, 147n18) also points out E2p6c and E5p1dem as evidence that Spinoza only wants to argue for one side of representational parallelism in E2p7dem. Compare also Gueroult (1974, 74-5), Peterman (2017, 229n44).

Note that introducing (4) does not require Spinoza to give up his previous, and foundational, claim that things can be understood through other things. It is not just ideas that are understood through each other. (Remember, for example, that modes are understood through their attributes.) Rather, what he is saying is that a thing is understood through another thing just in case the idea representing the first thing is understood through the idea representing the other thing. But this obviously does not render the first half of this biconditional false or incapable of applying to things that are not ideas. It does, however, straightforwardly imply an infinite progression of ideas of ideas in the attribute of thought; but Spinoza is unphased by this entailment and even affirms it openly (E2p21s, E2p43s).⁵⁰

This reconstruction of E2p7dem works on the assumption that E1a4 implies both *Causation* \rightarrow *Conception* and *Conception* \rightarrow *Causation*. This second claim – (3) in the argument – is used to establish (6) – which in turn is cited in (7) and (11) – and (8). By taking E1a4 to imply only the first of these two claims, Morrison (2013) and Lin (2020) are unable to derive anything more than the left-to-right, or thing-to-idea half of representational parallelism on the basis of E2p7dem. However, I think the more natural reading is to take E2p7dem to establish representational parallelism in full. The reference to E5p1dem strikes me as unconvincing.⁵¹ The fact that we can, by taking E1a4 to imply *Conception* \rightarrow *Causation*, use E2p7dem to establish representational parallelism in full, constitutes evidence in favor of the standard reading of that axiom.⁵²

⁵⁰ Because Spinoza takes ideas of ideas to be only conceptually distinct from their first-order ideas (E2p21s), this infinite progression of ideas does not conflict with his metaphysical parallelism, which arguably requires that there be only one mode of thought corresponding to its infinite corresponding modes in the other attributes.

⁵¹ In E5p1dem, Spinoza writes:

The order and connection of ideas is the same as the order and connection of things (by E2p7), and vice versa, the order and connection of things is the same as the order and connection of ideas (by E2p6c and E2p7).

So he cites E2p7 when stating the thing-to-idea half of representational parallelism, and E2p6c and E2p7 when stating the idea-to-thing half. Melamed and Morrison take this to support a reading on which Spinoza saw these two halves as different doctrines and thought that E2p7 was insufficient for establishing both of them. However, if he truly believed that, and thought that E2p6c by itself *was* sufficient for deducing the second doctrine, then he would not have cited E2p7 twice. It is more likely that he understands E2p6c to independently state the idea-to-thing half – which I grant it does – but also thinks that E2p7 is sufficient by itself to establish both halves of representational parallelism. He cites both to indicate that he has an additional proof in E2p6c, not to indicate that E2p7 is insufficient.

⁵² However, this evidence is not decisive so long as some of the other challenges by Morrison go unanswered. This is not the place for doing so.

5. E2p7c: *God's powers and three identities*

E2p7c: From this it follows that God's [NS:⁵³ actual] power of thinking is equal to his actual power of acting. I.e., whatever follows formally from God's infinite nature follows objectively in God from his idea in the same order and with the same connection.

That E2p7e is a statement of representational and not metaphysical parallelism becomes even more clear from the corollary that Spinoza deduces from it. In this corollary, Spinoza uses a distinction between formal and objective being to discuss the intentional relation between ideas and their objects. He states that God's power of *representing* things (his "power of thinking") is "equal" (*aequalis*) to his power of *producing* them (his "power of acting") and, by saying that whatever follows from the one power also follows from the other, makes the surprising claim that (adequate) ideas are identical to the things they represent.

Before discussing these claims, it will be helpful to recall the basics of the formal-objective distinction that Spinoza employs here.⁵⁴ Following Descartes, Spinoza takes the "formal being" of a thing to be its existence as a thing, independently of being thought about, whereas its "objective being" is its existence as an object of thought.⁵⁵ This means that a body's objective being is its being as the content of an idea of that body. In a well-known example, Descartes writes that

the idea of the sun is the sun itself existing in the intellect – not of course formally existing, as it does in the heavens, but objectively existing, i.e. in the way in which objects normally are in the intellect. (*First Replies*; CSM II, 75; AT VII, 102)

Contrary to current-day use, then, the "objective being" of a thing is its existence as the content of an idea, and not its existence outside of the mind. Similarly, an idea's objective being is its existence as the content of a second-order idea of that idea. Spinoza's invocation of this

⁵³ NS = *Nagelate Schriften*, the Dutch translation of Spinoza's *Opera Posthuma* in which the *Ethics* first appeared (both 1677). Additions marked NS thus indicate differences between the Dutch and Latin editions of the posthumous works. They are cited from Curley's translation (Spinoza 1985-2016).

⁵⁴ See Ayers (1998) for a discussion of the formal-objective distinction and an overview of some early modern theories of intentionality, including Spinoza's. See also Garrett (2008, 2018c) and Hübner (2019, 2022) for recent discussions of Spinoza's theory of intentionality.

⁵⁵ In the early *Treatise on the Emendation of the Intellect*, Spinoza speaks of formal and objective "essences": see esp. TIE, 33-42. At TIE, 41, Spinoza expresses an early version of representational parallelism as follows: "the idea is objectively in the same way as its object is really." In context, what he is stressing is that the idea stands in the same causal and conceptual relations as its object, i.e. is parallel to it.

terminology in E2p7c – as well as in the runup to E2p7, in E2p5, E2p6 and E2p6c – clearly indicate that he is talking about the representational relation of ideas to their objects here.

Three questions now arise about the relation of objective and formal being in connection to ideas and things:

- (1) Is x , considered objectively, one and the same thing as x , considered formally?
- (2) Is x , considered objectively, one and the same thing as $I(x)$, considered formally?
- (3) Is x , considered formally, one and the same thing as $I(x)$, considered formally?

I think that, surprisingly, Spinoza's answer to all three questions is 'yes'.⁵⁶ I therefore agree with Ayers (1998, 1077) that "Spinoza attempts [...] a coherent theory according to which the real object, the intentional object, and the mode of thought are all three truly identical with one another, distinguished only by distinctions of reason." When Ayers invokes a "distinction of reason" here, I take him to refer to the same distinction that holds between modes of different attributes. At any rate this is how I will understand the distinction. In other words, just as (formally) a mode of extension and a mode of thought can be "one and the same thing, conceived in two ways" (see discussion of E2p7s below), so also an (adequate) idea and the thing it represents, or a thing and its (adequate) representation in thought, are "one and the same thing, conceived in two ways." I take Spinoza to answer all three questions in E2p7c. In E2p7s, he then develops an additional way of answering the third.

To see how Spinoza can argue for these identity claims, we should first consider his account of God's powers of thinking and acting.⁵⁷ The wording of the corollary suggests that God's power of acting should be connected to what "follows formally from God's infinite nature," while his power of thinking relates to what "follows objectively in God from his idea." That suggests that in general, a thing's power of acting relates to what formally follows from that thing's essence, whereas a thing's power of thinking relates to what follows objectively from its idea (i.e., from what it represents). As one might expect, in the case of God both powers are absolutely infinite, and so everything that has formal or objective being falls under his powers. Without wanting to take a stance here on how these powers relate to traditional notions of

⁵⁶ Again, this puts significant restrictions on the scope of Spinoza's deductions in E2p7, because the idea of God (a mode) arguably cannot be identical to God himself, because that would imply that a mode were identical to a substance. For two contrary takes on this problem, see Melamed (2013b, 182-3) and Hübner (2022, 62-3).

⁵⁷ Spinoza characterizes God's powers of acting and (in the NS) thinking as "actual". I take him to do this merely to indicate that God does not have any unactualized power. See also Lærke (2017).

divine omniscience and omnipotence, I simply note that they show clear affinities with them, and so will use these terms as shorthand for the two powers:⁵⁸

Omnipotence: All things formally follow from God's essence.

Omniscience: All things objectively follow from God's idea.

Together, these principles state that for any thing x , God produces both its formal and its objective being. And this implies that x , considered objectively is one and the same thing as x , considered formally. This answers the first question above.

It also answers the second question, but only on the condition that $I(x)$, considered formally, is nothing over and above its content, which is the objective being of x .⁵⁹ That is, if ideas are individuated by their contents, then it makes sense to identify them with these contents. This is because Spinoza takes things to be individuated by their essences, and believes that things are nothing over and above their essences (E2def2). Now, plausibly, the content of $I(x)$ qualifies as its essence. The definition of essence at E2def2 states that if it is necessary and sufficient to "posit" or "conceive" x in order to "posit" or "conceive" some thing y , then x "belongs" to the essence of y . Now, when positing or conceiving the content of $I(x)$, this is sufficient for positing or conceiving of $I(x)$; it is also necessary, because it is hard to see how $I(x)$ could be posited or conceived in the absence of its object. For example, to form the idea of a triangle, it is sufficient to think of a triangle; whereas it is not possible to form this idea without thinking of a triangle. It is also difficult to see how there could be anything else that "belongs" to the essence of $I(x)$ apart from the representation of x . So the identification of x , objectively considered, and $I(x)$, formally considered, seems secure.

That leaves the third and most interesting question. Spinoza's affirmative answer to the first two questions implies his affirmation of the third as well. For if $I(x)$, considered formally, is one and the same thing as x , considered objectively, and that in turn is one and the same thing

⁵⁸ Spinoza himself refers to God's infinite power of acting as "omnipotence" in E1p17s (G II, 62). The term "omniscience" does not occur in the *Ethics*, but it does occur in KV I, 7 and CM II, 7.

⁵⁹ There is a technical complication here which I will ignore, which is that if Melamed (2013b, ch. 6) is right that one and the same idea represents its corresponding mode under any attribute, then $I(x)$ will also contain the objective being of x' , x'' , etc. (where x' and x'' are the third- and fourth-attribute mode corresponding to x , resp.).

as x , considered formally, then by transitivity it follows that $I(x)$ is one and the same thing as x .⁶⁰

Notably, this identity claim is an instance of Spinoza's well-known "mode identity" thesis – his view that corresponding modes of different attributes are "one and the same thing" – just restricted to identities between an idea and the mode it represents. That Spinoza is committed to mode identity is widely accepted on the basis of evidence from E2p7s.⁶¹ The discussion so far, however, has shown that Spinoza is already in a position to argue for part of mode identity on the basis of his views about representation, because representational parallelism also supports the identities of things and ideas.

Naturally, the theory of representation indicated in E2p7c raises a number of big questions about – to name just a few – how Spinoza can account for truth and falsity, for distinct ideas with the same objects, and for reference to nonexistents. These cannot be discussed here. E2p7c itself also raises further questions, particularly about how God's powers of thinking and acting interrelate and what Spinoza means by saying they are "equal." For reasons of space, these questions must also be deferred.⁶² What I hope to have shown in this section, however, is that there is good textual support for ascribing the three identity claims to Spinoza. This support derives from E2p7c's crucial assertion that "whatever" is produced by one of God's powers is also produced by the other.⁶³

⁶⁰ Garrett (2017) argues that Spinoza does not accept the transitivity of identity. But his worries focus on the distinction between the attributes: he ascribes to Spinoza the view that Thought = God, and Extension = God, yet Thought \neq Extension. I personally do not think that this is Spinoza's view; but at any rate, even if Spinoza rejects transitivity in relation to the attributes, I don't see any reason for thinking that he rejects it in relation to modes. So if mode of extension e = mode of thought m , and m = a mode of the third attribute t , then I see no reason why Spinoza would deny that $e = t$.

⁶¹ However, two dissenters to this consensus are Marshall (2009) and Renz (2018).

⁶² For discussion of the relation between God's powers, see Deleuze (1990, 113-28), Gueroult (1974, 74-8), Garrett (2008, 14-5) and Koistinen (2018). For the notion of "equality," see esp. Jaquet (2018).

⁶³ From the vantage point of E2p7c, it is clear that E1p16 had already stated both *Omnipotence* (in its main clause) and *Omniscience* (in the bracketed clause): "From the necessity of the divine nature there must follow infinitely many things in infinitely many modes (i.e., everything which can fall under an infinite intellect)." However, E1p16 does not yet designate these as powers, nor does it clearly distinguish them. More substantively, E2p7c adds that *one and the same thing* can be considered under both powers: "whatever" follows from the one also follows from the other.

6. E2p7s: *Metaphysical parallelism*

Both the enunciation and the corollary of E2p7 state views that relate to God's power of thinking: his ability to represent, or have an idea of, everything he produces. It is only in the scholium that Spinoza sets out the view that is standardly taken to be his (only) parallelist view. This view relates to the order and connection whereby modes in each of the attributes are (formally) produced. I will call it *metaphysical parallelism*.⁶⁴ It is the view that the modes in each attribute are parallel to the modes of any other attribute.

Spinoza's argument for this view is once again brief and in need of elaboration. He begins E2p7s by recalling his substance monism, according to which there is only a single substance with infinitely many attributes (E1def6, E1p11, E1p14), and his view that each attribute is "what the intellect perceives of a substance, as constituting its essence" (E1def4) and must be "conceived through itself" (E1p10). He derives from this the idea that the different attributes correspond to independent ways of conceiving the single substance:

the thinking substance and the extended substance are one and the same substance, which is now comprehended under this attribute, now under that.

Then he immediately generalizes and states that the same must be true for the modes of this substance. Spinoza's use of "So also" (*Sic etiam*) suggests that he takes the two arguments to be completely analogous:

So also a mode of extension and the idea of that mode are one and the same thing, but expressed in two ways.⁶⁵

This claim is a version of Spinoza's mode identity thesis. It is the same view that E2p7c partially establishes. But in the present context, Spinoza deduces it from substance monism and his views about attributes, instead of from any concerns about representation. In other words, that the identity here holds between a mode and the *idea* of that mode is no longer important – what matters is the identity of corresponding modes under different attributes generally. This is

⁶⁴ Melamed (2013b) calls it "inter-attribute" parallelism; Deleuze (1990, 113) calls it "ontological" parallelism. Note that, by calling this view "metaphysical," I don't want to deny that the representational parallelism is also in part a metaphysical doctrine.

⁶⁵ Spinoza uses the concept of expression in this sentence, but later in the scholium he describes the same relation in terms of conception. In E2p21s and E3p2s (G II, 141) he makes the same point as the quoted sentence in terms of the way the mode is conceived. See Marshall (2009, 900-2) for an inventory of these different claims.

confirmed by the way Spinoza goes on to generalize this idea-object version of the mode identity thesis to the unrestricted version of mode identity:

For example, a circle existing in nature and the idea of the existing circle, which is also in God, are one and the same thing, which is explained through different attributes. Therefore, whether we conceive nature under the attribute of Extension, or under the attribute of Thought, or under any other attribute, we shall find one and the same order, *or* one and the same connection of causes, i.e., that the same things follow one another.

The question now is how Spinoza can argue from what may be called his *substance identity* thesis to the mode identity thesis:

Substance identity: The substance with attribute A and the substance with attribute B are one and the same substance.

Mode identity: Mode *a* of attribute A and the corresponding mode *b* of attribute B are one and the same thing.

Although there are several ways of spelling out this connection, the one I will present here centers on the way modes relate to their attributes.⁶⁶ As their name implies, modes are ‘modifications’ of attributes: they are, as Spinoza sometimes puts it, a ‘certain and determinate expression of God’s nature’ (e.g., E2p10cdem). So a mode of extension is a certain definite modification of extension. The essences of these modes follow from the attributes whose modifications they are. Now, by *Substance identity*, each attribute belongs to one and the same substance. At the same time, by E1def4, each attribute constitutes the essence of substance.⁶⁷ Each attribute is an expression of God’s single and indivisible nature. Therefore, any other attribute, while being a different expression of God’s nature, expresses itself in an exactly analogous way. Everything that follows from God’s nature under one attribute follows from it under any attribute. So any other attribute will also instantiate the same modification that produced the mode of extension. As the two modes are produced by the same modification of the same substance – albeit ‘expressed’ in two different attributes – they are one and the same thing, expressed in two ways.

⁶⁶ For alternative proposals, see Curley (1988, 67-70), Della Rocca (1996, 129-40), Melamed (2013b, 148-51) and Lin (2019, 88). My presentation of the argument is close to that of Nadler (2006, 128-9).

⁶⁷ This is so assuming that the ‘subjectivist’ interpretation of the attributes, associated with Wolfson (1934), is rejected. For discussion, see Gueroult (1968, 428-61), Shein (2009) and Melamed (2018, 94-95).

This way of spelling out the argument makes clear how the mode identity thesis and metaphysical parallelism are related. As Spinoza points out in the last quoted passage from E2p7s, the attributes correspond to different ways of “conceiving” or “explaining” the single order of nature. This order of nature is also described by him as a *causal* order: an “order, *or* [...] connection of causes.” The modes of each attribute stand in one and the same causal order. So they are causally (and, via *Causation* \rightarrow *Conception*, conceptually) parallel.

There are two further things to say about the scholium. First, it is sometimes thought that the mode identity thesis implies metaphysical parallelism more or less immediately, via an invocation of Leibniz’s Law.⁶⁸ The idea is that if a mode *a* of attribute A causes an effect *e*, and *a* = a mode *b* of attribute B, then, because identical things share all their properties (including their causal properties), *b* causes *e*; this argument can then be extended to show that *a* and *b* are causally parallel. Regardless of whether Spinoza *could* argue this way, I think it is clear that this is not his argument in E2p7s. For Spinoza takes care to explain how it is that modes of different attributes stand in the same causal order. If he wanted to use Leibniz’s Law, he would not have added this explanation, because in that case the attribution of causal properties would be independent of the attributes. Instead, Spinoza appears to derive *both* mode identity *and* metaphysical parallelism from the claim that the attributes are different expressions of God’s nature. This claim implies both mode identity – because modes are produced by modifications of God’s nature, expressed in each attribute – and metaphysical parallelism – because there is only one order of nature, expressed in each attribute. Therefore, it appears that if anything, it is Spinoza’s view of the attributes as expressions that is at the basis of metaphysical parallelism, and not his mode identity thesis. If the latter implies metaphysical parallelism, it is because being “one and the same mode” for Spinoza means being produced in the same way regardless of attribute.

The second issue that arises, now that we have traced the arguments for both representational and metaphysical parallelism, is what relation there is between these views. Melamed has claimed that “there is hardly any overlap in the justifications of the two [parallelism] doctrines, and [...] Spinoza does not rely on either doctrine in proving the other” (2013b, 146). As the discussion so far has shown, I agree. Representational parallelism depends on Spinoza’s views about conception, causation, representation, and the ways these are interrelated. Metaphysical parallelism instead depends on his substance monism and his views about the attributes. At the

⁶⁸ See esp. several recent writings by Lin (2019, 88; 2020, 226). Cf. Marshall (2009, 903-4).

same time, the distinction between the two views should not be overstated. Melamed argues that representational parallelism has no bearing on mode identity at all and that metaphysical parallelism relates only to the identity of modes under different attributes. He also claims that E1p16 plays a role in the derivation of metaphysical but not of representational parallelism (2013b, 146n16, 150-1). I disagree with both of these claims: E2p7 lays the basis for E2p7c's limited mode identity thesis, and E1p16, in foreshadowing E2p7c, is in the background of representational parallelism. Furthermore, it is noteworthy that E2p7s keeps shifting back between claims about representation and claims about modes under different attributes. So, after the statements of substance identity and mode identity quoted above, Spinoza reverts to considering the relation between "God's intellect" and what he "understands" – a claim that connects to E2p7c's statement concerning God's powers. He goes on to make claims about how nature must be "perceived" and "explained." The impression created is that, while it is true that representational and metaphysical parallelism are different and perhaps independent doctrines, Spinoza nevertheless is interested in relating them to each other.

7. E2p13: *Mind-body parallelism*

Let me end by pointing out an important advantage of clearly distinguishing representational from metaphysical parallelism. This is that it allows us to rethink E2p7's relation to mind-body parallelism.

Mind-body parallelism is often defined simply as the view that modes of thought are parallel to modes of extension. As we now see, Spinoza can deduce this view both from the representational parallelism of E2p7e and from the metaphysical parallelism of E2p7s. The thought behind taking it to be identical to mind-body parallelism is probably that human minds, for Spinoza, are ideas that represent bodies. And that is true. But there are two gaps that this interpretation leaves open. First, it assumes that mind-body parallelism is an instance of metaphysical parallelism, and so it does not strictly involve the idea that minds are *representations* of bodies.⁶⁹ At most it is the idea that minds and bodies stand in the same causal and conceptual relations, in line with the general form of parallelism. It takes the relation between minds and bodies to be exactly the same as that between bodies and modes of the third attribute. But

⁶⁹ E.g., Peterman (2017, 224) describes these as separate views. Della Rocca (2008b, 316) also does not include representation in his glossary entry on parallelism. See also Bennett (1984, 127).

Spinoza is clear that minds are defined at least partly in terms of what they represent (E2p13s, Ep32, Ep66). A bare, metaphysical parallelism cannot account for this feature of minds.

Second, Spinoza only establishes that the mind represents the body after E2p7, in E2p11-13. E2p11 establishes that the mind is an idea and that it has an actually existing object. E2p12 adds that this idea represents anything that happens in its object. E2p13 then finally establishes that “The object of the idea constituting the human mind is the body, *or* a certain mode of Extension which actually exists, and nothing else.” This means that it is only in E2p13 that Spinoza establishes that his ideas-bodies parallelism is in fact a *mind*-body parallelism. By his lights, E2p7 states views about the relation of *ideas* to bodies, but not of *minds*.

It is worth being strict on this point, because doing so allows us to connect mind-body parallelism to Spinoza's account of the union between mind and body in a way that is tighter than is usually done.⁷⁰ This union, as E2p13s and especially E2p21dem make clear, consists in the fact that the mind has the body for its object. As Spinoza writes: “We have shown that the mind is united to the body from the fact that the body is the object of the mind (see E2p12 and 13)” (E2p21dem).⁷¹ It can be argued that Spinoza appeals to this relation of direct intentionality, and *not* to metaphysical parallelism, to explain many central psychological functions, such as sense perception (E2p16), imagination (E2p17s), memory (E2p17c), and the affects themselves (E3def3; the ‘general definition of the affects’ at G II, 203). In each of these cases, a mental phenomenon is explained by the fact that a certain change in the body is perceived – represented – by the mind. Now, by representational parallelism, the existence of these intentional relations entails that changes in the body will correspond to changes in the mind. But this suggests that this intentional relation is prior to, and explains, these causal correlations, and not vice versa. At the same time, however, this relation is incorporated into Spinoza's account of the union through representational parallelism, which in itself explains this correlation between changes in ideas and in their objects at a more fundamental level of Spinoza's metaphysics. This shows that mind-body parallelism, therefore – provided we read it as incorporating

⁷⁰ On the union, see Curley (1988, 74-8), Hübner (2019, 6). Besides presenting her own interpretation of the mind-body union in Spinoza, Jaquet has criticized the ability of parallelism to account for it in several works (see Jaquet 2018, 12-9; 2021).

⁷¹ As far as I know, Spinoza never appeals to the *identity* of mind and body to explain the union. Even if he did, this identity could be argued on the basis of both representational and metaphysical parallelism.

this representational claim – is central to Spinoza's account of the mind-body union, while a bare (i.e., nonrepresentational) metaphysical parallelism is not.⁷²

As this chapter has argued, therefore, representational parallelism is an ineradicable part of mind-body parallelism. From metaphysical parallelism, we can only deduce that ideas are parallel to bodies, but not that these ideas represent these bodies. That also means that we cannot, in strictness, derive mind-body parallelism from it; hence, bare metaphysical parallelism is unable by itself to explain the mind-body union.

8. Conclusion

This chapter has clarified what it means for things to be parallel for Spinoza: it is for them to stand in the same "order and connection." It has also reconstructed Spinoza's arguments in E2p7 for his various parallelist doctrines and clarified the connection of these doctrines to the mind-body parallelism of E2p13. It has stressed the role of considerations about representation in these arguments and in mind-body parallelism itself. The resulting picture of parallelism is complex, but it is also rich.

A final word about the status of E2p7 in the *Ethics* as a whole. Early in Part Two, Spinoza is coordinating a number of different distinctions in his system: the relation between causation and conception, between the idea and what it represents, and between modes in different attributes. E2p7 presents his explicit account of how these relations interconnect. Seen in this light, Spinoza's deductions in E2p7 are as interesting for the relations he does *not* establish as for those he *does*. On the one hand, he positions each mode in the causal, conceptual and representational relations that connect it to other modes, showing how these relations intersect through the causal axiom and through his view about God's two powers. At the same time, Spinoza resists making several easy reductions. Although he sees any conceptual relation among things as mirrored by a conceptual relation between the *ideas* of these things, Spinoza does not commit to a reduction of conceptual relations to relations in the attribute of thought. He also does not incorporate his account of conception into his theory of representation. And, while his metaphysical parallelism allows him to establish that ideas are causally parallel to bodies, he does not reduce the relation of representation to a relation of causal covariation. It is

⁷² So I disagree with Colin Marshall's claim that causal metaphysical parallelism can explain the union (Marshall 2009, 913-4).

therefore important to clearly distinguish these different relations and the roles that they play in Spinoza's arguments. The resulting system, which is overdetermined but ostensibly coherent, lays the foundation for all of Spinoza's further deductions in the *Ethics* in philosophy of mind, epistemology, psychology and ethics.

*Spinoza's Rejection of Interactionism**1. Why did Spinoza reject interactionism?*

Spinoza is famous for rejecting Cartesian interactionism and adopting substance monism in response to Descartes' dualism. It is received wisdom that several of his most characteristic metaphysical innovations are adopted at least partly in response to perceived problems in their Cartesian alternatives.⁷³ So, against Descartes' view that a substance can have only one principal attribute, Spinoza argues that there is no contradiction in the idea of a single substance's having several attributes (E1p10s). Against Descartes' substance dualism, he argues for an identity theory, according to which the mind and the body are "one and the same thing, but expressed in two ways" (E2p7s). And he rejects Descartes' notorious interactionism, saying that "The body cannot determine the mind to thinking, and the mind cannot determine the body to motion, to rest or to anything else (if there is anything else)" (E3p2).

If all of these positions are put together, however, it can seem as if Spinoza has committed himself to a set of positions that is implausibly strong. Spinoza's view of the mind is that it is identical to ("one and the same thing as") the body. Besides this, however, he also adopts what has come to be called a *causal barrier* between the attributes – the idea that there are no causal relations between modes of different attributes (E2p6; see Della Rocca 1996). (This causal barrier in turn rests on a deeper *conceptual barrier* that is based on the idea that each attribute must be "conceived through itself" (E1p10), precluding the existence of conceptual relations between different attributes.) The mind is a mode of the attribute of thought and the body a mode of the attribute of extension, and so the causal barrier precludes their interaction. The

⁷³ See Curley (1988) for a classic argument for strong Cartesian influences on Spinoza.

causal barrier motivates the denial of interactionism in E3p2dem, which cites E2p6. Spinoza therefore denies that a mind can cause a body to move. But this sits uneasily with his identity theory, according to which my mind is identical with my body, and bodies naturally *do* cause other bodies to move. Mind and body are identical, then, and both enter into causal relations with other identical mind-body pairs; but these causal relations are nevertheless such that there are no causal relations from a mind to a body, or vice versa.

It appears, then, that Spinoza accepts the premises but denies the conclusion of the following argument.⁷⁴

The mental causation argument:

- (1) Mode of extension p_1 causes mode of extension p_2 .
- (2) Mode of extension p_1 = mode of thought m_1 .
- (3) Therefore, m_1 causes p_2 .

Now, the mental causation argument is highly intuitively compelling. It appears to be a consequence of the Indiscernibility of Identicals (one half of Leibniz's Law) that if x causes y , and x is identical to z , then z causes y . Descriptions or ways of referring normally do not make a difference to whether something is a cause of something else (although we will come back to this). And even if there are exceptions to this rule, the rule nevertheless holds in most cases. Yet Spinoza's views seem to require it to be broken on a systematic basis.

We can bring out how problematic Spinoza's position is by placing it in the context of the contemporary mental causation debate. This modern debate arguably has a version of the mental causation argument at its heart. The modern discussion centers around the intelligibility of positions such as property dualism and nonreductive physicalism. While these positions are not identical, what they have in common is a denial of an identity theory – the view that mental causes are identical to physical causes – which is replaced with the view that mental causes supervene on physical causes.⁷⁵ In other words, if “mode of extension” is replaced by “physical

⁷⁴ Delahunty (1985, 197); see discussion in Della Rocca (1996, 121-5). Compare Jarrett (1991, 470) and Davidson (2005, 300-1). Note that throughout this chapter, by “mental causation” will be meant *psychophysical*, i.e., mind-body and body-mind causation. Mind-mind causation is sometimes classed as a form of mental causation as well, but it is clear from E2p6 that Spinoza does accept this form of mental causation.

⁷⁵ As Bennett (2008) construes these positions, physicalism involves the acceptance of a global supervenience claim with metaphysical necessity. Any position that accepts a weaker supervenience claim will, by this definition, not count as a form of physicalism. It is then open for a property dualist to deny the identity of the causally efficacious properties, but accept only a weaker (i.e., nomological) form of necessity for the supervenience claim (e.g., Chalmers (1996) fits this description). Naturally, other definitions of these positions are possible as well; for example, Árnadóttir and Crane (2013, 252) distinguish between nonreductive

cause” and “mode of thought” by “mental cause” in the above argument, then these “nonreductivists” deny (2). It is this denial of identity that raises the familiar questions about how mental causes can have genuine causal powers if they are dependent on physical causes that are already, by themselves, assumed to be sufficient for bringing about the effect. (I return to these in section 3.) By contrast, an identity theory can, in virtue of identifying physical and mental causes, bypass all these worries altogether. Whatever problems such an identity theory faces, they are not problems about mental causation. And this is precisely because the identity theorist can appeal to an argument along the lines of the mental causation argument to bolster their claim that mental causes can have physical effects. The effect of all this is that it becomes even more puzzling why Spinoza should adopt an identity theory, yet still deny mental causation.

This chapter is in two parts. In the first part (sections 2-3), I argue that so far, there has not been a satisfactory explanation for why Spinoza adopts his causal barrier. I review five responses to the problem that have emerged in the literature, several of which stress the role of the causal barrier, and argue that none of them can be considered satisfactory (section 2). I also underscore the strangeness of Spinoza's commitments by situating them in relation to the contemporary mental causation debate (section 3). This analysis suggests that what Spinoza objects to is not so much the bare existence of mental causation as a certain substance dualist theory of mental causation. Then, in the second part of the chapter (sections 4-5), I show that Spinoza's own discussions do not clearly distinguish between a rejection of mental causation *per se* and a rejection of such a dualist theory of mental causation, and that several of his views actually put pressure on him to accept the existence of mental causation. I also show that there are several places where Spinoza appeals to his identity theory, and not to the causal barrier, in order to attack interactionist views. I end by emphasizing that Spinoza's causal-conceptual barrier is not the only feature of his views motivating his rejection of mind-body interaction.

2. *Five responses*

physicalists and emergentists, where these latter accept the same global supervenience claim as the nonreductive physicalists, but deny the further claim that all the fundamental facts are physical facts. These distinctions do not matter in the context of this chapter, because it is interested only in positions that reject the identity claim while adopting some form of supervenience claim, which all these types of nonreductivists are universally held to do, however their precise positions are defined.

I will begin by reviewing five existing responses to this problem.

2.1 *Accommodating apparent interaction*

A first response argues that even if Spinoza cannot accommodate *actual* mental causation, he can at least explain why it *appears to us* as if there is mental causation. So, even though Spinoza does, on this approach, truly deny (3), the counterintuitiveness of this denial can be assuaged by pointing out that, even if Spinoza denies the literal truth of (3), he can at least account for why it *appears* to be true. Thus, Schmidt writes that

no modes of different attributes can causally interact with each other. What the identity theory can explain, however, is their *apparent* interaction in the case of mind and body. If it seems that a mode of thought A causes a mode of extension B, the mode of thought A that apparently causes the mode of extension B is actually identical with the mode of extension A*. (Schmidt 2009, 96-7; his italics)

In addition, Bennett (1984, 132) points out that Spinoza could account for the appearance of interaction by the fact that we are normally ignorant of some of the causes in the two causal chains. If I am stabbed, there is both a physical and a mental sequence in which I am stabbed, feel pain, and then react. These three steps are represented in both the attribute of thought and that of extension, but ordinarily I will be aware only of the physical aspect of the stabbing and the reaction, and only of the mental aspect of the pain. So, Bennett suggests, it is natural that I interpret the stabbing as an example of mind-body interaction.⁷⁶

This response has serious problems, however. While Bennett's psychological explanation for the appearance of interaction is interesting, it does little to resolve the central problem of explaining *why* (3) does not follow in the mental causation argument. Bennett thinks there is no explanation of this fact: the causal separation of the attributes comes down to an "undefended assumption" (1984, 49). Schmidt's position is more problematic, because he uses the identity theory to argue for the *appearance* of mental causation, when ordinarily it is held to secure the real thing. We are owed an explanation of why identity cannot establish real mental causation in this instance.

⁷⁶ Recently Lin (2019, 87; 2020, 221) has also suggested that parallelism can explain the causal correlations between mind and body.

In addition, the response faces a deeper worry. In E2p35s, Spinoza writes:

men are deceived in that they think themselves free [NS: i.e., they think that, of their own free will, they can either do a thing or forbear doing it], an opinion which consists only in this, that they are conscious of their actions and ignorant of the causes by which they are determined.

This states that the way interaction *appears* to occur to subjects will frequently not be trustworthy: it will pick out illusory causes. It doesn't appear to say something that would support Bennett and Schmidt's solution, that the appearance is untrustworthy only in picking out a mode of one attribute for a cause when the real cause is its corresponding mode in another attribute. In other words, E2p35s does not rule out a situation where, if it appears to me that mode of thought *m* caused my action, the real cause is *not* its corresponding mode of extension *p*, but some other mode of extension *p** not parallel to *m*. This shows that in Spinoza's case – unlike, incidentally, in that of Leibniz – the appearance of mental causation need not reflect an underlying causal relation. If that is right, Spinoza's parallelism or his mode identity doctrine cannot be exploited for an explanation of the phenomenology of mental causation in the way Bennett and Schmidt intend to exploit it.

2.2 *Dualist responses*

The second type of response argues that Spinoza is not in fact an identity theorist. In other words, the mental causation argument may be valid, but Spinoza denies (2). Donagan (1988) has defended this position, and more recently Marshall (2009) has argued that Spinoza should not be read as committed to numerical mind-body identity.⁷⁷ Such a view faces the difficult task of explaining how Spinoza could have said that mind and body are “one and the same thing” without thereby wanting to say that they are numerically identical. However, I will not go into this problem here. In this chapter I am interested in the way Spinoza's apparent adoption of a mind-body identity theory impacts his arguments against interactionism. A position that argues that Spinoza is not committed to an identity theory in the first place thereby places itself out of the scope of this question, regardless of its other attractions. I will proceed

⁷⁷ See Garrett (2017) for a response to Marshall.

throughout this chapter on the assumption that E2p7s and E3p2s express a mind-body identity theory (in which I follow the majority of commentators).

2.3 *Causal contexts as intensional*

The third response defends an intensionalist interpretation of Spinoza's theory of causation and argues that the mental causation argument is invalid.

Jarrett (1991) and Della Rocca (1996) suggest that Spinoza construes causal contexts as "referentially opaque": he holds that co-referring terms in an expression involving a causal predicate are not substitutable *salva veritate*. In other words, 'causes' functions much like 'believes'. To use a stock example, given that Benjamin Franklin, the inventor of bifocals, and the first postmaster general of the United States were all the same person, the following inference is nevertheless not valid:

- (4) X believes that Benjamin Franklin is the inventor of bifocals.
- (5) Benjamin Franklin = the first postmaster general of the United States.
- (6) Therefore, X believes that the first postmaster general of the United States is the inventor of bifocals.

This inference does not go through because there is no guarantee that X knows that (5) is true. This is explained by saying that 'believes' in (4) creates a referentially opaque context.

The Jarrett-Della Rocca suggestion is that the same goes for expressions of the form '*x* causes *y*.' If that is right, it would account for the failure of the mental causation argument, because this argument also involves the substitution of co-referring terms. With the attribution of such an "intensional" conception of causation, in contrast to the more familiar "extensional" conception, Spinoza would be able to give an account of his denial of mental causation.

There is some evidence that causal contexts may be intensional. The truth value of a sentence of the form '*x* causes *y*' in some cases does seem to depend on the way *x* or *y* is described. Consider the following example (adapted from Davidson 1963):

- (7) My act of turning on the light alerted the burglar.
- (8) My act of turning on the light = my act of flipping the switch.
- (9) Therefore, my act of flipping the switch alerted the burglar.

The conclusion might be said to be false, because it seems to imply that the burglar was alerted by the flipping of the switch itself, rather than by the light coming on. So even though they refer to identical events, 'my flipping the switch alerted the burglar' and 'my turning on the light alerted the burglar' have different truth values. This can be explained analogously to the previous example, by saying that 'alerted' in (7) is a causal verb and so creates an opaque context.⁷⁸

There is also some evidence that Spinoza may have held an intensional theory of causation. Della Rocca (1996, 122-3) points to Spinoza's use of "insofar as" (*quatenus*) in E2p6. This proposition states (my italics): "The modes of each attribute have God for their cause only *insofar as* he is considered under the attribute of which they are modes, and not *insofar as* he is considered under any other attribute." So while a mode of thought may be numerically identical to a mode of extension, it is still true that it is caused by God only *insofar as* he is considered as a thinking substance, and not insofar as he is considered as an extended substance.

While E2p6 only covers the relation of immanent causation whereby God causes modes, something similar holds for transient, inter-modal causal relations. Della Rocca (1996, 123-4) reconstructs a Spinozist argument for this claim on the basis of E1a4, but Spinoza himself is clear enough in E3p2dem:

All modes of thinking have God for a cause, insofar as he is a thinking thing, and not insofar as he is explained by another attribute (by E2p6). So what determines the mind to thinking is a mode of thinking and not of extension, i.e. (by E2def1), it is not the body.

The demonstration goes on to give a similar argument for modes of extension.

Della Rocca argues that this shows that the truth value of causal statements in Spinoza is relative to the way the cause is described, and hence that such descriptions are referentially opaque.

This interpretation, then, can account for the invalidity of the mental causation argument, has some intrinsic plausibility, and is compatible with some of Spinoza's claims about causation. For these reasons it has become influential. Yet it also suffers from some problems.

The chief among these is that, while the interpretation makes good sense of Spinoza's denial of (3) in the context of his other commitments, it does not motivate this denial in its own right.

⁷⁸ See also Jarrett (1991, 472; 1982). Other defenders of intensionalist theories of causation include Achinstein (1975) and Anscombe (2005).

Simply put, accepting the validity of the mental causation argument is massively more intuitive and less committing than the Jarrett-Della Rocca solution of subscribing to an intensionalist theory of causation. It is just not clear that the arguments in its favor are strong enough to make up for its continued denial of (3). The position does not give us an independent argument why it is not possible to say that mental events ever stand in a causal relation to physical events, in the face of our apparent everyday experience that they do so (cf. Marshall 2009, 907). To be sure, a Spinozist might dispute this experience, but that requires something more than just adopting an intensionalist theory of causation: it requires an argument that shows why this theory is correct. In fact, Della Rocca appears to agree with this assessment, given that he repeats Bennett's conclusion that the causal barrier between the attributes that is stated in E2p6 and that drives Spinoza's adoption of an intensionalist theory of causation amounts to no more than an *assumption* made by Spinoza, the reasons for which he ultimately considers "not good" (Della Rocca 1996, 15).

It is also questionable to what extent the evidence I just pointed out for an intensional theory of causation is really available for Spinoza to appeal to. For one, his view extends the referential opacity of causal contexts to *all* causal contexts involving minds and bodies. In spite of the limited evidence for intensional causal relations pointed out above, it seems clear that there are also many causal relations which *are* transparent. In this way, too, the interpretation appears to be overly strong. Furthermore, if all causation is intensional according to Spinoza, it appears that questions about what caused an event *as it is in itself* become unanswerable (Koistinen 1996, 32). Even in the light flipping switch case above, it is natural to assume that there is an underlying causal order of events that produces the effect, and which the two descriptions refer to. But the Jarrett-Della Rocca reading makes no place for such an underlying order of events, because it treats causation only at the level of causal descriptions. (We will return to this in the next subsection.) Apart from being unattractive in its own right, this also does not appear to do justice to Spinoza's own language when talking about causation. For example, the quotation from E3p2dem that was given above speaks, in very quick succession, of causal relations as holding "insofar as [God] *is* a thinking thing," "insofar as he is *explained* by another attribute," and then describes them as relations of "determination". The fact that Spinoza moves so

quickly from ontic to epistemic terms when describing causation arguably shows that he does not take himself to be speaking only at the level of causal descriptions.⁷⁹

2.4 Distinguishing causation and causal explanation

Dissatisfaction with an exclusively intensionalist interpretation of Spinoza's views on causation drives the next response. This response is close to the previous one in that it sees Spinoza's conception of causation as closely linked to explanation; but it argues that Spinoza also has, or is in a position to develop, an extensional concept of causation. It argues that 'cause' in the mental causation argument has to be read in this second, extensional sense, so that the argument comes out valid and (3) can be accepted by Spinoza. Different versions of this response have been defended by Davidson (2005), Koistinen (1996), and more recently Lin (2019, 88-91; 2020). Of course, if Spinoza does make this distinction, then that would put him quite close to contemporary writers on causation, and in particular to Davidson himself, who relies on it to formulate his anomalous monism (Davidson 1970).

One way to articulate the response is to say that the mental causation argument equivocates between two different notions of causation. Strictly speaking, the argument should be formulated like this:

- (1') Mode of extension p_1 (intensionally and extensionally) causes mode of extension p_2 .
- (2) Mode of extension p_1 = mode of thought m_1 .
- (3') Therefore, m_1 (extensionally) causes p_2 .

(1') can be read using either concept of causation, because the intensional causal relation that Spinoza explicitly accepts implies an extensional relation among their relata irrespective of what attribute they are conceived under. Because (3') similarly applies to such a purely extensional relation, Spinoza is in a position to accept it, even if he would reject its intensional variant, (3''):

- (3'') Therefore, m_1 (intensionally) causes p_2 .

Defenders of this response take Spinoza's denial of (3'') to be a different way to formulate the anti-interactionism put forth in E3p2: it is a rejection of the idea that mind-body causal relations are *explanatory*, not a rejection of their *existence*. As the distinction between extensional and

⁷⁹ Marshall (2009) and Garrett (2017) make the stronger claim that the notion of intensional causation is incoherent and so should not be ascribed to Spinoza.

intensional notions of causation is supposed to underwrite a distinction between causation and causal explanation, this argument concludes that there may be causal relations between modes of thought and modes of extension, but that causal explanations are still restricted to modes of the same attribute. As Davidson puts it:

The point of E3p2 is not, then, to deny that mental events can cause physical events, but to deny that they can explain them (and conversely, of course). Nothing in this picture of the relations between mind and body, the mental and the physical, rules out what we would call the causal interaction of particular physical events with particular mental events. We therefore do not have to saddle Spinoza with the logical absurdity that would result from holding that the physical event of a bell ringing cannot cause a mental awareness of the ringing even though that mental awareness is identical with a physical event in the brain caused by the ringing. (Davidson 2005, 306)

What is the evidence that Spinoza distinguishes between these two notions of causation? We have already seen that Jarrett and Della Rocca read the important propositions E2p6 and E3p2 as advancing the idea that all causation is intensional for Spinoza. By contrast, evidence that Spinoza also recognizes an extensional order of causation is much more rare. Such an order would need to exist independently from the way in which it is conceived. Both Koistinen (1996, 37) and Lin (2019, 91) point to a remark at the end of E2p7s as a possible reference to such an extensional order of causes:

So of things as they are in themselves, God is really the cause insofar as he consists of infinite attributes. [*Quare rerum, ut in se sunt, Deus revera est causa, quatenus infinitis constat attributis*]

This is one of the more obscure sentences of the *Ethics*. How does it support the idea that there is an extensional form of causation? Lin gives the following gloss:

Although this text is hardly pellucid, I take it to be Spinoza's (somewhat clumsy) way of saying that things, apart from how they are thought about, stand in causal relations to God independently of how he is conceived. This suggests that, in this context, *cause* is not expressing a relation that depends on how its *relata* are conceived. Thus, although it is not center stage, Spinoza has a notion of causation that is objective and mind-independent. (Lin 2019, 91; his italics)

This interpretation – similar to the one by Koistinen (1996, 37) – depends on paraphrasing “insofar as [God] consists of infinite attributes” as “independently of what attribute he is conceived under.” “Things as they are in themselves” similarly has to be interpreted as “things, independently of how they are conceived.” E2p7s, then, on this interpretation, asserts there to be causal relations between God and the modes independently of the attribute under which either of the two is conceived.

Unfortunately, however, it is not clear that this interpretation of the sentence is correct. Both of the proposed glosses are contestable. First, let's consider “insofar as [God] consists of infinite attributes”. Remember that “insofar as” is taken by Jarrett and Della Rocca to signal an opaque context. Its occurrence in E2p6 is an important piece of evidence for their claim that Spinoza holds an intensional theory of causation. Now, Koistinen and Lin also take E2p6 to set out such a theory; they just think that Spinoza recognizes an extensional form of causation as well. But in E2p7s, they take “insofar as” to indicate precisely the opposite of what it does in E2p6, namely a transparent context. This saddles Spinoza with severe terminological inconsistency. Also an arguably unnecessary one, as the standard intensional reading of “insofar as” makes sense when applied to E2p7s. On this reading, Spinoza is saying that God is the cause of things under every possible description, not that he is its cause regardless of any description. Provided we accept the theory of intensional causes, this seems a plausible and even a natural reading, and it avoids the terminological inconsistency.

Similarly, I think the Koistinen-Lin reading of “as they are in themselves” is unconvincing. Elsewhere in the *Ethics* Spinoza uses the same expression in a way that cannot be reconciled with their interpretation. E2p44 and its demonstration read as follows (my italics):

E2p44: It is of the nature of reason to regard things as necessary, not as contingent.

Dem.: It is of the nature of reason to perceive things truly (by E2p41), viz. (by E1a6) *as they are in themselves* [*ut in se sunt*], i.e. (by E1p29), not as contingent but as necessary.

Taking perception here to be a form of conception, what the demonstration says is that reason can conceive things as they are in themselves, and given that reason among other things conceives modes of specific attributes, this implies that “things in themselves” in this instance include these modes. Taking “as they are in themselves” to mean “independently of what attribute they are conceived under” here would have the seemingly absurd consequence that

reason could never adequately conceive modes of specific attributes, such as a certain body, *as* a body. Even God could not form such ideas. But this seems unacceptable.

Note that in E2p44dem, Spinoza cites E1a6 in connection to the phrase “as they are in themselves”. E1a6 states: “A true idea must agree with its object.” So to conceive a thing “as it is in itself” is to conceive it truly, and to conceive it truly is to form an idea of it that “agrees” with the actual thing. This suggests that when Spinoza speaks of things as they are in themselves, he is really stressing that true ideas represent their objects truly. (Naturally, some of these objects are modes of particular attributes.) This reading makes sense when extended to the sentence from E2p7s. On this reading, what the sentence says is just that God causes things under each of his infinite attributes. This does not entail that the things in question have some attribute-independent way of existing.

Note that claims about representation and the relation between ideas and objects play a big role in E2p7's demonstration and corollary, so it is not strange that Spinoza would use the “as it is in itself” locution in the scholium.⁸⁰ To employ terms used earlier in E2p7 and the scholium, Spinoza uses the phrase “as it is in itself” to refer to a thing's *formal* being, as opposed to its *objective* being (its being as the content of a representation). He is stressing that the actual things themselves, whose formal being is produced by God in infinite attributes, corresponds to the adequate ideas of these things. He is *not* opposing “being in itself” to “being in a specific attribute,” as Lin and Koistinen think. All in all, then, the E2p7s sentence does not seem to support the idea that Spinoza recognizes an extensional form of causation.

Apart from E2p7s, there is one other text that these authors appeal to. Koistinen (1996, 34-5) refers to a passage in E4pref where Spinoza describes someone's desire to build a house as the “efficient cause” of that house's being built (at G II, 207). This appears to be a genuine example of psychophysical causation in Spinoza and as such it is a fascinating find. However, it also illustrates a deep problem with the view that causation and causal explanation are separated by Spinoza. For, clearly, the desire to build the house here very much functions as an explanation for why the house was built. If causal explanation is entirely confined to intra-attribute relations, as Koistinen wants to hold, then this is an example of a place where Spinoza contradicts himself – or perhaps is speaking loosely instead of from within the context of his broader systematic

⁸⁰ See Chapter 2 for more on the role of representation in E2p7 as a whole.

views about causation – but not an example of him using a (merely) extensional notion of causation.

In fact, so long as this separation between intra-attribute causal explanations and cross-attribute extensional causation is held to, no case of psychophysical causation will ever be explanatory for Spinoza. But an important reason to want mental causation is precisely because such relations *are* (at least sometimes) explanatory. To return to Davidson's example, it does not seem unproblematic to say that the bell's ringing does not just *cause*, but actually plays a part in *explaining* my mental awareness of the bell. To that extent, this interpretation does justice to the letter of (3), but not to its spirit.⁸¹

2.5 Causal pluralism

The final proposal to be considered is Don Garrett's suggestion that Spinoza is a "strong ontological pluralist" about the attributes and that this is what motivates his rejection of (3) (Garrett 2017).

Strong ontological pluralism about the attributes (SOPA) is the view that each attribute is a distinct manner of existence. So it is the idea that extension, for example, is a fundamentally different manner of existing from thought. On this view, God exists in infinitely many ways because he possesses every attribute, and humans exist at least as extended and thought. Crucially, Garrett takes strong ontological pluralism to entail that existence as such can only be understood through any of the fundamental manners of existence. This means that the concept of existence in its attribute-neutral sense – in the sense in which both thought and extension can equally be said to exist – is derivative from a more fundamental, attribute-relative sense.

As he puts it:

"Existence" is an attribute-neutral term, but it can only be understood by understanding one or more particular attributes – such as Extension or Thought – that are not explicitly

⁸¹ Davidson recognizes this, and tries to establish a distinction between two sorts of explanation, one of which is nomological and holds between modes in the same attribute, and the other one of which is not nomological and can hold between modes in different attributes. Noting that this view is close to his own anomalous monism (2005, 308), Davidson tries to defend it as an interpretation of Spinoza. However, it is not at all clear that Spinoza recognizes any type of explanation that is not nomological (as, e.g., the earlier citation of E2p44 makes clear).

mentioned. In this sense, the concept of existence is, we might say, “disjunctively derivative” from the concepts of particular attributes. (Garrett 2017, 29)

This proposal connects to the mental causation argument through Garrett’s claim that ‘cause’ is also a “disjunctively derivative” term for Spinoza. That is, it too can only be understood through attribute-specific concepts of causation. I’ll indicate such concepts by a subscript E for extension and T for thought. Then, ‘ m_1 causes_T m_2 ’ and ‘ p_1 causes_E p_2 ’ state true propositions. However, (3) ‘ m_1 causes p_2 ’ does not (nor do its substitutions with either cause_T or cause_E).

Unfortunately, this suggestion suffers from problems when it is applied to the case at hand, as well as from textual problems. To begin with how it deals with the mental causation argument: here it faces the same problem as the Jarrett-Della Rocca response, that it is only successful if the doctrine on which it rests can be given an adequate defense. Now, Garrett’s concerns in his paper are (understandably) interpretive: he advocates SOPA as a way of making sense of Spinoza’s views. The present concern, however, is whether Spinoza has any arguments that support his rejection of (3); the fact that (3) is incompatible with SOPA, even if he did accept SOPA, is of little help with this in the absence of arguments for SOPA itself. It is not intuitive to think that causation by the mind is a different ‘manner’ of causation altogether than causation by bodies, and so arguments to defend this are needed.

Garrett’s interpretation also suffers from textual problems. For one, Spinoza never says or suggests that he takes attributes to be different manners of existence. What is more, the idea that ‘cause’ is disjunctively derivative from the concepts of causation in specific attributes is in apparent conflict with Spinoza’s conceptual procedure in the *Ethics*, which starts instead from causal relations between substances and only later comes to consider causation under specific attributes. Especially noteworthy here is E1def1, the opening definition of the work, which defines *causa sui* in a way that does not make reference to attributes. If this notion was disjunctively derivative, we would have expected Spinoza to begin by defining an attribute-specific notion of causation, or at any rate to introduce the notion of attribute before introducing the notion of causation; but he does neither of these things. (In fact, he does not explicitly prove that God has the attributes of thought and extension until the opening propositions of Part Two.)

Garrett's reading also appears to be in conflict with E1p10dem. E1p10 states that "Each attribute of a substance must be conceived through itself" – that is to say, the concept of any attribute does not depend on any concept other than itself. Its demonstration reads as follows:

For an attribute is what the intellect perceives concerning a substance, as constituting its essence (by E1def4); so (by E1def3) it must be conceived through itself, q.e.d.

The demonstration proves that attributes are conceived through themselves by appealing to the definition of substance, E1def3, which states that substances are conceived through themselves. It is because attributes "constitute the essence" of substance, and the essences of substances are conceived through themselves, that attributes are also conceived through themselves. Here, then, Spinoza derives one of the essential characteristics of attributes from substance, instead of the other way around. This shows that 'being conceived through', at any rate, is not disjunctively derivative. Now, Garrett does not claim that it is. But he does claim that 'inherence' is disjunctively derivative, and 'being conceived through' is arguably an inherence relation. Moreover, it is difficult to see what would prevent Spinoza from running an analogous argument to prove that every attribute is *caused* by itself.

In light of these textual problems, it is not clear that SOPA does motivate Spinoza's rejection of (3); even if it did, it is not clear what in turn motivates Spinoza to accept SOPA.

3. The minimal exclusion problem and Spinoza

The discussion so far has left us at an impasse: various ways of trying to account for the causal barrier between the attributes have come up short. Spinoza's adoption of this causal barrier, while clearly following from some of his fundamental commitments (particularly E1a4 and E1p10), is nevertheless very difficult to motivate in terms that would persuade someone who did not already hold these commitments. In fact, we saw that both Bennett and Della Rocca in the end conclude that Spinoza bases the causal barrier on nothing more than a rock bottom *assumption*. In this section I want to stress how problematic the causal barrier is by positioning Spinoza's view in relation to the contemporary mental causation debate. So let us take a step back and briefly review what is at stake in this debate.

There is not just one mental causation problem; however, the problem that has received most discussion in the past decades is the causal exclusion problem.⁸² This argument can be considered as a way to raise worries about mental causation for “dualist” positions – where “dualism” is simply the denial of the numerical identity of mental and physical causes. The dualist positions in question are nonreductive physicalism and property dualism, as was mentioned in the introduction.⁸³ Such positions typically accept the causal closure of the physical, which is the idea that every physical effect has a sufficient physical cause. But given that mental causes are assumed to be distinct from physical causes, it follows that any mental cause of a physical effect must now apparently work ‘in addition’ to the physical cause. The most natural way to think of this is through the idea of overdetermination. Unproblematic cases of overdetermination are those in which there are multiple independent sufficient causes of one event, like a Covid infection resulting from attending a party with many people that have Covid. But a view on which mental and physical causes are independent, but somehow consistently always together produce their overdetermined effects is held to be wildly implausible. (In addition, the nonreductive physicalist will want to say that the mental cause in fact is not independent from the physical cause at all; they will appeal to a relation like *realization* in order to say that the mental cause is dependent on the physical one.⁸⁴)

Bringing this problem back to its most basic form, it can be reconstructed as follows.

The minimal exclusion problem:

- i. *Dualism*: Mental causes are distinct from physical causes.
- ii. *Mental causation*: Some mental causes are sufficient causes of physical effects (and vice versa).
- iii. *Causal closure of the physical*: Every physical effect has a sufficient physical cause.
- iv. *No overdetermination*: Physical effects do not (normally) have more than one sufficient cause.

This reconstruction of the problem is ‘minimal’ because it includes only the bare essentials needed to cover the debate. In particular, *No overdetermination* (iv) combines two claims that are often separated: a rejection of overdetermination, and a rejection of the idea that effects can

⁸² For an overview of the different problems associated with mental causation, see Bennett (2007). Classic texts for the exclusion problem include Malcolm (1968) and Kim (1993, 1998).

⁸³ See fn. 75 above.

⁸⁴ See Shoemaker (2007, 2013) for more on realization.

have more than one sufficient cause. This latter claim is more controversial than the former, and much of the debate addresses it – by effectively claiming that an effect can have more than one sufficient cause without being overdetermined.⁸⁵ But for our purposes we can make do with the simpler (iv). Additionally, I sidestep the issue of whether causal relations are taken to hold between *token* causes (events or tropes) or *types* (properties), and instead speak simply of ‘causes.’

The problem takes the shape of an inconsistent tetrad. Someone who accepts all of (i)-(iv) is faced with the problem that mental causes seem to be “excluded” from their view of the world: it is unclear how a mental cause could be the sufficient cause of any physical effects, if these latter already have sufficient physical causes and mental causes are not allowed to overdetermine their effects. The debate, accordingly, is over which of (i)-(iv) to give up. As noted, it is (iv) that most of the debate centers on. (ii) and (iii) are held to be less controversial, although both are occasionally disputed. While (iii), causal closure, is normally considered to be a part of scientific common sense, and therefore as something that is not to be doubted, it is sometimes thought that the principle is stronger or more contentious than it is normally taken to be.⁸⁶ Likewise, one response to the problem, although not a popular one, is to deny mental causation, (ii). It is possible for a nonreductivist to adopt a view where higher-level causes (including mental causes) are realized by fundamental, physical causes but have causal relations only to other higher-level effects, and where the fundamental physical causes similarly never have higher-level effects. Such a position can therefore reject (ii).⁸⁷

Yet the simplest response to the argument is to deny (i) and adopt an identity theory of mental and physical causes. The very existence of mental causation is taken both by Papineau (2002, 17-8) and Kim (2011, 111-4) to be a strong argument in favor of such an identity theory. Kim, in effect, argues from (ii) and (iii) to the denial of (i). (Papineau’s argument depends on a stronger version of (iii), but is otherwise similar.) While not strictly valid, such an argument is elegant and intuitive. If mental causes can have physical effects, but every physical effect has a sufficient physical cause, then the easiest way to assign a causal role to the mental cause is by

⁸⁵ See Bennett (2003, 2008) as well as many of the essays in Gibb, Lowe and Ingthorsson (2013).

⁸⁶ Papineau (2001) presents a defense of causal closure. The principle is questioned by Lowe (2003), Montero (2003) and Gibb (2015).

⁸⁷ For two examples of such a view, see Gibbons (2006) and Zhong (2014).

identifying it with the physical cause. The mental cause m_1 is a sufficient cause because the physical cause p_1 is, and m_1 is identical to p_1 .

Note that this argument makes no assumptions about whether mental causes are efficacious “qua mental,” nor about causal explanation. Nor should it, because these further questions already take us beyond the scope of the minimal exclusion problem. All that it takes to respond to this argument is to find a bare sufficient causal role for mental causes. To accuse identity theorists of failing to find a place for intrinsic mental efficacy arguably begs the question against them, because it in effect forces them to distinguish the mental and physical cause.

When Spinoza's position is considered in the context of the minimal exclusion problem,⁸⁸ it quickly becomes clear how problematic his position is. For, at first glance, he seems to deny both (i) *and* (ii). From the perspective of the contemporary debate, this puts him at a disadvantage in relation to positions that employ greater parsimony and deny just one of these claims. In particular, the disadvantage, of course, results from his denial of the highly intuitive (ii).

What should be done? Spinoza's denial of mental causation is, as far as I can see, consistent; that is, it does not lead to contradictions internal to his own metaphysics. Yet, as we have now seen, it flows out of a causal barrier that Spinoza apparently simply assumes without defending and that interpreters have had trouble defending. Moreover, it has entailments that are eccentric when compared to current-day writers on mental causation; to the extent that the present discussion makes sense to one, Spinoza's own views will probably strike one as strange, perhaps as unmotivated. I think it is fair to say that at present it is simply not understood why Spinoza denies the conclusion of the mental causation argument.

Still, I do not want to leave the discussion here. Instead, in the next section I want to point out a source of anti-interactionist arguments in Spinoza that does not depend on the causal barrier. This text, the often neglected scholium to E3p2, occurs in a polemical context and so is designed to directly attack interactionists. This means that it may have persuasive force that the

⁸⁸ I take it that this can be done, i.e., that (i) to (iv) have their (approximate) counterparts in Spinoza's system. It is especially “distinctness” in (i) as well as the principle of causal closure (iii) that require interpretation in order to relate them to Spinoza's views. But if distinctness is interpreted as *numerical* distinctness, then it is just the denial of a (numerical) identity view, which is the view I am presupposing Spinoza to hold (as I noted in section 2.2 above). As for causal closure, aside from E2p6 it can also be derived from Lemma 3 after E2p13s, which says that bodies can only be determined to motion and rest by other bodies (as pointed out by Nadler 2006, 132). Spinoza is, of course, independently committed to the view that everything has a sufficient cause.

causal barrier between the attributes by itself does not have. In addition, in section 5, I discuss two further ways, besides his adoption of an identity theory, in which Spinoza's views are especially germane to accommodating mental causation.

Studying Spinoza's other anti-interactionist arguments will bring out a surprising feature of them: unlike E2p6 and E3p2, they do not entail a denial of all forms of mental causation. They only target *interactionism*, a view which from now on I will understand as a combination of substance dualism together with acceptance of mental causation. It is a stronger position than (ii) by itself, because the latter only posits a bare causal relation between a mental and a physical thing. Interactionism goes beyond this by saying that the mental cause has to be capable of acting by itself, independently of any subvening physical cause or realizer. The view can, then, be defined as follows:

- v. *Interactionism*: Some mental causes are sufficient causes of physical effects independently of any physical causes or realizers.

As I will now argue, the way Spinoza restates E3p2's argument in its accompanying scholium, and the arguments of the scholium itself, show that he takes himself to be targeting this stronger view. Rejecting this view clearly does not necessarily involve rejecting (ii). Spinoza rejects (v) for being incompatible with (iii), causal closure. He also argues that interactionist positions involve mistaken ideas about free will and the explanation of behavior. If my reconstruction of his argument is correct, then there is nothing in E3p2s, at least, that forces him to reject (ii).

4. *Spinoza against interactionism*

So far, this paper has considered E3p2 mainly in the context of its demonstration and the other propositions, such as E2p6, that it cites. But besides a demonstration, E3p2 also has one of the most extensive scholia in the *Ethics* (4 pages in Curley's edition). Can it help us to understand Spinoza's position?

The scholium opens with a surprise. In its first paragraph, Spinoza writes:

These things are more clearly understood from what is said in E2p7s, viz. that the mind and the body are one and the same thing, which is conceived now under the attribute of thought, now under the attribute of extension. The result is that the order, *or* connection, of things is one, whether nature is conceived under this attribute or that; hence the order

of actions and passions of our body is, by nature, at one with the order of actions and passions of the mind.

We have until now been working on the understanding that Spinoza's adoption of an identity theory is in tension with his anti-interactionism: something had to be done to reconcile them. But looking at Spinoza's actual words in the scholium, it is clear that he sees no tension whatsoever. In fact, he appears to believe that the argument given in E3p2dem is an equally powerful alternative to – and hence compatible with – an argument based in identity. The transition from the demonstration of E3p2 to its scholium is so direct that it is difficult to believe that Spinoza could have failed to see this most basic connection between his identity theory and his anti-interactionism. So an interpretation needs to be found on which they are genuine alternatives, and can equally be used to argue against Spinoza's target in E3p2. However, under what conditions could that be the case?

I believe that it is only on the assumption that Spinoza is not distinguishing interactionism (v) from mental causation (ii) that the transition from E3p2dem to E3p2s is intelligible. In the scholium he is targeting interactionism, not mental causation *per se*. Again, if mental causation had been the target, the appeal to an identity theory would have been mysterious, if not counter-effective. The identity of mental and physical causes is typically seen as an intuitive way – perhaps the *most* intuitive way – of explaining how mental causes can be efficacious. To say that m_1 cannot be the cause of p_2 because $m_1 = p_1$ is, at best, an apparently confused argument that needs further elucidation. However, it is not similarly confused to take the adoption of an identity theory to have significant consequences for the idea that the mind can, entirely by itself, cause motion in the body.

It is not surprising, then, that after its first paragraph, E3p2s is largely given over to a critique of the freedom of the will. Spinoza attacks the idea that the mind could, by a free and independent act, cause any changes in the body – the idea, as he puts it, “that the body now moves, now is at rest, *solely* from the mind's command, and that it does a great many things which depend *only* on the mind's will and its art of thinking.” (G II, 142; my italics) I take it that this view is, if not a restatement of interactionism, then at least a view that implies it. In other words, it involves the idea of a mental cause that acts on the body without a corresponding physical cause.

Spinoza's opposition to this view centers around two claims. One we have just encountered: it is his assertion that “the order of actions and passions of our body is, by nature, at one with

the order of actions and passions of the mind.” The other is his famous claim that “no one has yet determined what the body can do.”

I take it that the target of the claim about actions and passions is the substance dualist view on which the body directly influences the soul. On a traditional interpretation of action and passion, such a view would then entail that, the body being the active agent and the soul the one being acted on, the one would be active and the other passive. In that case, then, the “order of actions and passions” in our bodies and minds would necessarily depart from each other. Descartes states this traditional view in the second article of the *Passions of the Soul*:

I note that we are not aware of any subject which acts more directly upon our soul than the body to which it is joined. Consequently we should recognize that what is a passion in the soul is usually an action in the body. (CSM I, 328; AT XI, 328; see Deleuze 1988, 18)

Spinoza's claim that what is an action in the mind is also an action in the body precludes, then, a traditional interactionist construal of the mind-body relation. If the mind acted on the body (or vice versa) then necessarily during that action one would be active and the other passive. By denying that conclusion, Spinoza rules out interactionism.

Moreover, the principle supports an identification of mind and body, as that would give a natural interpretation to the idea that they are active and passive at the same time. However, note that it could in principle be reconciled with a nonreductive physicalist position. So long as the mental effect is held to supervene on a physical effect, both could be passive at the same time in response to a shared cause. Both could then also be “active” at the same time – in the sense of having a shared effect. This would raise worries over overdetermination, precisely as it does in the context of the minimal exclusion problem, but it does not undermine the shared order of actions and passions. A direct action of mind on body without a subvening physical base would, however, undermine it. The target of Spinoza's principle, then, is interactionism and not mental causation per se.

The same is true for the claim that “no one has yet determined what the body can do.” To be effective in its polemical context within the scholium, this claim needs only to be read as saying that a physical explanation for any bodily event is available, not that no mental one is. Spinoza associates interactionism with the view that our consciousness can directly determine how our minds move our bodies, without the need for a corresponding investigation into the powers of the body. Much like in the case of the claim about the actions and passions, the sting of

Spinoza's response to this idea lies in his parallelism: in the idea that it is not just that the body has to be investigated in order to establish the mind's power, but that it will turn out that the mind's power will be equal, in some sense, to that of the body. But leaving this point aside, Spinoza's examples in the scholium invoke the availability of physical explanations, not their exclusivity. Two central paragraphs make this clear:

They [the defenders of interactionism] will say, of course, that it cannot happen that the causes of buildings, of paintings, and of things of this kind, which are made only by human skill, should be able to be deduced from the laws of nature alone, insofar as it is considered to be only corporeal; nor would the human body be able to build a temple, if it were not determined and guided by the mind.

But I have already shown that they do not know what the body can do, or what can be deduced from the consideration of its nature alone, and that they know from experience that a great many things happen from the laws of nature alone which they never would have believed could happen without the direction of the mind—such as the things sleepwalkers do in their sleep, which they wonder at while they are awake. (G II, 142-3)

Spinoza's criticism is directed at the attempt to support interactionism through a claim of human exceptionalism. The interactionist argues that certain things humans do or produce are so complex that they could not be produced by physical causes alone, and that they must hence be the result of independent mental causes. Against this, Spinoza points out that many equally complex behaviors result from processes that are allowed to be entirely natural and to occur "without the direction of the mind," such as the behavior of sleepwalkers. He replaces the exceptionalist claim with the programmatic statement that we do not know yet what the body can do, so that it would be a mistake to rule out the possibility of physical explanations of complex behaviors and, a fortiori, a mistake to accept interactionism. However, note that the underlying principle of Spinoza's argument – that every form of human behavior has a physical explanation – is not itself at odds with the idea that it could have a mental explanation as well. Another way to put this is that Spinoza is attacking the interactionist's denial of the causal closure of the physical, when mental causation itself is entirely compatible with causal closure.

5. *Two potential sources of mental causation*

The previous section showed that Spinoza mobilizes his identity theory to attack interactionism. Here I want to highlight three further components of Spinoza's theory of the mind that arguably put pressure on him to adopt some form of mental causation, while being incompatible with interactionism. These are his psychophysical theory of appetite and desire and his theory of intentional action.

5.1 *Appetites as psychophysical causes*

Spinoza's psychophysical theory of appetite and desire puts pressure on him to accept mental causation, while being incompatible with interactionism. Spinoza conceives of desire as a psychophysical state which is expressed both in the mind and in the body. He defines it as "appetite together with consciousness of the appetite" (E3p9s; italics removed). Appetite in turn is defined as *conatus* "when it is related to the mind and body together" (E3p9s). (*Conatus* is the "striving" to persevere in one's being (E3p6) that Spinoza takes to be one manifestation of the essence of each thing (E3p7).) So the desire in this case is not a purely mental phenomenon, but something that also has a physical aspect. As LeBuffe (2010a, 130) writes, "Spinoza is somewhat indifferent about the way in which he refers to striving [i.e., *conatus*]: it is the essence of the person, which may be labeled by terms that emphasize mental, physical, psycho-physical, or conscious aspects of it but which, properly understood, is all of these together."

Now, if Spinoza truly denied mental causation, this would lead him into the contrived position of holding that there are cases where it may be true that

- (10) appetite A – being a psychophysical state – causes an effect *e*,

while the related

- (11) appetite A, *insofar as it is mental*, causes *e*

would be false. Note that, because one appetite may cause another, *e* may itself be a psychophysical state. In that case, (12) would be false but not (13):

- (12) appetite A, *insofar as it is mental*, causes *e*, *insofar as it is mental*.

Say that appetite A is directed at apples. There are apple trees in the garden. My appetite toward apples then causes an appetite to go to the garden. Similarly, my desire for an apple – which

just is my appetite, but now described in relation to my mind – causes a desire to go into the garden – which, again, is the same as its corresponding appetite. But my desire does not cause the appetite, nor vice versa. Again, it does not seem *contradictory* for Spinoza to adopt this view. But it is highly contrived, and his own conception of strivings as psychophysical states puts serious pressure on him to recognize mental causation, so that he can accept (12) and not just (13).⁸⁹

5.2 *Intentional action*

A related worry arises from Spinoza's sparse comments about intentional action. Considering it will reveal some of the best evidence that he understands the pressure on him to accept some forms of mental causation.

It is so natural as to be almost unavoidable to explain the outcomes of behavior by a (mental) intention to produce these outcomes. Conversely, it is also natural to say that the object of an intention, that what it aims to do or achieve, is an actual thing outside of the mind, instead of a mental entity. But if Spinoza held strictly to the causal barrier between the attributes, both these things would be impossible for him to say. He would have to say that my desire for an apple would produce my *idea* of eating the apple, but not the eating of the apple itself. And if an apple gets eaten, he would have to say that, insofar as it is a physical event, its cause is only a complex physical state, instead of my desire. These are highly counterintuitive and unattractive positions to take.

Thankfully, however, we have already seen that Spinoza does not take them, or at least does not take them consistently. In E4pref, after all, he is happy to say that a desire (a mental entity) is an efficient cause of the (physical) construction of a house. This passage does not stand on its own. In fact, it is underwritten by a crucial passage of E3p28dem:

We strive to imagine, as far as we can, what we imagine will lead to joy (by E3p12), i.e. (by E2p17), we strive, as far as we can, to regard it as present, or as actually existing. But the mind's striving, or power of thinking, is equal to and at one in nature with the body's striving, or power of acting (as clearly follows from E2p7c and E2p11c). Therefore, we

⁸⁹ The psychophysical nature of *conatus* is emphasized in some earlier interpretations: see Hampshire (2005, 61-70, 101-6) and Wartofsky (1973). See also Jaquet (2018).

strive absolutely, or (what, by E3p9s, is the same) want and intend [*appetimus et intendimus*] that it should exist.

This part of the demonstration makes several important points. First, it corrects a misunderstanding that E3p12 and 13 could have generated. (E3p13 is cited in the part of the E3p28dem that I have not quoted.) These propositions state the following:

E3p12: The mind, as far as it can, strives to imagine those things that increase or aid the body's power of acting.

E3p13: When the mind imagines those things that diminish or restrain the body's power of acting, it strives, as far as it can, to recollect things that exclude their existence.

At first sight, these propositions suggest precisely the counterintuitive view. They appear to say that mental strivings are directed only at further mental acts: they are strivings to imagine things but not to produce the things that correspond to them. E3p28 and its demonstration correct this first impression by pointing out that this mental striving is identical to a striving of the body. So the mental act of trying to think of something is identical with a bodily act of actually pursuing it and bringing it about. My act of trying to think of an apple, then, is identical to my physically going to get one.

This goes some way toward explaining how the mind's strivings produce physical actions. However, it does not constitute a complete answer. For Spinoza still holds that the mind's striving, *insofar as it is mental*, is directed only at an idea. He is still describing claims of type (13) and not (12). However, he then goes on to make two important additional moves. First, he introduces a notion of striving "absolutely." He does not explain what he means by this term, but a plausible gloss is the following. Humans consist of mind and body, and both the mind and the body strive to do things. "Absolute" striving is most naturally interpreted as the striving of the individual human being as a whole. The term seems designed to point out that there really is no difference between the striving of the mind and that of the body (they are "equal" and "at one in nature"), so that the human being as a whole may be said to have a single, absolute striving. Spinoza's second move, his identification of this absolute striving with a "wanting and intending," confirms this analysis. By referring back to the definitions of appetite and desire in E3p9s, Spinoza indicates that "wanting" (*appetimus*) is supposed to be read as indicating the presence of an appetite, or (in the words of E3p9s) *conatus* "related to the mind

and body together". More clearly than E3p9s, this confirms that Spinoza would be happy to say that humans strive to bring about physical effects.

This also coheres well with his definition of the affects as involving "man's essence" and "man's passage" from one degree of perfection to another (E3DeffAff1-3). As Jaquet (2018, 136) points out, "man" (*homo*) for Spinoza explicitly refers to a human being composed of mind and body (E2p13c). So to say that "man" undergoes something or does something is to indicate that both the mind and the body are involved. Spinoza is happy to speak in this loose way of human actions. And in at least one case – E4pref – Spinoza speaks more specifically of a desire's causing a physical effect.

Spinoza's argument in E3p28dem turns on the fact that desires are identical to appetites. It goes roughly as follows. (Just like in Spinoza's text, this reconstruction has an obvious suppressed premise.)

- (13) The mind strives to form the idea of x .
- (14) The mind's striving to form the idea of x = the body's striving to produce x .
- (15) Therefore, we absolutely strive to produce x .

Stated like this, the argument's form comes extremely close to that of the mental causation argument. However, it is not the same, because it does not conclude with:

- (16') Therefore, the mind strives to produce x .

Moreover, it remains unclear if Spinoza thinks that (16) implies (16'). However, the argument does establish that whenever our mind strives to think of x , we absolutely strive to produce x . This shows that Spinoza recognizes a form of action explanation that connects intentions (strivings) to physical effects. When the apple gets eaten, it is not *just* a physical cause that produces this result: it can validly be said to result from my absolute striving. And my desire for the apple results, absolutely, in my eating the apple. So Spinoza notices the tight interrelation between mental and physical strivings, and exploits it to move toward a more acceptable account of intentional action. This interrelation, again, turns on his identity theory.

6. Conclusion

Spinoza denies mental causation in prominent places in the *Ethics* (E2p6 and E3p2). Many commentators have attempted to use his conceptual-causal barrier between the attributes to explain this feature of his views. The problem, however, is that it is not clear why Spinoza adopts the conceptual-causal barrier in the first place. Because this is not clear, it is also not clear how strong his arguments against mental causation are. I have tried to bring this out by comparing his views with standard assumptions of the contemporary mental causation debate. Without a clear defense of Spinoza's conceptual-causal barrier, it's not clear why a believer in mental causation should not just reject the barrier out of hand.

In light of this situation, it is important to notice that Spinoza does not seem to systematically distinguish between mental causation and the stronger thesis of interactionism. It turns out that several of Spinoza's arguments target this latter view and not mental causation per se. Furthermore, these arguments are independent of the causal-conceptual barrier and instead depend on his identity theory. Spinoza himself clearly recognized the incompatibility of his identity theory with interactionism. Moreover, there is significant pressure on Spinoza to accept some mental causation from his psychophysical theory of appetite and his account of intentional action. In the last case, Spinoza recognizes this and himself allows for mental causation.

Spinoza rejected interactionism because he believed that it implied false views about human psychology, particularly in relation to free will and motivation. Because these views, in his time, were mainly supported by a substance dualist metaphysics, he opposed that metaphysics as well. Insofar as contemporary theories in the mental causation debate can be stated without implying such views, they fall outside of the scope of Spinoza's criticisms. In the absence of a compelling account of the conceptual-causal barrier, Spinoza's relation to contemporary assumptions about the nature of mental causation remains unclear. Instead, his arguments are directed against a view, substance dualism, that is not as often discussed in this debate, which rather centers around views that accept some form of supervenience claim of the mind on the body. If Spinoza does have contributions to make to this wider debate, they will perhaps center around his criticisms of the ordinary phenomenology of action (e.g., E1app, E2p35s, E3p2s, Ep58) and his conception of how a certain form of naturalism informs a theory of psychology, and not around his precise views on the metaphysics of causation.

CHAPTER FOUR

Spinoza's Deterministic Theory of Motivation

*For there is no creature whose inward being is so strong
that it is not greatly determined by what lies outside it.*

George Eliot, *Middlemarch*

1. Introduction

The first part of Spinoza's *Ethics* sets out a ruthlessly deterministic – even necessitarian – system. It formulates a metaphysics in which there is a single substance – *Deus, sive Natura* (E4pref), i.e., God or Nature – in which all other things inhere as modes. “From the necessity of the divine nature there must follow infinitely many things in infinitely many modes (i.e., everything which can fall under an infinite intellect.” (E1p16) These modes depend on God both for their essence and their existence (E1p25), and all finite modes are determined by other modes to behave in the way that they do in an infinite causal chain (E1p28). Spinoza draws the necessitarian⁹⁰ conclusions of his system in two further propositions:

In nature there is nothing contingent, but all things have been determined from the necessity of the divine nature to exist and produce an effect in a certain way. (E1p29)

Things could have been produced by God in no other way, and in no other order than they have been produced. (E1p33)

⁹⁰ Curley and Walski (1999) and Martin (2010) dispute the ascription of (strict) necessitarianism to Spinoza; Garrett (1991) and Griffin (2008) defend it.

As the demonstration of the latter proposition points out, any change in the order in which things had been produced would have entailed a change in the nature of God, which Spinoza holds to be inconceivable (see also E1p20c2). There is therefore no way in which either the modes or God himself can diverge from the necessary order of nature.

In spite of what claims such as these might lead one to believe, however, Spinoza does not go on to declare the existence of freedom a complete illusion. He does attack the concept of *free will*, which he understands as the idea that there is a faculty of willing that is (i) independent of the intellect (E2p49), (ii) capable of “willing things or not willing them” (E2p48, Ep58),⁹¹ and (iii) can immediately cause the body to move (E3p2, E5pref). Spinoza rejects this conception of free will as absurd and he rejects all three components.⁹² Due to (ii), this understanding of free will is also incompatible with determinism. Spinoza writes: “In the mind there is no absolute, or free, will, but the mind is determined to will this or that by a cause which is also determined by another, and this again by another, and so to infinity.” (E2p48) Elsewhere, Spinoza famously calls the mind a “spiritual automaton” (TIE, 85), whose operations are just as strictly subject to natural laws as those of the body.

However, Spinoza does not deny the existence of *freedom* per se.⁹³ He even goes to great lengths to redefine this concept in a way that is compatible with his deterministic system. His definition of freedom is as follows:

That thing is called free which exists from the necessity of its nature alone, and is determined to act by itself alone. But a thing is called necessary, or rather compelled, which is

⁹¹ In contemporary terms, this means that Spinoza takes the Principle of Alternate Possibilities to be part of the concept of free will. It also appears to entail that free will presupposes a view of the will as a so-called “two-way power” for Spinoza, i.e., a power that can either act or refrain from acting (see Lowe 2013, Frost 2020).

⁹² Some of the main textual loci for Spinoza’s criticisms of free will are E1p17s, E1p32, E1p33, E1p33s2, E1app, E2p35s, E2p48, E2p49, E3p2s, TP II, 6-7, Ep21, Ep58.

⁹³ This means that Spinoza’s position is difficult to phrase in terms of the familiar opposition between compatibilism and incompatibilism. Klein (2019) stresses the incompatibility of Spinoza’s thinking with a traditional understanding of free will. Kisner (2011, 18) writes that Spinoza is a compatibilist about freedom. They are both right: Spinoza believes that *free will* and determinism are incompatible, but not that *freedom* is. Because current debates tend to define compatibilism as a claim about *free will* and not freedom (see van Inwagen 1983 and McKenna and Coates 2020), from that vantage point it is perhaps better to describe Spinoza as an incompatibilist. But this is a terminological decision only. Note that if hard determinism is understood as the idea that free will does not exist because determinism is true, then Spinoza is a hard determinist, even while also being a compatibilist about freedom. See below for my suggested label for Spinoza’s position.

determined by another to exist and to produce an effect in a certain and determinate manner. (E1def7)

An agent is free when they are “determined to act by [themselves] alone,” and an action is free when it similarly follows from the agent’s nature.⁹⁴ This definition places necessity at the center of the concept of freedom. Insisting on an apparent paradox, Spinoza writes: “I place freedom not in a free decree, but in [...] free necessity.” (Ep58) In their titles, Parts Four and Five of the *Ethics* – which refer to the unfreedom (“bondage,” *servitus*) that humans experience due to the passions they suffer from and “the power of the intellect, or human freedom,” respectively – both announce the central role of freedom in Spinoza’s ethical project. They make it clear that Spinoza, in spite of his determinism, does not think it is impossible for humans to attain a certain measure of freedom.

Spinoza’s definition of freedom articulates a sourcehood condition on freedom:

Source: An action A is free if its only source is the agent.

However, how can he accept *Source* as a determinist? Contemporary compatibilist attempts at accommodating Source often define what it is to be the source of A in terms of a psychological or a normative, not a causal, relation. For example, they hold that an agent is the source of their action if they are “reasons-responsive,” or if they “identify” with the reasons that caused them to act.⁹⁵ Notably, the ability to respond to or to identify with reasons is here taken to be capable of occurring in spite of determinism. These compatibilists claim that agents can act freely even in situations in which their actions have been fully determined by external forces.

Libertarians have responded to these arguments by insisting on a causal understanding of *Source*: “True sourcehood – the kind of sourcehood that can actually ground an agent’s freedom and responsibility – requires [...] that one’s action not be causally determined by factors beyond one’s control.” (O’Connor and Franklin 2020, §2.5) Surprisingly, Spinoza fully agrees with this intuition, as on his account of freedom an agent must act “according to the necessity of their own nature” and not be determined by external factors. However, doesn’t determinism entail that agents are always determined by external factors, namely by prior causes in the causal

⁹⁴ Note that, throughout this chapter, I will use the term ‘action’ the way it is now normally used, to refer to anything an agent does or any part of their behavior. Spinoza uses the term in a stricter sense to refer to truly *free* actions (E3def2). Hence, many of the things I will call actions would for Spinoza really be passions.

⁹⁵ See O’Connor and Franklin (2020, §2.4) for further discussion of these two strategies.

chain? How can Spinoza reconcile his causal understanding of *Source* (or *causal Source*, for short) with determinism?

This chapter discusses this problem. Answering it will help to then respond to a second problem, which I'll call the motivation problem. Compatibilists have traditionally struggled to assign a genuine causal role to the will. They have tended to hold either that the will is entirely determined and itself causally impotent (epiphenomenalism), or that it does cause actions but is at the mercy of other, non-volitional and often irrational mental states (a "secondary cause" model of the will). Neither view is very attractive. Spinoza's response to the source problem extends to an interesting response to this problem as well.

The structure of this chapter is as follows. The core of a Spinozan response to the source problem is found in Spinoza's theory of causation, which assigns an intrinsic power to things' essences. The next section sets out this theory. The third section then shows how this account of causation answers the source problem. Section 4 introduces the motivation problem and sets out Spinoza's response to it, as well as the novel, distributive theory of motivation that underlies this response. Section 5 defends the legitimacy of this theory by comparing it to a 'vector' model of agency. Section 6 describes how Spinoza's theory incorporates a form of mind-body parallelism and section 7 is a brief conclusion.

2. Intrinsic power and causation by essences

That a thing's essence plays a role in determining what it is capable of is not such a strange idea. For example, the fact that something is a ball determines that it can roll down an incline. Here, what the thing *is* settles what it can *do*. However, while this is part of what Spinoza means when he says that things' essences play a role in their actions, he goes beyond it by saying that these essences exert actual causal power. He identifies the essences of individual things with a *conatus*, or "striving," to persevere in existence (E3p7). Here I want to briefly set out how he arrives at this claim. In sketching his arguments, I pay particular attention to the closely inter-related views about causation, explanation and power that underlie them.

It is well known that Spinoza sees close connections between causation and explanation. For Spinoza, causal relations imply conceptual relations. That is, he accepts the following principle:

Causation → *Conception*: If x causes y , then y can be understood through x .⁹⁶

Causation → *Conception* is entailed by E1a4: “The knowledge of an effect depends on, and involves, the knowledge of its cause.” This traditional scholastic principle entails that whenever x causes y , “knowledge” (i.e., *cognitio*, which can also be translated by “cognition”) of the effect depends on knowledge of the cause somehow. Spinoza applies this principle in a more demanding way than the scholastics, however, because his conceptual relation is a relation of a priori entailment: for him, if one understands the cause, one can deduce the effect. We can express this by saying that Spinoza holds that causation is *scrutable to reason* (cf. Lin 2019, 37-9):

Causal scrutability: If x causes y , then the idea of y can be deduced from the idea of x .

All causation for Spinoza therefore underwrites these entailment relations. By contraposition, if y cannot be deduced from x , x cannot be y 's (total) cause.

From the beginning of the *Ethics*, Spinoza's claims about power stress its interrelation with essence. Spinoza equates following from a thing's essence with being produced by that thing's power. E1p34 states that “God's power is his essence itself.” Its demonstration is as follows:

For from the necessity alone of God's essence it follows that God is the cause of himself (by E1p11) and (by E1p16 and E1p16c) of all things. Therefore, God's power, by which he and all things are and act, is his essence itself, q.e.d.

The demonstration hinges on the principle that if something follows from God's essence, it is produced by God's power. Because all things follow from God's power, he is absolutely powerful. A similar principle is at work in the case of things that are not absolutely powerful – i.e., modes. These things inherit God's power, which they “express” through their own essence (E1p36). As Spinoza writes in E4p4dem:

The power by which singular things (and consequently, man) preserve their being is the power itself of God, or Nature (by E1p24c), not insofar as it is infinite, but insofar as it can be explained through man's actual essence (by E3p7). Man's power, therefore, insofar

⁹⁶ The idea that Spinoza holds that x causes y iff y can be understood through x has been challenged by Morrison (2013). Note however that the present chapter does not assume this stronger, biconditional claim, but only the weaker, conditional one that if x causes y , then y can be understood through x . It is uncontroversial that Spinoza accepts this claim. Note also that I distinguish conceptual from ‘conceiving through’ relations, for reasons suggested by Melamed (2013b, ch. 3): Spinoza seemingly uses the conceiving through relation exclusively for discussing the relation between modes and their attributes. If so, it does not apply between modes (unlike the conceptual relation).

as it is explained through his actual essence, is part of God or Nature's infinite power, i.e. (by E1p34), of its essence. (Translation modified)

The principle at work is therefore a general one:

Essential powers: If x follows from y 's essence, then x was caused by y 's power.

This principle can be used to establish a connection with *Causal scrutability*: anything produced by y 's power is also caused by it, and hence can be deduced from the idea of y 's essence. For Spinoza, therefore, the laws of physics, insofar as they enable the making of deductions from physical cause to effect, by that very fact make appeal to causal powers.

So far, we have seen some relations between Spinoza's views about causation, conception and power. To complete the picture, it remains to say something about the role of essences (hence, of *conatus*). To do this, however, it will help to introduce Spinoza's view of the human being and its affections.

In Spinoza's system, humans are understood as "individuals" (E2def7) consisting of a mind and a body (E2p13c). The mind and the body are both modes of substance, belonging to the attributes of Thought and Extension, respectively. There are no causal or conceptual relations between modes of different attributes, which means that ideas in the mind cannot be understood through, or cause, motions in the corresponding body, and vice versa.⁹⁷ The mind, however, is just the idea that has the body for its object (E2p13), and moreover is identical to the body: they are "one and the same thing, expressed in two ways" (E2p7s).⁹⁸ Changes in the mind follow "the same order and connection" as changes in the body (E2p7s), a claim that can be understood either through the fact that mind and body are 'expressions' of one and the same thing, or through the fact that the mind perfectly represents the body, and hence its changes.

Both mind and body are modes of substance. Spinoza defines substance as a fundamental entity that is "in itself and is conceived through itself" (E1def3). Modes, by contrast, are those things that are "in another through which it is also conceived" (E1def5), i.e., those that are in and conceived through a substance. Their dependence on substance opens up a gap between a

⁹⁷ This is the view Spinoza states in E2p6 and E3p2. However, it is not clear that Spinoza truly rules out *all* cross-attribute conceptual and causal relations. See Chapter 3 for further discussion.

⁹⁸ I assume that Spinoza holds mind and body to be numerically identical, together with most commentators. Two recent dissenters are Marshall (2009) and Renz (2018).

mode's essence and its existence. Substance is "cause of itself" (*causa sui*), which is to say that it cannot be conceived except as existing (E1def1, E1p7). Modes' essences *can* be conceived without existence, and need a cause to come into existence; their essence does not imply their existence (see E1p24, E2a1). Their existence must be caused by other modes and will eventually also be ended by the causal influence of other modes.

Whether something is *causa sui* or not influences the kind of freedom it is capable of. For, by E1def7, an action is free when it follows from a thing's nature, or essence, alone. It follows that God is "absolutely" free, because anything that God does can be understood through his nature alone (E1p17, E1p17c2). The same is not true for the modes (E1p24, E1p25). But, even so, modes' actions must still be understood to follow at least partly from their essences. For, if a mode's action could not be understood to follow from itself in any way, then it could not be an action of that mode at all (by the contrapositive of *Causal scrutability*). This allows Spinoza to introduce a *graded* notion of freedom, according to which an action is more free to the extent that it follows more exclusively from an agent's essence. While only God is absolutely free, modes are free to some (limited) extent.⁹⁹

When an action is not completely free, it is partly determined by something outside of the agent. Spinoza explains how this goes with his theory of affection. Any mode is causally affected by other modes. In such an affection, the human body is affected by another mode of extension, while its mind, working in parallel to the body, forms an idea of the affection (E3def3). As *Causal scrutability* implies, this affection must be at least partially understood through the affecting cause – otherwise, there would not be causation at all. But the result cannot be entirely understood through the affecting cause; the body itself makes a contribution as well. Intuitively, as I suggested before, the nature of the affected body makes a difference in determining the outcome of the affection. Spinoza writes that the affection "must involve the nature of the human body and at the same time the nature of the external body" (E2p16). That is, it involves the essence of both bodies. Spinoza's sketch of a physical theory after E2p13 presents the basic rules determining the outcome of any such interaction.

A thing's essence plays a role, then, in any of its actions. As was mentioned before, Spinoza identifies this essential power, in the case of finite things, with *conatus* (E3p6, E3p7). The precise

⁹⁹ See Della Rocca (2003b, 205-6) and Kisner (2011, 32-3) for related distinctions.

way in which Spinoza establishes this claim is controversial.¹⁰⁰ But a central element is that each thing “is opposed to everything which can take its existence away” (E3p6dem). Things try to manifest their own nature and oppose those things that prevent them from doing this. I take ‘manifesting one’s own nature’ to involve, at minimum, producing effects that follow from one’s essence. As E3p7dem notes, “things are able [to produce] nothing [*nec res aliud possunt*] but what follows necessarily from their determinate nature”. So things cannot produce anything that would go against their nature (as E3p4 also notes), at least not insofar as their nature is involved. Possibly equivocating a lack of power to do one thing with a positive power to do its opposite, Spinoza concludes that things’ essences can only produce those things that help them to remain in existence. If things act against their own interests, this is because they have been overpowered by external forces.¹⁰¹

Things, therefore, inherently possess a degree of power, or *conatus*, which manifests to varying degrees depending on the extent to which their actions depend more on themselves and less on the affections that they experience. As modes, however, these things themselves depend on God. This raises an important issue. The idea that things have intrinsic powers can be described as ‘agent causation’.¹⁰² But genuine agent causation is normally taken to require that the agent is a *substance* that has *irreducible* powers. The Spinozan agents we are considering here don’t seem to meet either criterion: they are obviously not substances, and their power does not appear irreducible either but to reduce to God’s power. If finite things’ power derives ultimately from God, then can they be said to have any power of their own? Exploring this question will help to determine the status of these powers; it will also clarify Spinoza’s position by relating it to current debates.

My argument here has been that finite things for Spinoza have *intrinsic* powers – powers that they necessarily possess. Even if it turned out that these powers are reducible to God’s powers, it would remain true that they are intrinsic to the finite things that have them. So my argument in what follows is unaffected if the agent-causal terminology is set aside. Still, I think there is a sense in which it fits Spinoza’s view. Let’s reconsider the two criteria, irreducibility and

¹⁰⁰ For some interpretations, see Garrett (2002), Carriero (2011, 2019), Viljanen (2011, ch. 4) and Lin (2019, ch. 6).

¹⁰¹ See E3p4. Spinoza’s position on this point has strong implications for his conception of suicide, among others: this too can only occur as a result of outside forces. See E4p18s, E4p20s.

¹⁰² See Jorati (2017) for an argument that Leibniz is an ‘agent-causal compatibilist,’ a label that may fit Spinoza as well – with the provisos to be noted. For a contemporary agent-causal compatibilist, see Pereboom (2015). For agent causation more generally, see Griffith (2017).

substancehood. First, it is true that finite modes have no power that is not God's power. However, it remains true as well that, in case of a free action A resulting from an agent's essence S, the statement 'S caused A' is literally true, and there is no other mode M such that the statement 'M caused S to cause A' is true. Additionally, it should also be noted that S cannot be identified with God *simpliciter*; instead, it can only be identified with, as Spinoza tends to put it, 'God, insofar as he constitutes S.'¹⁰³ But if God can cause A only insofar as he constitutes S, then it seems fair to say that S has an irreducible role in the production of A. (Compare: if a piece of plastic can only roll down a slope insofar as it constitutes a ball, then its constituting a ball plays an irreducible role in its rolling down the slope.) To this extent at least, S's power remains irreducible.¹⁰⁴

Second, while it is of course clear that modes are not substances for Spinoza, it is also true that his requirements for something to be a substance are famously strict. In current-day discussions, these requirements are looser. Meghan Griffith writes: "Agent causation is *irreducible substance causation*. The agent, the person, is a metaphysical substance whose causal activity grounds autonomous agency." (Griffith 2017, 73; her italics) Such a view is to be contrasted with *event-* and *state-causal* views, which base autonomous agency instead in causation by mental events or states. Note that the salient feature of this division is whether agency derives from certain mental processes *in* the agent or from the agent's power *as such*. In terms of that division, Spinoza is closer to the agent-causal view than to the others, even if it remains true that modes are not substances.¹⁰⁵ Moreover, it seems fair to say that the agent's causal activity "grants autonomous agency" for Spinoza.

¹⁰³ For examples, see E2p11c, E2p12dem, E2p13dem, and many other occurrences. For an extensive discussion of why Spinoza blocks these so-called *secundum quid ad simpliciter* inferences, see Douglas (2018).

¹⁰⁴ The issues raised here concerning the relation between finite and divine power tie in with the broader problem of 'acosmism': the accusation that Spinoza cannot accommodate the reality of the modes because they wholly depend on God. However, this is a problem that cannot be discussed here. Suffice it to say that if one takes finite things' dependence on God to not prevent their reality, then one should similarly also not deny the reality of their essences and hence of their power. The charge of acosmism originates in Maimon's *Autobiography* (Maimon 2018, 63-4) and Hegel's *Lectures on the History of Philosophy* (Hegel 1990, 3:162-3). For recent discussions of the problem, see Melamed (2010), Newlands (2011) and Lin (2019, 128-32).

¹⁰⁵ Garrett (2002) argues that modes, to the extent that they are free, are "quasi-substances," because they are to that extent "in" themselves. I don't follow his suggestion here because a necessary condition to be a substance for Spinoza is to be independent from other substances. While a maximally free mode would be independent of other modes, it is hard to see how this would make it something independent from God. It would, for example, still be conceived through a certain attribute, hence "in" the substance with that attribute (and not "in" itself). However, even if the notion of 'quasi-substance' does not seem to make sense within Spinoza's system, perhaps it does remain fair to say that, insofar as modes have intrinsic causal powers, there is something 'substance-like' about them according to some more relaxed conception of substance.

There is, then, a case to be made for describing Spinoza as an agent causationist – more precisely, as an agent-causal compatibilist – even if there are also ways in which his views do not quite fit the contemporary terminology. More importantly, however, this section has shown how modes' intrinsic essential powers play a role in their actions. As the next section will make clear, this claim can be used to develop a distinctive answer to the source problem.

3. *Answer to the source problem*

Recall that the source problem asks how a causal determinist can accept a causal understanding of *Source*, i.e., accept that free actions are those whose only causal source is the agent. The problem can be reproduced in argument form as follows:

The source problem:

- (1) If determinism is true, then all things are causally determined.
- (2) A free action is one whose only causal source is the agent.
- (3) If x is the only causal source of A , x cannot itself be determined.
- (4) Therefore, if determinism is true, nothing is the only causal source of anything.
- (5) Therefore, if determinism is true, there are no free actions.

The argument is valid. (1) is a statement of determinism and (2) restates *causal Source*. (3) articulates a commonsensical constraint on what it means for something to be the causal source of something. It is based on something like the thought that, if something is caused by something else, then it depends on this other thing; but in that case, it would not be the (full, ultimate) causal source.¹⁰⁶

Compatibilists claim that determinism is true and that there are free actions, and so they must find a way to resist (5). As we have seen, they ordinarily do this by rejecting (2) and finding some non-causal way of formulating *Source* (or by finding a different conception of free action altogether).

Spinoza, however, accepts (2). He instead finds a way to deny (3). The argument goes through only if we take the causal determination in (3) to be determination *by something other than the agent*.

¹⁰⁶ Compare Galen Strawson's argument that for someone to be "ultimately responsible" for their actions, they would have to have been the cause of their own existence (Strawson 1986). (Strawson refers to being "*causa sui*" in this connection.) Analogously, being the only causal source of something precludes being dependent on anything else.

But if the agent is self-determining, then they can be the cause of an action even while being determined themselves. Indeed, one could go further and say that this self-determination is a possible determinist gloss on what it means to be the causal source of something in the first place. Hence Spinoza's claim that God, who is the ultimate causal source of everything, is self-caused. (3), therefore, is false, because something can be the only causal source of its actions if it is self-determined.

Spinoza's conception of self-determination is consistent with both determinism and the existence of free actions. Additionally, because actions can depend to a greater or lesser extent on the agent's *conatus*, Spinoza is also in a position to introduce a graded notion of *freedom to a degree*. An action becomes more free to the extent it is less dependent on external forces. While it is impossible to gain ultimate power over the affections, Spinoza's ethical therapy, which is intended to teach us to gain such control to whatever degree possible, is nevertheless not pointless, because it can help us to become *more* free. That is, it can help to ensure that our actions depend on our *conatus* to a greater extent.

In this connection, it is helpful to distinguish Spinoza's concept of the body's *power of acting* from that of the *conatus*. Because the *conatus* is simply the human being's essence and this essence, naturally, cannot change without the human being of which it is the essence also becoming something different, I take it that the *conatus* always manifests a fixed degree of power. However, Spinoza claims that the affects increase or decrease the body's *power of acting* (*agendi potentia*, E3def3). He goes on to note that the body is capable of maintaining its nature – and so its *conatus* – while changing its power of acting.¹⁰⁷ The body's power of acting, in turn, is determined by the degree to which the body is capable of being an “adequate” cause (E3def1, E3p1) – that is, the extent to which its actions follow from it alone.¹⁰⁸ The degree to which an action is free therefore depends on the body's power of acting. Power of acting can in turn be rephrased as the degree to which the *conatus* – and not any external force – is involved in the action. To the extent that we are unfree, our *conatus* is involved in actions that do not wholly follow from our essence and that may even harm us. The attempt to increase our freedom is

¹⁰⁷ E3post1. That he is concerned with the continuity of the body's nature is clear from the postulate's reference to Lemmas 5 and 7 after E2p13s.

¹⁰⁸ Because the body's power of activity is determined by the relation between the body and its actions and passions, it is fair to say that Spinoza holds a *relational* view of freedom. See Armstrong (2009), Ravven (2014) and Armstrong, Green, and Sangiacomo (2019) for the claim that his view of *autonomy* is relational.

therefore an attempt to make sure that our power is not used in ways that go against our own interests (as these are determined by our nature).

As a metaphor for the relation between *conatus*, power of acting, and external determination, consider the parallelogram of forces. The *conatus* can be construed as one force in the model, and the external affection represents another force (which derives from the *conatus* of the affecting mode). The combination of these forces determines the resulting action.¹⁰⁹ It also determines the power of acting, which can vary both by becoming stronger and weaker (thus resulting in the affections of joy and sadness, respectively; E3p11s) as well by being ‘directed’ either in the direction of the agent’s own *conatus* or away from it.¹¹⁰ The model makes it intelligible how one of the forces in the model – the agent’s *conatus* – can remain constant, while its power of acting varies with the ‘force’ and ‘direction’ of the affection (i.e., its power and the extent to which it, as Spinoza puts it, “agrees” or “disagrees” with us).

Spinoza therefore has an interesting and original response to the source problem. It is also clear that he allows for an indefinite approximation to a state in which our actions are truly free, i.e., self-caused (even if he is also clear that this state can never be attained). For all that, however, it is worth repeating that this is still a rigidly deterministic system. While agents may attain a degree of freedom, they cannot directly influence the way in which their *conatus* contributes to any action, no matter how free it is (see E1p27). Given a certain affection, its effect on the *conatus* and hence the resulting action are necessary and not under direct control. Human efforts to become more free therefore revolve around attempts to control the affections to which they are exposed, and the attempt to learn to respond to them in a more rational way; this is the core topic of the ethical theory finally developed in Parts Four and Five of the *Ethics*.¹¹¹ Becoming more free is never a matter of becoming undetermined or of leaving the necessary order of nature behind (E4p4). Even a relatively free mind remains a “spiritual automaton”.

¹⁰⁹ For a fuller discussion of the relation between power of acting and *conatus* that also situates these terms in relation to Descartes’s and Spinoza’s physics, see Sangiacomo (2015). His account of their connection nevertheless appears to differ slightly from mine. He writes that “insofar as external causes determine [a] thing to exist in a certain time and place, they determine it to bring about the effects that follow from its essence in that time and place, that is, by interacting in a certain causal network that will affect the thing’s causal efficacy.” (536) In the terms introduced below, this seems to commit Spinoza to a ‘secondary cause’ model of external determination and a ‘triggering’ model of the will, and not to the ‘distributive’ and (quasi-)‘vector’ models that I defend.

¹¹⁰ Conceivably, one and the same affection can both increase our power of acting and, as it were, turn it against us, thus making us more passive in the long run.

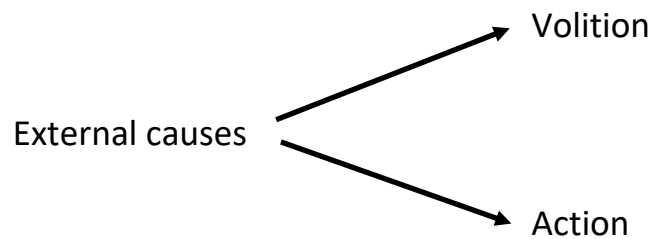
¹¹¹ For general discussions of this theory, see Garrett (2021a), LeBuffe (2010a) and Kisner (2011) for generally optimistic discussions and Lin (2009) for a more pessimistic one.

4. *The motivation problem*

Spinoza's peculiar conception of the role that self-determination plays in free action also provides him with a response to another problem for determinists. This is the problem of finding an adequate causal role for the will. It trivially follows from determinism that the will, in any action, must be causally determined. But how can we conceive of this kind of determination in such a way that it does not render free action impossible?

Unlike the source problem, this is a problem in moral psychology. The problem is how to find an adequate theory of motivation and agential control that does not conflict with determinism. It will help to bring the problem in view if we first sketch two possible models of determinist theories of motivation and understand why they fail. These are the *epiphenomenal* and *secondary cause* models of willing, which can be schematically represented as follows.¹¹² In both models, solid arrows represent causal determination relations; that is, the arrows relate sufficient causes to their effects.

Epiphenomenal model:



Secondary cause model:



Both models represent the will as entirely determined by external forces. On the epiphenomenal model, there is a shared cause both for the volition and for the resulting action, and the volitions themselves do not cause actions. On the secondary cause model, by contrast, the volitions *do* cause actions, but they have themselves been fully determined by a prior cause.

As drawn, the models are agnostic as to the precise nature of the external causes that determine the will. Katsafanas' discussion of a Nietzschean moral psychology takes them to be the drives

¹¹² Leiter (2007) describes both models in the context of a discussion of the motivation problem as it occurs in Nietzsche. My discussion, however, is more indebted to Katsafanas' discussion of Leiter's models (Katsafanas 2016). The diagrams are reproduced, with changes, from Katsafanas (2016, 149).

and affects of the agent (2016, 149). However, they could also be construed as brain states of the agent or as things in the agent's environment. More broadly, they could be understood as 'external' to – in the sense of independent from – the agent's will. I take it that these different ways of understanding external causes make no difference for the general point I am making here. As we have seen, however, the relevant external causes for Spinoza are the affections (or the passions); in other words, like for Nietzsche, they are other, irrational mental states. In what follows I'll therefore take the external causes to be affections.

Both of these models can be taken to be responses to the general problem for deterministic theories of how to assign a fitting causal role to the will. A basic principle that no theory of motivation should deny is the following:

If the agent's volition had been different, the agent's action would have been different.¹¹³

Denying this results in a form of fatalism, where the agent's actions are fixed regardless of the agent's actual volitions.¹¹⁴ Determinists can consistently reject this kind of fatalism, however, because they only claim that mental states have their own causes, not that these mental states do not cause any action. (In the case of the epiphenomenal model, a change in the will, while not itself the cause of any action, will presumably result from a change in its cause. Given that the cause also produces the action, the counterfactual relation between the state of the will and the action can therefore be maintained on this model.)

Both models are therefore compatible with accepting the principle above. That is not where they go wrong. Instead, their mistake is that they can accommodate the principle only by denying any influence to the will itself independently of other mental states. In both models, if the will changes, this is entirely due to the fact that it has been caused to change by prior changes in the agent's other mental states.¹¹⁵ Put in Spinoza's terms, then, these determinists accept this principle:

The agent's volition is entirely determined by the agent's affections.

¹¹³ A principle of this form was traditionally used as a compatibilist analysis of the freedom to do otherwise (for example, see Hume 1978, sect. 73). In that role it is open to serious objections (see McKenna and Coates 2020, §2 for a brief overview). But here it is not used as an analysis of freedom, but as a minimal condition on acceptable theories of the will.

¹¹⁴ This is a nonstandard understanding of what fatalism is, but it is illustrative in the present context.

¹¹⁵ Again, as Sangiacomo's interpretation appears to entail this view as well, I think it is unattractive. See fn. 109 above.

As Katsafanas notes (2016, 154), this leads the will to take up a very peculiar position in the agent's mental economy: commonly seen as the primary cause of deliberate action, it is relegated to a position where it has no causal powers at all – unlike the affects, which turn out to be the real, and total, causes of the agent's actions. Both models, then, end up by making the will the slave of the agent's affections and render it incapable of acting on its own.¹¹⁶

Precisely because they render the agent dependent on external forces, both models are incompatible with *causal Source*, and so Spinoza cannot accept them. Naturally, rejecting the models does not involve denying that affections influence or even determine the will¹¹⁷ – just that the will is *always, entirely* determined by the affections. In other words, it creates space to say that the will has some independent contribution to our actions. However, because a determinist should nevertheless hold that the will is determined by some cause, Spinoza should find a source that can supply the 'missing' determination as well. Even if the will is not (entirely) determined by the affections, there should still be a total cause – one that presumably incorporates affection as a part – that does determine it.

The prior discussion has already made it clear where Spinoza can find this additional source of determination: he can find it in his doctrine that an agent's essence, or *conatus*, makes a causal contribution to their actions. This solution is available to him because he closely connects the will to *conatus*. Their connection is stated in E3p9s, where Spinoza defines will as *conatus* “related only to the mind.” The will is one and the same thing as “appetite,” which is *conatus* “when it is related to the mind and body together.” Will and appetite are different ways of considering *conatus*, according as we place greater emphasis on the mental part of any action – such as the other decisions, wishes, or associations of the agent that play a role in its causation – or instead emphasize features that are not easily classified as either mental or physical – such as subconscious drives, habits, and instincts. Spinoza goes on to say: “This appetite, therefore, is nothing but the very essence of man, from whose nature there necessarily follow those things that

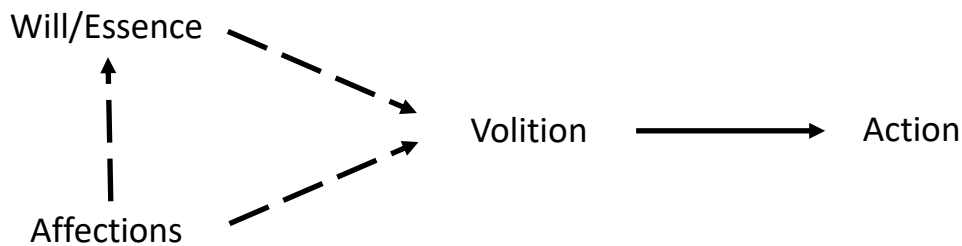
¹¹⁶ Spinoza sometimes appears to assert a position quite similar to this. In Ep58, for example, he writes that “experience teaches quite abundantly that there is nothing less in man's power than to restrain his appetites” (G IV, 266; cf. E3p2s), seemingly suggesting that the will is at the mercy of these appetites. However, I take it that Spinoza is here appealing to a traditional understanding of ethics (and freedom) as the attempt to overcome and control one's passions. While Spinoza allows this to be quite difficult, he does not say that the will is absolutely powerless to act independently of the appetites.

¹¹⁷ If this is taken to mean, however, that the agent's response is *wholly* dependent on the affections, so that the agent is *entirely* passive, then Spinoza would deny it. Because – as we saw in section 2 – in that case the agent would no longer make *any* contribution to the action, so that it would no longer be the agent's action at all.

promote his preservation. And so man is determined to do those things.” The will itself, however, considered as a particular way of conceiving the agent’s essence, can be distinguished from its volitions, or states, in the same way that the human mind can be distinguished from its ideas.¹¹⁸

The resulting picture is one on which the agent’s *conatus* and affections both play a role in determining the production of actions. As the agent becomes more free, the causal role of the affections correspondingly decreases, and vice versa. Because the affections in turn are causal influences on the *conatus* – as Spinoza writes in E2p48, “the mind is determined to will this or that by a cause which is also determined by another, and this again by another, and so to infinity” – the resulting model takes a triangular shape.

Distributive model:



In this model, solid arrows indicate causal determination and striped arrows indicate causal influence. The affection and will (or *conatus*) jointly determine a volition, which then itself causes the action. (The two arrows leading to the volition could therefore be fused into a single, determining arrow that relates the will and affection together to the action.) Will and affection are both involved in the production of any volition, and through it also in that of actions. The model is ‘distributive’ because the degree to which a given action is free depends on the degree to which it was caused by the will as opposed to the affection; the way causal influence is distributed between will and affection determines the degree of freedom of the action.

This distributive model satisfies the desiderata for a response to the motivation problem: it shows how any volition can be entirely determined in spite of the will’s having its own causal

¹¹⁸ It is sometimes said that Spinoza has a “bundle theory” of the mind, such that the mind is nothing over and above a collection of ideas. It is true that Spinoza rejects a faculty psychology (see E2p48 and scholium) and, of course, the mind is not a substance for him. However, I think it is helpful to distinguish between the idea the human mind *is* and the ideas it *has* (cf. Deleuze 1990, 145-6). Because the idea it *is* has a close connection to its essence, I take the mind’s ideas to be unified in a way that is stronger than the bundle theory reading suggests (although it admittedly depends on how the reading is fleshed out). Similarly, I take it that the will, as an avatar of the *conatus*, provides a sort of unity to the particular volitions. See also Della Rocca (1996, 41-3).

role that is irreducible to that of the affections (or other external forces). Its wider attractiveness depends, naturally, on a substantive account of how Spinoza's conception of freedom operates. A more extensive discussion of what it concretely means to be free according to Spinoza, and how his conception of essences informs his view of freedom is owed. Unfortunately, this is too great a task to be accomplished in this chapter.¹¹⁹ In the next section, I will, however, attempt to tackle a slightly smaller issue, which is to defend Spinoza's apparently highly unorthodox theory as a genuine conception of willing. Doing this will help to show, in section 6, how the theory exemplifies a form of mind-body parallelism.

5. The distributive model as a vector model of willing

Some powers often ascribed to the will are the following:

- the ability to intervene in behavior;
- the ability to act independently of any motivating factors – i.e., to “suspend” their influence (Katsafanas 2016);
- the ability to act or to refrain from acting – i.e., the will is a “two-way power” (e.g., Steward 2012, Lowe 2013, Frost 2020).

In addition, episodes of willing are typically described as occurrent mental states, which means that they are both (i) (typically) conscious and (ii) different from other mental states, such as thoughts, desires, etc.

During the brief discussion of Spinoza's rejection of free will in the introduction, it already became clear that Spinoza rejects most of these features. Spinoza insists that the body is always moved by corporeal causes, and so he denies that the mind can be the exclusive cause of any change in behavior (E1app, E2p35s, E3p2 and scholium). He also rejects the idea of ‘suspension’: given that the mind is being affected, its effects on the mind are necessary and cannot be controverted or undone by the mind (E1a3). Similarly, Spinoza would reject the existence of two-way powers (E1p17s, E2p48).

¹¹⁹ This issue is the focus of a quickly growing literature. See Kisner (2011), LeBuffe (2010a) and James (2020), as well as the collections Kisner and Youpa (2014), Naaman-Zauderer (2019), and Armstrong, Green, and Sangiacomo (2019).

It is more difficult to assess Spinoza's position on the status of volitions as particular mental states. On the one hand, Spinoza is clear that the mind is conscious of the affects (E3p9); as the affects involve changes in the *conatus*, and the mind's *conatus* can be described as the will, it appears to follow that the mind is conscious of changes in the will too.¹²⁰ At the same time, because the will is continually affected, it also follows that the will as such is not an "occurrent" mental state in the ordinary sense of the term, given that occurrent mental states are mental episodes with determinate beginning and end points. (Of course, volitions *can* be described as occurrent.) Just like the intellect, which according to Spinoza is the same thing as the will (E2p49c), the will is one fundamental guise of the mind's essence, and every mode of thinking in the mind can be expressed as a determination of the will.

To somewhat lessen the counter-intuitiveness of this conception of willing, it helps to introduce a distinction due, like the two models of the will considered in the last section, to Paul Katsafanas' recent work on Nietzsche's moral psychology. Katsafanas distinguishes between "triggering" and "vector" models of willing. On the triggering model, the will plays a necessary and sufficient role in the causation of action, which it 'triggers'. While various motivational states can influence the will, they cannot by themselves produce an intentional action without the will's consent. Katsafanas writes that the view is that "[m]otives are capable merely of inclining us to act, and choice alone is causally efficacious" in producing action (Katsafanas 2016, 160). He adds that a common reading of Kant ascribes the triggering model of the will to him (160; see references given at 143n13). Katsafanas notes that it "has become our common-sense conception of action" (160); it is also compatible with many of the powers ascribed to the will that were listed at the beginning of this section.

By contrast, the vector model of willing is less familiar. On such a model, the will is only one determinant of action among others: "the individual's action is determined by the vector of motives, including the will." The will "does not initiate action; rather, it intervenes in a continuous stream of behavior that is prompted by the agent's motives [affections]. It reshapes and redirects this stream, but does not generate it *ex nihilo*." (2016, 160) Such a model does not deny that the will has a causal role independent from the affections – unlike the epiphenomenal and secondary cause models – or that the will's decision determines the action to occur. However, it denies that the affections cannot produce action without the involvement of the will.

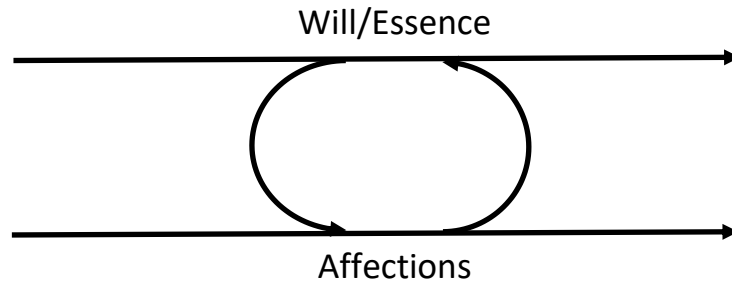
¹²⁰ Garrett (2008) and Marshall (2013) both stress the role of affects in the Spinozan mind's consciousness. See also Chapter 5, section 7.

According to Katsafanas, Nietzsche holds such a vector model of willing. And it is easy to see that Spinoza's distributive theory of motivation also has many features of the vector model. As the model prescribes, Spinoza holds the will to be one causal factor in the production of action and to not be the exclusive source of the agent's behavior.

There is, however, an important respect in which the distributive model is different from the vector model. This is that even in the case of actions that are only free to a degree, Spinoza assigns a greater causal role to the will than is strictly compatible with the vector model. The vector model allows for the existence of actions (under the general heading of 'behavior') in which the will had only a minimal or perhaps no role at all. Katsafanas writes that on the vector model, "conscious thought, episodes of decision, and motives [affections] are all treated as causal forces interacting with one another. None enjoys a privileged position in the production of action." (2016, 162) By contrast, for Spinoza, the will *does* enjoy a privileged position, because it determines the extent to which the action is free, and because there are no actions that do not involve the will. However, his theory of motivation shares with the vector model a basic view of the will as a continually operating factor in the production of action, instead of the more commonsensical but, for Spinoza, deeply mistaken idea that the will is triggered to act by external forces.

6. *A parallelist theory of motivation*

The distributive model of the will depicts agency as a matter of the interaction between an agent and their affections. In any action of a finite being, there is a role for both; and so, both these affections and the activity of the will can be described as ongoing factors in the agent's behavior. These two streams of activity causally interact. As we have seen, the affections causally influence the will. The agent's capacity to change their affections is more indirect, but not less real; Spinoza's ethical theory revolves to an important extent around the attempt to change one's environment so as to undergo more beneficial affections. If, therefore, we take a slightly wider perspective on agency in Spinoza, we see that it exhibits a form of parallelism, as follows:



The arrows here indicate causal influence. The model shows the mutual causal influence that the will and the affections have on each other. The two straight lines indicate the ongoing causal activity of the will and the affections. (Of course, just as in the diagram of the distributive model, when it comes to the production of any action these two causal contributions are only conceptually distinct.)

The diagram is parallelist not because it describes a relation between mind and body – it doesn't, it describes a relation between a thing's essence and its affections – but because of its organization. For there to be a parallelism, two conditions have to be met. First, things are parallel to each other if they stand in the same order and connection, such that corresponding items have corresponding causes in terms of which they can be explained and neither series can be explained in terms of the other.¹²¹ In the present case, every change in affections entails a change in volitions. The agent's will, meanwhile, partly determines the affections to which they will consequently be subject. So the causal structure among states of will and the affections is preserved. Second, parallelism also requires that the items cannot be reduced to each other. This condition is also met. It is true that in the production of any given action, the affection causally influences the will, and so to that extent the will depends on, and hence is less fundamental than, the affection. However, this dependence does not allow for a wholesale reduction of the will to affections. Similarly, the causal contribution of an affection to an action cannot be reduced to that of the will.

One reason this model is interesting is that it shows that there is a role for parallelism in Spinoza's psychology aside from the paradigmatic case of mind-body parallelism. Another is that a similar model could be created for other psychological phenomena. The central feature of the model is that we construe those elements of the agent's mental life that we take to be crucial for freedom in parallel to its other, non-free features, while still allowing for these latter features

¹²¹ For this characterization of parallelism, see Chapters 1 and 2.

to influence the former. The model is not beholden to the specific way in which Spinoza understands the will or the affections. The model allows for realism regarding both of its central elements and for a causal influence or dependence of one on the other, without reducing them. It also allows the 'higher' element to have a strictly lawful role in the production of its effects. The price that is paid for assuming this model is the ascription of essential causal power to this 'higher' element – however, at least in the case of Spinoza's model of willing, this power is not *sui generis* or unnatural, but is in fact shared by everything.

7. Conclusion

Ascribing causal power to the agent's essence allows Spinoza to give a surprisingly rich account of free action and the will, one that does not prevent him from also developing an uncompromising critique of traditional conceptions of free will. This central metaphysical innovation also enables him to accept one of the central intuitions that drive support for libertarian theories of free will, *causal Source*.

Some of the most surprising features of Spinoza's theory derive from the way it incorporates a parallelism. By construing the agent and the will as determinable, but not determined, by external forces, and so as changing in parallel to changes in their affections, Spinoza is able to make sense of how the will can have its own causal power and yet be fully subject to natural determination (by its own essence and external forces).

There appears to be a widespread desire in discussions surrounding free will for a theory that incorporates a substantive (i.e., metaphysically underpinned and nonreductive) conception of freedom while nevertheless being naturalistic. Much of the work in the free will literature deals with the threat that a scientific, mechanistic worldview appears to pose for the existence of free action.¹²² Provided that those interested in such a view are willing to make peace with its radically anti-libertarian implications – admittedly a tall order – the interpretation of Spinoza's thinking about freedom sketched in this chapter appears to meet these desiderata. It holds that freedom to some degree is real and that it is a natural feature of the world.¹²³ Spinoza's views are therefore relevant to these contemporary discussions. If it is fair to say that many

¹²² A poor selection of relevant works might include Anscombe (2005), Nagel (1986, ch. 7), and Steward (2012).

¹²³ For more on Spinoza's naturalism, see Garrett (2008), particularly in connection to a view he calls "incremental naturalism." See also Garrett (2018a) (in which the 2008 paper is reprinted).

discussions surrounding free will are trapped in an oscillation between on the one hand a reduction of agency to a merely physical mechanism and on the other hand the ascription of implausible, libertarian powers of intervening in the natural world, then Spinoza's distributive model of the will seems to provide a way out of this deadlock.

Spinoza's Representationalist Theory of Consciousness

All sudden understanding closely resembles an acute incomprehension.

No. All sudden understanding is finally the revelation of an acute incomprehension.

Clarice Lispector, *The Passion According to G.H.*

1. Introduction

Two of the main philosophical questions raised by consciousness are, first, why are certain physical states accompanied by consciousness (and why specifically with these conscious states); and second, what does consciousness consist in – what is it like to have a certain experience or, to use the philosophical jargon, what is phenomenal quality? This chapter will explore Spinoza's answers to these questions. It argues that Spinoza is a representationalist about phenomenal consciousness: he believes that what an experience is like is determined by its representational content. I will also show that Spinoza's parallelism provides him with a peculiar way to accommodate explanatory gaps in his philosophy of mind. This parallelist strategy plays a role not just in his answer to the first question about consciousness, but also informs the kind of representationalism he defends.

Spinoza's thinking about consciousness has recently received much attention.¹²⁴ But as I will argue, this work leaves the question of what phenomenal consciousness consists in largely unaddressed. Meanwhile, there has also been a tendency to argue that Spinoza's theory of consciousness is underdeveloped, radically implausible, or that he does not have a theory at all.¹²⁵ By contrast, I will argue that Spinoza's metaphysics and philosophy of mind are deeply informed by his thinking about consciousness. At the same time, it is true that he generally emphasizes different features of consciousness than are commonly addressed in contemporary theories, or even in those of Spinoza's own contemporaries. Moreover, his views about specifically phenomenal consciousness remain to a large extent implicit. Reconstructing his views will therefore require a deep dive into Spinoza's views on the mind-body union, the affections, and representation.

Such a long discussion will benefit from having a guiding thread. In the next section, I will introduce this thread, which takes the form of a question: why does Spinoza never present a theory of the sensible qualities? Pursuing this question will allow me to raise and answer the questions about phenomenal consciousness in turn. I begin by considering the first question – why are certain physical states accompanied by consciousness (and why specifically with these conscious states)? – which revolves around the *explanatory gap* between physical and conscious states. I sketch both the contemporary discussion around this gap (section 3) and its early modern ancestor as it occurs in Descartes (section 4). I also show how Spinoza's criticisms of Descartes anticipate the problems associated with the explanatory gap. These problems are that the gap suggests that the relation between mind and matter is just *arbitrary* and that it threatens to open up an *ontological* gap between them. However, Spinoza has several strategies for resisting these conclusions (section 5). In the second half of the chapter, I go on to consider Spinoza's account of phenomenal quality. After introducing the notion of phenomenal quality (section 6) and sketching Spinoza's theory of affection (section 7), I investigate Spinoza's theory of confused ideas and show how it supports a kind of representationalism about phenomenal quality (sections 8 and 9). I conclude by answering the question why Spinoza had so little interest in sensible qualities (section 10).

¹²⁴ Some of the main texts are Della Rocca (2008b, 108-18), Garrett (2008), Nadler (2008), LeBuffe (2010b), Marshall (2008; 2013, ch. 4).

¹²⁵ Skeptics include Wilson (1999b), Bennett (1984), Miller (2007) and Garber (2021). Hübner (2018) questions Spinoza's views on subjectivity and the first-person perspective more broadly. Even some of the work that defends Spinoza's theory of consciousness nevertheless admits that his views may not be fully developed or consistent; for example, see Nadler (2008, 585).

2. Does Spinoza have a theory of sensible quality?

Spinoza's views about the sensible qualities have received almost no attention.¹²⁶ One explanation for this is that it can appear that he has nothing to say. In Spinoza's entire works, there are only a few texts that address the sensible qualities, all of them indirectly. Some of these texts are early and occur in discussions of other philosophers' views (Ep6, CPP 2a7, 2p1-2), whereas the only other one mentions the sensible qualities as instances of a wider phenomenon relating to quality and occurs in a highly polemical context (E1app). This relative silence is interesting and needs to be explained; but when this is done, it reveals that Spinoza in fact has a rich and subtle view.

As the name implies, sensible qualities are qualities experienced through the senses, i.e., qualities like color, smell and taste, heat and cold. The status of these qualities – do they exist in bodies or only in the mind? if the latter, what is their relation to the bodies that seemingly cause them and that they apparently represent? – was much discussed in early modern philosophy, as anyone with a passing familiarity with the period knows. The new science suggested that the material world could be entirely explained in terms of purely quantitative relations, but sensations of sensible qualities appear to have nothing in common with, or in no way to resemble, anything quantitative. This suggested that sensible qualities must be in the mind somehow without existing in reality in conformity with how they were perceived. Yet this raised two major questions. First, what was the cause of the connection between these inner perceptions of sensible qualities and the external objects that caused them? And second, how and why do we come to form representations of sensible qualities?

That Spinoza was aware of these discussions emerges from the early texts in which he mentions sensible qualities. First, there is an offhand mention in a critical discussion of Robert Boyle's attempts to validate a mechanist view of them on experimental grounds. In this context, Spinoza writes that

all the tangible qualities depend only on motion, shape, and the remaining mechanical affections [...]. [This] has already been more than adequately demonstrated by Bacon and later by Descartes. (Ep6; G IV, 25)

¹²⁶ But see Gueroult (1974, 228-9) for a helpful albeit brief discussion.

(‘Tangible’ qualities are qualities like hot and cold that can be perceived through touch. In Descartes’s classification at *Principles of Philosophy* IV, 191-5, tangible qualities are a subset of the sensible qualities.)

It is important to note here that Spinoza is making an ontological claim about what the tangible qualities depend on; he is not talking about how they are experienced. Spinoza’s statement here should be seen in light of his other early discussion of the sensible qualities, which occurs at the beginning of Part Two of his geometrical reconstruction of Descartes’s *Principles*. There he writes:

CPP 2a7: Beyond the sensible qualities there is nothing in body except extension and its affections, enumerated in Part I of the *Principles* [cf. *Principles* I, 66-70].

CPP 2p1: Even though the hardness, weight, and the rest of the sensible qualities are separated from a body, the nature of the body will still remain whole.

CPP 2p2: The nature of body, or matter, consists in extension alone.

Dem.: The nature of the body is not taken away when the sensible qualities are taken away (by 2p1). Therefore, they do not constitute its essence (by 2a2). Nothing remains, then, except extension and its affections (by 2a7), [which (by 2a6) cannot be conceived without extension].¹²⁷ So if extension is taken away, nothing will remain that pertains to the nature of the body, but it will be entirely taken away. Therefore, the nature of body consists in extension alone (by 2a2), q.e.d.

Spinoza here repeats Descartes’s view – which he himself endorses – that the nature of bodies consists in extension alone. He also repeats Descartes’s argument for this view (cf. *Principles* II, 4), which moves from the claim that we can *conceive* the nature of bodies without conceiving of their sensible qualities to the claim that these qualities can be *separated* from bodies, i.e., that it is possible for bodies to exist without their sensible qualities. As will become clear over the next few sections, this move ties Descartes’s view on the sensible qualities in with his substance dualism more broadly. It will also become clear that Spinoza’s rejection of this move as invalid informs his own, parallelist take on the sensible qualities.

Spinoza’s only mention of the sensible qualities in one of his mature writings occurs in E1app and is highly polemical. It also does not address the sensible qualities directly, but only mentions

¹²⁷ The passage in brackets derives from the Dutch translation of the CPP; see Spinoza (1985, 224).

two of them as part of a broader discussion of why certain qualities should not be appealed to in the explanation of natural phenomena. One relevant passage reads:

After men persuaded themselves that everything that happens, happens on their account, they had to judge that what is most important in each thing is what is most useful to them, and to rate as most excellent all those things by which they were most pleased. Hence, they had to form these notions, by which they explained natural things: *good, evil, order, confusion, warm, cold, beauty, ugliness*. And because they thought themselves free, those notions have arisen: *praise and blame, sin and merit*. (E1app; G II, 81; original italics)

In this passage, Spinoza makes reference to two of the central illusions that E1app (and much of the rest of the *Ethics*) attacks, the belief in final causes (or of things in nature ‘happening on human beings’ account’) and that in free will.¹²⁸ In light of this, warm and cold are on this list not so much as representative sensible qualities, but as instances of a broader process by which humans refer the way external bodies affect them to objective properties of these things themselves. Spinoza sees this process as a fundamentally misguided, anthropocentric way to approach nature. Rather than sensible qualities, Spinoza is treating warm and cold here as instances of what might be called ‘affective qualities,’ qualities considered in relation to their pleasant or unpleasant effects on us. This is what emerges from the unrelentingly naturalistic explanation of these qualities that Spinoza goes on to suggest:

For example, if the motion the nerves receive from objects presented through the eyes is conducive to health, the objects by which it is caused are called beautiful; those which cause a contrary motion are called ugly. Those which move the sense through the nose, they call pleasant-smelling or stinking; through the tongue, sweet or bitter, tasty or tasteless; through touch, hard or soft, rough or smooth, etc.; and finally, those which move the ears are said to produce noise, sound or harmony. Men have been so mad as to believe that God is pleased by harmony. Indeed there are philosophers who have persuaded themselves that the motions of the heavens produce a harmony. (E1app; G II, 82)

Here odor, smell and sound are explained in terms of their affective effects on us. Although warm and cold are not mentioned again in the appendix, presumably something similar would

¹²⁸ These illusions can themselves plausibly be said to be part of the operations of our consciousness according to Spinoza, as is argued by Deleuze (1988, 20). Cf. Marshall (2013, 160-5) and Melamed (2017) on the illusion of free will. I leave this aside here, however, as it is not directly relevant to the sensible qualities and to phenomenal consciousness.

go for them. (Perhaps Spinoza would have said that they are both pleasant relative to the body's temperature within certain limits, and unpleasant beyond them.) In the end, Spinoza writes, these qualities are only "beings of the imagination" (*entia imaginationis*; G II, 83) and as such are unfit to figure in a scientific explanation of the world.

Spinoza's mention of some sensible qualities as instances of affective qualities is telling. It suggests that he is interested in these qualities insofar as they affect us – insofar as they are experienced as pleasurable or not. It also shows that he makes a clear separation between the subjective effects of such qualities and a scientific way to explain nature. Finally, his designation of them as *entia imaginationis* is suggestive. At the same time, E1app can hardly be said to address the sensible qualities in full. This is because, in focusing on affectivity, Spinoza is shifting his attention away from a broader conception of phenomenal quality. It is the latter that plausibly motivates interest in the sensible qualities (and that continues to drive interest in phenomenal quality today). Granted that colors can be discussed in terms of their more or less pleasant effects on our experience, there remain the questions of what distinguishes experienced red from experienced yellow or blue and of what experienced color itself consists in. And so the question of what Spinoza believes phenomenal quality consists in, and what its ontological status is for him, remains open.

Disjointed as Spinoza's mentions of sensible qualities may seem, together they raise many of the central issues that arise in relation to them. This by itself suggests that Spinoza was aware of these issues and had thought them through. The coming sections will first detail Spinoza's thinking about the link between conceivability and possibility (sections 3-5). Afterwards, from section 6 onwards, I will try to find Spinoza's answers to the questions that E1app leaves open.

3. *The modern explanatory gap*

One of Descartes's arguments for substance dualism employed a move from conceivability to possibility that put down deep roots in subsequent thinking about the nature of mind and body. Exploring connections to the contemporary debate here will provide some helpful terminology and context with which to approach Spinoza's views. First among these is the notion of an explanatory gap.

In its broadest sense, the 'explanatory gap' (Levine 1983) refers to the sense that there is a 'gap' of some sort between any physical explanation of a mental state and what it is trying to explain.

A physical explanation of an experience that involves color may, for example, invoke particular features of the retina and the brain. Yet no amount of information about these physical processes seems sufficient to explain why the experience feels the way it does – why, in the terms used by present-day philosophers, it has the ‘phenomenal quality’ it does.

The sting of the problem is in its seeming possible to conceive the physical state in the absence of the phenomenal quality, and vice versa. We can conceive the neural state that underlies my perception of an apple, but accompanied by the experience of an orange, or even by no experience at all. Meanwhile, it also seems possible to conceive the experience, but accompanied by an entirely different physical structure, or none at all. In the light of these apparent facts about what we can conceive, the question arises *why* a particular physical (or mental) state should be accompanied by the particular mental (or physical) state that it in fact is accompanied by. Why this instead of any other? And more extremely, why is it accompanied at all? The explanatory gap asks for an *intelligible connection* between these two states: for an answer to these why questions.

Apart from a sense in which physical explanations are incomplete or, as Levine (1983, 357) puts it, not “fully explanatory” so long as the gap remains, there have also been attempts to develop the explanatory gap into an *ontological gap*, thereby showing that physicalism is false. These arguments have a common structure. They begin by stating that a certain mental (or physical) phenomenon is conceivable without its accompanying physical (or mental) phenomenon. From this, they infer that it is possible for the former to occur without the latter. This possibility is then used to undermine some physicalist thesis. Schematically:

The conceivability argument:

- (1) x can be conceived without y .
- (2) If x can be conceived without y , then it is possible for x to occur without y . (I.e., there is a possible world that contains x but not y .)
- (3) If it is possible for x to occur without y , then some physicalist thesis P is false. (E.g., ‘ $x = y$ ’ is false, or x does not necessarily depend on y .)

The details of the argument will differ depending on what x and y are, as well as depending on how P is understood. But for example, if physicalism is the claim that mental state x = physical state y , then the argument suggests that this is false, because identity is generally understood to be necessary and the conceivability argument just given argues that there are possible worlds in which $x \neq y$. A version of this argument was famously used by Kripke (1980). If physicalism

instead is the view that the physical is fundamental, such that a “minimum physical duplicate” (Jackson 1998) of this world – a world that contains all the same physical states as the actual world – would be a complete duplicate of this world (including its mental and other higher-order states), then a conceivability argument can be set up that starts by conceiving a minimum physical duplicate world that does not contain mental states, and argues from there to the falsity of physicalism.

The responses to this argument have fallen into two main groups, depending on which of its premises is denied. Some physicalists deny its first premise. They argue that, after all, it is not really true that mental states can be conceived without physical states. These physicalists – ‘Type A’ physicalists in the taxonomy of Chalmers (2003) – in effect argue that the explanatory gap can be closed and there is an intelligible connection between mental and physical states after all. However, most contemporary physicalists are ‘Type B’ physicalists who deny the second premise, rejecting the claim that the conceivability of x without y implies the possibility of their separate existence. We’ll see traces of both of these moves in Spinoza’s response to an early modern ancestor of the conceivability argument. Let’s now first turn to this early modern version of the conceivability argument itself.

4. *The conceivability gap in Descartes*

It is often pointed out that Locke discusses an early form of this problem, the form that is now known as the ‘inverted qualia’ problem (see *Essay* II.xxxii.15). But before Locke, the same problem is, famously, a fundamental issue in Descartes’s metaphysics. In the *Meditations*, Descartes of course argues explicitly that it is possible to conceive the soul without the body – this is an important result of his *cogito* argument. Descartes also argues that a physical body of similar complexity to that of a human being may nevertheless not have a soul – this is the case if the body is that of a non-human animal – in which case it is possible to conceive such a body in purely physical terms (see *Discourse on Method* V). In the Sixth Meditation, he uses these claims to argue that the soul and the body are ‘really distinct,’ where the distinctive feature of really distinct substances is that they are separable: God could bring it about that the soul (a thinking substance) existed apart from the body (an extended substance) and vice versa. He writes:

I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can

clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated at least by God. The question of what kind of power is required to bring about such a separation does not affect the judgement that the two things are distinct. Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing. It is true that I may have (or, to anticipate, that I certainly have) a body that is very closely joined to me. But nevertheless, on the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing; and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it. (CSM II, 54; AT VII, 78)

Descartes's argument can be reconstructed as follows:

The separability argument:

- (1) I can clearly and distinctly conceive my soul apart from my body.
- (2) If I can clearly and distinctly conceive x apart from y , then God is capable of separating them.
- (3) If x and y can be separated, then they are really distinct.
- (4) Therefore, the soul is really distinct from the body.

(Descartes's argument incorporates further claims about the power necessary to separate distinctly conceived things and about the essence of the soul and the body, but this reconstruction will do for present purposes.)

The argument exhibits the structure of the conceivability argument. It begins with a claim that it is possible to conceive one thing without conceiving something else. Descartes then brings in a principle that enables him to deduce a claim about possibility from this premise. This principle is that God is capable of bringing about whatever can be clearly and distinctly conceived. For Descartes, this principle follows from the fact that God is not a deceiver: if it were possible to clearly and distinctly conceive something that nevertheless could not possibly exist, this would mean that God gave us a faculty of knowledge that was mistaken, because it represented something impossible as possible. Through his claim that two things are really distinct if they can be separated, Descartes is able to reach his desired conclusion: the real distinction between soul and body.

For Descartes, then, mind and body are conceived entirely independently and are, in fact, distinct substances. Nevertheless, they are also intimately “united” in living human beings, which he takes to entail the existence of various psychophysical relations, causal relations especially. As anyone even slightly familiar with Descartes knows, this leads to intractable problems relating to his ability to give an account of these various relations. One fundamental worry – known as the Heterogeneity Problem – is how it is possible for radically distinct substances to stand in any relations at all.¹²⁹ For there to be a relation between two things, it is natural to think they must have something in common. But thinking and extended substances don’t have anything in common. So how can they be related? This worry is already pressed in Descartes’s lifetime by Gassendi and Princess Elisabeth.¹³⁰ In his replies to them, however, Descartes shows himself relatively unfazed. It appears he simply rejected the main assumption of the problem, that different substances must have something in common in order to be related.¹³¹

Spinoza also developed a number of fundamental criticisms of Descartes’s account of the mind-body union, and particularly of the infamous claim (in the *Passions of the Soul*) that the pineal gland is the site of mind-body interaction. In the preface to Part Five of the *Ethics*, Spinoza criticizes Descartes for being unable to explain “how many degrees of motion the mind can give to that pineal gland, and how great a force is required to hold it in suspense” – a worry similar to the Heterogeneity Problem. His fundamental concern, however, is not so much about the *possibility* of mind-body relations as about their *intelligibility*. As he writes:

What, I ask, does he [Descartes] understand by the union of mind and body? What clear and distinct concept does he have of a thought so closely united to some little portion of quantity? Indeed, I wish he had explained this union by its proximate cause. But he had conceived the mind to be so distinct from the body that he could not assign any singular cause, either of this union or of the mind itself. Instead, it was necessary for him to have recourse to the cause of the whole universe, i.e., to God. (E5pref; G II, 279–80)

Spinoza points out that on Descartes’s view, there is no “proximate cause” of the mind-body union – the only way to explain why *this* body is joined to *this* soul is by reference back to God’s decision to pair them up. There is no way from understanding the nature of the mind or the

¹²⁹ The term derives from Richardson (1982). See also Rozemond (1999).

¹³⁰ For Gassendi’s objections, see *Fifth Objections* (CSM II, 237; AT VII, 341). For Elisabeth’s questions, see her letters of 6 May, 10 June, and 1 July 1643 (in Shapiro 2007).

¹³¹ See esp. the letter to Clerselier of January 1646 (CSM II, 275; AT IX, 213). For discussion, see e.g. Della Rocca (2008a).

body themselves to understanding why they are coupled. A different way to put this is that they are not *intrinsically* but at best only *extrinsically* related: their relation is not due to themselves but to a third factor, God. For Spinoza, this is unacceptable. On his own view, the mind-body union is explained through the fact that the mind necessarily represents the affections of the body (E2p12, E2p13, E2p21s). His criticism in E5pref suggests that he takes this representational connection to constitute the “cause” of the union.¹³²

5. *Intrinsic connection without conceptual connection: Spinoza's conceptual barrier*

This shows that Spinoza was well aware of the problems with intelligible connections that Descartes's dualism entailed. Without intelligibility, on Descartes's picture the mind-body union comes to seem simply *arbitrary*. What is to be done in response to this situation? One option is to find an intelligible connection after all: to close the explanatory gap. But Spinoza develops a different reaction. He instead finds a way to hold that the mind-body connection is not arbitrary but (intrinsically) necessary, even though it remains true that there is no intelligible connection between a particular mind and the body it is related to.

Spinoza begins by purifying Descartes's ‘soul’ and ‘body’. Descartes takes it for granted that through the *cogito* we do not just conceive thinking itself but the *soul*, a thinking substance. While Spinoza agrees that we can conceive the ‘thinking substance’ independently from the ‘extended substance’, he no longer identifies these substances with mind and body. These latter are not substances themselves but only modes of substance. Furthermore, thought and extension are, for Spinoza, attributes of substance. With these qualifications in place, however, Spinoza accepts that thought and extension are conceived independently from each other. He expresses this by saying that they are “conceived through themselves” (e.g. E1p10), “have nothing in common” (E1a5, E1p2), and do not “involve” each other (e.g. Ep2; G II, 7-8). Spinoza clarifies in E2p49dem that “to say that A must involve the concept of B is the same as to say that A cannot be conceived without B.” So, if thought and extension do not involve each other, then extension can be conceived without thought and vice versa. In fact, when Spinoza states that attributes are “conceived through themselves,” he appears to mean that the concept of any particular attribute is independent of any other concept (see, e.g., E2p6dem).

¹³² For more on the mind-body union in Spinoza, see Jaquet (2021). For the connection between it and his mind-body parallelism, see Chapter 2, section 7.

Unlike the concept of a particular mode of extension (such as a triangle) which requires the concept of extension to be thought, the concept of extension itself does not require any other concept to be thought. The same is true of any other attribute. Spinoza concludes that no attribute can be conceived through any other. In the literature, this is known as Spinoza's 'conceptual barrier' between the attributes (see esp. Della Rocca 1996). Descartes's claim, then, suitably abstracted and generalized, is endorsed by Spinoza as a basic tenet about the conceptual relations between attributes.

However, in spite of their independent conceivability Spinoza does not think it is possible for attributes to exist independently of each other. Instead of belonging to separate substances, attributes are only "what the intellect perceives of substance, as constituting its essence" (E1def4). For Spinoza, the only substance that exists is God, who has all the attributes (E1def6). The concepts of different attributes are, therefore, only different (independent) conceptions of God's nature.¹³³ Because God exists necessarily (E1p11), it is not possible for one attribute to exist without any other. In other words, the independent conceivability of attributes does not entail their possible separation. Similarly, although mind and body can be conceived independently through their respective attributes of thought and extension, it is not possible for the mind to be separated from the body. Because the attributes exist necessarily, and particular modes 'follow' from them in a necessary order, it is not possible for a mode of one attribute to occur without its corresponding mode in another attribute. They are necessarily metaphysically parallel (E2p7s).

On the most straightforward way of reading him, therefore, Spinoza blocks the inference from conceivability to possibility, similarly to current-day Type-B physicalists. However, there is also a sense in which he denies that thought can be conceived without extension in the first place. As Garrett (2009) has pointed out, the claim '*x* can be conceived without *y*' is ambiguous. On one reading, it means that it is possible to *conceive* *x* without *conceiving* *y*; this is the reading we have been considering so far. But it can also mean that it is possible to conceive *x* *without* *y*: i.e., to conceive *that* *x* \neq *y*. In context, this would mean conceiving an extended substance that is not a thinking substance, for example. Here there is good reason to think that Spinoza rejects that this is conceivable. While attributes can be conceived independently of each other, it is not

¹³³ My interpretation of Spinoza here naturally supports a recent trend in the literature that sees the attributes as only rationally distinct. For different articulations of this view, see Shein (2009), Melamed (2018) and Lin (2019, ch. 4).

possible to conceive a substance that has fewer than all the attributes. This results from the way in which substances are individuated. For Spinoza, things are individuated in terms of their essences (E2def2; cf. Della Rocca 2002). As we saw, he also holds that a conception of an attribute is a (true)¹³⁴ conception of what constitutes the essence of a substance (E1def4) and that God possesses all the attributes (E1def6). The result is that a merely extended substance – call it E – would share its single attribute with God. Now, Spinoza is clear that it is not *possible* for substances to share attributes (E1p5). But he arguably believes it is not even *conceivable*. For conceiving the attribute of extension ought, by E1def4, to be conceiving the essence of E. But it isn't, because conceiving extension is also a way of conceiving the essence of God and so doesn't allow for distinguishing E's essence from that of God. So E cannot be individuated in terms of its attribute. So how can it be individuated? The only way this is possible is by conceiving something other than E, and then negating it of E. For example, we can conceive the attribute of thought and then conceive that E does not have thought. The fact that E does not have thought sets it apart from God. But to distinguish it in this way we need to conceive E in terms of something 'outside of' itself, contrary to Spinoza's definitional claim that substances are conceived through themselves (E1def3). The argument can be repeated for a substance with any finite combination of attributes. Arguably, therefore, none of these substances are conceivable.¹³⁵

Spinoza's view, then, is that although mind and body can, in one sense, be conceived independently of each other, they are nevertheless necessarily connected in a sense stronger than on Descartes's view. Mind and body, for Spinoza, do not have a *conceptual* connection, if by this is meant a connection that allows you to deduce the concept of the one from that of the other. They are independently conceived and it is possible to have the concept of one of them without that of the other. Still, their connection is *intelligible* in the broader sense that there is an intelligible reason why they are connected. (There has to be, or Spinoza would be violating his commitment to the Principle of Sufficient Reason.) In fact, there appear to be at least two reasons: one is God's necessary existence as a being with infinite attributes, and the other is the mind's being the adequate representation of the body. To use slightly anachronistic terms, their relation

¹³⁴ I am assuming here that so-called 'subjectivism' about the attributes is false. For two recent discussions, see Shein (2009) and Melamed (2018, 94-5).

¹³⁵ For further discussion, see Noorloos (2021).

is known as a synthetic and not an analytic *a priori* truth. Their relation is, furthermore, intrinsic, in the sense that it follows from the things they are that they are connected.

It is now possible to return to the first opening question: why are certain physical states accompanied by conscious states (and why specifically with these conscious states)? So far, I have given Spinoza's answer to a part of this question – the part outside of the brackets. It turns out that Spinoza has a general strategy to argue that the relation between conscious states and physical states is necessary, without however providing a conceptual connection in terms of which it would be possible to deduce *a priori* which physical state corresponded to any mental state (or vice versa). Still, the question why particular physical states are accompanied by particular conscious states has not been answered yet. To deal with this question, it is first necessary to move on to the second question and consider Spinoza's account of phenomenal consciousness.

6. *Phenomenal consciousness as what it is like*

In the coming sections, I will argue that Spinoza has a representationalist theory of consciousness. But before doing so, it is important to give a slightly more extensive characterization of the kind of consciousness that will be at issue. There are two reasons for this, one historical and the other relating to prior work on Spinoza's theory.

It has often been pointed out that the modern concept of consciousness was still in the process of crystallizing in the 17th century. The English term 'consciousness' and its cognates in Romance languages derive from the Latin *conscientia*. As the term suggests, its original meaning was closer to 'conscience'; the Scholastics took it to refer to a moral capacity to pass judgement on one's own actions, and the term is occasionally still used in this manner in Spinoza's time.¹³⁶ By then, the concept had developed to the point where it could indicate a direct awareness of one's mental acts, but even so it was often used unsystematically. For example, consider Descartes' definition of thought in the *Principles* as "everything which we are aware [*consciis*] of as happening within us, in so far as we have awareness [*conscientia*] of it."¹³⁷ The fact that thought

¹³⁶ For the history of *conscientia* as a moral term, see Davies (1990), Lewis (1960), Thiel (2011, 7-11), and Jorgensen (2020, §1). Spinoza uses *conscientia* as 'conscience' in his affect of *conscientiae morsus* (E3p18s2), translated by Curley as 'remorse' but literally 'the bite of conscience.'

¹³⁷ *Principles of Philosophy*, I, 9 (CSM I, 195; AT VIIa, 7). See also definition 1 in the geometrical exposition at the end of the *Second Replies* (CSM II, 113; AT VII, 160).

is defined in terms of consciousness suggests that consciousness plays an important role in Descartes' philosophy.¹³⁸ At the same time, this is a definition of *thought*, not of consciousness; consciousness as such is not defined by Descartes in the *Principles*.¹³⁹

Some authors have suggested that, because there is no fixed terminology for expressing what consciousness is in this period, there is also little prospect for discovering theories of consciousness in 17th-century authors. This argument is also made by reference to Spinoza's inconsistent.¹⁴⁰ However, the absence of a fixed *terminology* for consciousness does not entail that philosophers in this time may not have had important things to say about the *phenomenon*.¹⁴¹ Spinoza certainly does. At the same time, it does imply that there are clear limits on the usefulness of attempts to derive Spinoza's views about consciousness from his use of the terms *conscius* and *conscientia* alone. These terms likely mean different things in different contexts; moreover, Spinoza should be expected to make important observations about consciousness without using these terms.¹⁴² This puts a burden on the interpreter to extract and systematize his relevant remarks.

It also means that it is important for the interpreter to be self-reflexive and upfront about their own understanding of consciousness. Three different aspects of consciousness are sometimes distinguished in work on early modern theories of consciousness: those relating to *awareness* (including phenomenal character and subjectivity), *self-reflexivity*, and *intentionality* (including the notion of having a perspective on the world).¹⁴³ Clearly, these categories intersect and overlap in various ways, but they provide a rough way of dividing up the terrain. While I will tangentially address all three of these aspects, my main emphasis will be on the notion of awareness, and particularly on the question of what, if anything, the *phenomenal character* of experiences consists in for Spinoza.

¹³⁸ See also Simmons (2001).

¹³⁹ Miller (2007, 205). He adds that the term *conscientia* never occurs in the *Meditations*, while *conscius* appears only once.

¹⁴⁰ See Wilson (1999b), Bennett (1984, 188-91), Miller (2007), Hübner (2018) and Garber (2021). Wilson, Bennett and Hübner treat the absence of a theory of consciousness as a problem for Spinoza; Miller and Garber go further and suggest that Spinoza neither *has* nor *needs* a theory of consciousness.

¹⁴¹ This is also pointed out by Perler (2014).

¹⁴² For this reason, Garber's (2021) recent attempt to interpret Spinoza's views about consciousness on the basis of his use of the terms *conscius* and *conscientia* appears to me to run into serious methodological problems. Some important texts for Spinoza's theory do not use these terms at all, but speak instead of 'feeling' (E2a4), 'perception' (e.g., E2p12) or even 'knowledge' (e.g., E2p19).

¹⁴³ For instance, see Heinämaa, Lähteenmäki, and Remes (2007b, 10-23) and Jorgensen (2020, §2.2).

The notion of phenomenal character is itself contested, however, and so it requires some more explanation. With this term, following Thomas Nagel's influential article (Nagel 1974), I intend to refer to *what it is like* to have an experience. What it is like to have a certain experience is something over and above what an exhaustive physical description of the experience can reveal; for example, what it is like to see red is inexplicable solely in terms of any set of physical facts about this experience. Nagel was concerned to argue that this renders proposed "reductionist" explanations of subjective experience in terms of physical facts unintelligible: he argued that "we have at present no conception of what an explanation of the physical nature of a mental phenomenon would be" (Nagel 1974, 436).¹⁴⁴ He used the question 'what it is like' to refocus attention on issues related to subjectivity and the first-person point of view in the explanation of mental phenomena. To understand what it is like to experience something involves taking up the perspective of the subject experiencing it; in other words, describing this experience from their point of view. I will use the terms 'what it is like,' 'what-it's-like-ness,' 'phenomenal character' and 'phenomenal quality' interchangeably, and take them to refer primarily to the 'feel' of experiences but also to their subjective character in a broader sense.

It is important to see what does and does not follow from an understanding of consciousness as what-it's-like-ness. In particular, it does not imply any stance on the existence of *qualia*. Qualia – also called "raw feels" – are sometimes taken to be simply synonymous with phenomenal character itself. However, here the term will be used in a narrower sense. On this understanding, qualia are the "intrinsic, consciously accessible features [of sensory experiences] that are non-representational and that are solely responsible for their phenomenal character" (Tye 2018). Qualia, considered like this, are *intrinsic* to specifically *sensory* experiences and are *contentless* (non-representational). While someone who adopts the what-it's-like-ness understanding of consciousness *may* also commit themselves to the existence of qualia, it should be clear that this is a further commitment not necessitated by the notion of phenomenal quality as such. This is a good thing, because the notion of qualia arguably is a fairly recent invention, so that it is an open question whether it can be ascribed to early modern authors.¹⁴⁵ So long as this notion of qualia is set aside, however, emphasis on phenomenal character does not have to be anachronistic. While the phenomenal character of a state naturally involves its 'feel,' it does not

¹⁴⁴ As has already emerged, this point resonates with Spinoza, whose conceptual barrier between the mental and the physical also prevents there from being physical explanations of mental states.

¹⁴⁵ Crane (2019) separates the notion of qualia from that of phenomenal character and argues that the first originates in early-20th-century philosophy.

follow that any interpretation that considers these feels is *ipso facto* committing itself to the existence of qualia.¹⁴⁶ On the interpretation defended here, Spinoza is a representationalist about phenomenal character: he explains phenomenal character in terms of what and how experiences represent things. This in itself shows that he has no use for qualia as I understand them here.

The value of distinguishing phenomenal character from other understandings of consciousness becomes clear when previous interpretations of Spinoza's theory are considered. Surprisingly, recent work on Spinoza's theory of consciousness has not systematically investigated phenomenal quality. Instead it has been mainly focused on determining the *extent* of consciousness, both within minds and in nature. Proposals for how to do this, however, have tended to be presented as accounts of consciousness as a whole, hence as also explaining phenomenal character. But it is not clear that they succeed in this. For example, Steven Nadler argues that, for Spinoza,

consciousness is simply the reflection within thought of a body's internal relations in extension (just as the mind itself is the reflection in thought of the body's basic reality). [...] [F]or Spinoza, human consciousness just *is* the greater complexity of the human body as this is manifested under the attribute of Thought. (Nadler 2008, 589, 591; his italics)

This is suggestive as an account of *sentience* – that is, of the ability of being able to consciously respond to reality at all. However, that it fails as an account of phenomenal character is clear from considering that two different conscious states – say, that of seeing green and that of seeing red – may be equally complex, yet obviously differ in phenomenal character.

Similarly, Don Garrett's claim that Spinoza can account for differing degrees of consciousness “simply by identifying degrees of consciousness with degrees of power of thinking” (2008, 23) can account for sentience. Perhaps it can also account for what is now known as ‘access consciousness’ (the kind of consciousness that consists in being available for reasoning and action; Block 1995): it says that ideas with more power of thinking are the ones accessible to the mind. But it is hard to see how it can account for phenomenal character, because here too it is easy

¹⁴⁶ Marshall (2008) argues that Spinoza does account for the “phenomenal feel” of mental states, which he locates in their affective character, i.e., the specific way in which they exemplify joy or sadness. However, he also argues that for Spinoza, the affective feel of an idea is identical with that idea: “the affective feel and the idea are in fact one mode of thought, involving each other essentially, just as the volition and the idea are.” (2008, 20) Because the affects are “cognitive,” i.e., have content, their ‘feels,’ being identical to them, are not qualia.

to think of experiences with equal power of thinking that nevertheless differ in phenomenal character.¹⁴⁷

A view that comes closer to the one that will be defended here is that by Eugene Marshall (2008; 2013, 118-20). Marshall proposes that consciousness consists in “affectivity,” or an idea’s being involved in the mind’s *conatus*, its striving to persevere in its being (see E3p6 and below). While closely related to Garrett’s view – as Marshall acknowledges (2013, 117) – I think this proposal improves on Garrett’s by clearly connecting the consciousness of an idea to how it affects us, and hence to the way it is experienced.¹⁴⁸ In this way it connects to the notion of phenomenal character under discussion here. Moreover, because red and yellow bodies affect the human body differently, Marshall is in a position to explain phenomenal quality in terms of affectivity. Nevertheless, Marshall does not develop this line of thought in his own work, and he also does not have much to say about the role of representation in consciousness. I see the interpretation I develop here as compatible with his.

These proposals, then, all have more or less serious limitations as accounts of phenomenal character. Accordingly, in giving my own account of Spinoza’s theory of phenomenal quality in the next sections, I will shift attention from their accounts by focusing in particular on the role of representation – particularly *mis*representation, or ‘confusion’ – in Spinoza’s theory. This will be necessary to support my contention that Spinoza has a ‘representationalist’ theory of consciousness. In the coming sections, I will first give an overview of Spinoza’s theory of affection (section 7); then I will introduce representationalism (section 8); I will argue that there is a sense in which, even though phenomenal qualities are ‘real’ for Spinoza, they nevertheless cannot be explained (section 9); before concluding by suggesting an explanation for his silence about the sensible qualities (section 10).

¹⁴⁷ See Garber (2021, 316-7), for further critical discussion of Nadler’s and Garrett’s proposals. LeBuffe (2010b) raises a worry similar to mine relating to what claim the ‘degrees of consciousness’ notion has to adequately represent phenomenal consciousness. In response to LeBuffe, Garrett appears to admit that he was not attempting to address phenomenal consciousness: he separates the power of ideas from their “sensory vivacity” and claims he was explaining only the former (2018b, 422). It’s not entirely clear how sensory vivacity relates to phenomenal consciousness, but presumably an account of the latter should at least address the former to some extent. So by this light, Garrett’s interpretation does not address phenomenal consciousness as I understand it.

¹⁴⁸ For their own discussions of how their views are related, see Marshall (2013, 114-7, 130) and Garrett (2018b, 423n5).

7. *A sketch of Spinoza's theory of affection*

Spinoza believes that human consciousness fundamentally revolves around our affections and appetites. As we saw, the mind and the body are modes of the attributes of thought and extension, respectively (E2p13c). They stand in a representational connection because the mind is the idea that has the body for its object (E2p13). This connection is fundamental to Spinoza's view of the mind and its operations. In the first place, it underlies Spinoza's notorious claim in E2p12:

Whatever happens in the object of the idea constituting the human mind must be perceived by the human mind, *or* there will necessarily be an idea of that thing in the mind; i.e., if the object of the idea constituting a human mind is a body, nothing can happen in that body which is not perceived by the mind.

This states that the mind “perceives” anything that “happens in” its body. This claim has puzzled commentators, because it appears to make the profoundly counterintuitive statement that the mind is, at some level, aware of the complete state of the body – down to “the smallest molecules, the atoms, the electrons of each of the millions of cells,” as Margaret Wilson (1999b, 129) has complained. Don Garrett (2008, 19-20) makes an important corrective to this reading by noting that “happens in” in E2p12 should be read to refer only to the affections of the body, and not to all of its states. On his reading, “in” denotes not the relation of spatial containment but Spinoza's metaphysical relation of “inhering” in something. Modes inhere in attributes, and affections (which are second-order modes) inhere in the mode that is the body. Garrett goes on to add that these affections are changes to the body's state. Spinoza's sketch of a physical theory after E2p13 makes clear that parts of the body may change or be replaced without any change to the functions these parts perform.¹⁴⁹ If that is so, then there can be many changes to the body that are not consciously experienced.

Much more could be said about this worry. And other issues arise as well. Spinoza is explicit that *every* body has a corresponding mode of thought that functions as its ‘mind’, which raises the question what mindedness consists in for plants, rocks and buildings (E2p13s). Commentators by and large agree that Spinoza needs a theory of “selective consciousness” to account for *which* minds, and which *ideas* in minds, are conscious, and have produced several proposals

¹⁴⁹ See Lemmas 5 and 7 after E2p13s. See also the classic article by Jonas (1965). See Lloyd (2017) for a short discussion that also emphasizes the role of physiology for Spinoza's theory of consciousness.

for such theories.¹⁵⁰ Here I will not produce yet another proposal, as my goal is instead to get clear on what consciousness consists in according to Spinoza. I will leave the question of selective consciousness for another occasion.¹⁵¹

From the perspective taken here, what is important about E2p12 is that it suggests that consciousness for Spinoza fundamentally revolves around awareness of the body's affections. E2p12 says that the mind is aware of all of the affections of the body. It resonates with Spinoza's claim in E2a4 that "We feel that a certain body [NS: our body] is affected in many ways." Note that Spinoza does not say that we are aware of the *body* – he says we are aware of the body's *affections*. Statements like these, which occur right when Spinoza is just beginning to set out his concept of the human mind early in Part Two of the *Ethics*, suggest that the human mind's conscious awareness of the affections is the most fundamental feature of human consciousness, in terms of which its other features must be explained.

What does this awareness consist in? In particular, what sets it apart from God's own ideas of these affections? Per E2p7c, God has an adequate idea of everything that exists, and this includes bodily affections. Does this mean that the mind's awareness of the affections is identical to God's? It is sometimes suggested that something like this must be right, with the result that human subjectivity appears to disappear from Spinoza's philosophy.¹⁵² However, I read Spinoza as making a fundamental distinction between the God's-eye view of the human mind and

¹⁵⁰ The term originates in Bennett (1984, 191). All the recent texts (see fn. 124) address this issue. They all appeal to Spinoza's claim in E2p13s that different minds can be conscious to different "degrees". While LeBuffe (2010b) agrees that Spinoza needs a theory of selective consciousness, he has some important reservations about whether the notion of 'degrees of consciousness' can help to attain such a theory.

¹⁵¹ Still, I cannot help briefly making a suggestion. It seems clear that Spinoza happily calls beings conscious that have only a very limited mental life. He does this on the basis of clearly stated principles, mind-body parallelism especially. However, nothing would prevent him from also recognizing a more demanding notion of consciousness, one that is closer to how the term is commonly used. Two requirements to make of this stronger notion could be that beings conscious in this sense must possess sensory organs (or, perhaps, a nervous system) and memory (roughly, a bodily structure for retaining past affections; cf. E2p18s and the references to "traces" at Postulate 5 after E2p13 and E3post2). This would allow Spinoza to exploit his own theory of the bodily functions that underlie consciousness to rule out many creatures as unconscious in this stronger sense. This leaves worries about the counter-intuitiveness of the wide extension of the weaker notion of consciousness and the apparent lack of a clear cut-off point between weakly and strongly conscious creatures (although this depends somewhat on how clearly we can identify the bodily structures required for strong consciousness). But given Spinoza's rationalist inclinations, I don't see how the former consideration could have much weight for him; and the second point may, instead of being seen as an objection, be taken as just a fact about the common notion of consciousness: the fact that it has vague boundaries. (This in turn reinforces Spinoza's choice for weak consciousness as the better grounded, non-vague notion.) This suggestion needs to be developed further, but it appears promising as a way of fulfilling one requirement of a theory of selective consciousness, that of determining which minds are conscious.

¹⁵² See in particular Hübner (2018) for some remarks in this direction. See also Miller (2007) and Garber (2021).

the point of view of the embodied human subject.¹⁵³ Spinoza's theories of both consciousness and subjectivity rest on this distinction. For Spinoza, the mind just is the adequate representation of the body; this claim is part of his mind-body parallelism (E2p7 and scholium, E2p13). But we may distinguish between the idea that we *are* and the ideas that we *have*.¹⁵⁴ We are normally only conscious of the latter. And even of these ideas, our conscious awareness is limited.

To see this, consider E2p17s, where Spinoza distinguishes between “the idea of, say, Peter, which constitutes the essence of Peter's mind, and the idea of Peter which is in another man, say in Paul.” The idea that ‘constitutes the essence of Peter's mind’ is a perfect representation of Peter's body; it is the mode of thought parallel to Peter's body. This idea is in God's mind but not in Paul's. Paul's idea of Peter, by contrast, originates in the imagination. Imagination, for Spinoza, involves the mind considering things on the basis of “images,” which are “the affections of the human body whose ideas present external bodies as present to us” (E2p17s). Paul's idea of Peter, therefore, is not parallel to Peter's body, but only to the sensory affections of Paul's body produced by Peter. Note, however, that Peter's idea of *himself* will also, much like Paul's idea, be based in affections. Peter does not have the same understanding of himself that God has. This is confirmed by E2p23: “The mind does not know itself, except insofar as it perceives the ideas of the affections of the body.” E2p23 belongs to a string of propositions that relate to the mind's ability to reflexively form ideas of its ideas.¹⁵⁵ Spinoza is saying here that the only way the mind can “know” itself is through this process of reflecting on its affections. The mind does not know itself immediately and the perfect representation of Peter's body is not itself in Peter's mind.

E2p17s introduces a fundamental distinction between two ways of understanding the world. To use Spinoza's own terms, we can understand the world either according to the “order and connection of the affections of the human body” or the “order of the intellect” (E2p18s; see Wilson 1996, 103). I will speak of ‘affective’ and ‘intellectual’ ideas, where these are ideas

¹⁵³ Renz (2018) and Renz and Hutchins (2021) make the stronger claim that human subjectivity as such is unavailable to God.

¹⁵⁴ Cf. Deleuze (1990, 145-6).

¹⁵⁵ See E2p20-23. This theory of ‘ideas of ideas’ has been considered to be the locus of Spinoza's theory of consciousness, but faces serious objections when cast in that role. For defenders of the theory, see Curley (1969, 128), Rice (1990), Martin (2007). For criticism, see Wilson (1999b, 135-8; Wilson criticizes it while also ascribing it to Spinoza), Bennett (1984, 184-91), Nadler (2008, 581-5), Marshall (2013, 108-11). Although it cannot by itself be the whole of Spinoza's theory, however, it is still clearly a part of it as his account of the mind's power to reflect on itself.

formed according to the order of the affections and the order of the intellect, respectively. Note that these are relational definitions and that one and the same idea can be 'affective' in my mind and 'intellectual' in God's.

This gets us at the sense in which human consciousness centers on the affections and the ways in which it is subjective. Two further points must be discussed: the sense in which affective ideas represent external bodies and the relation between affections and appetites.

Affective ideas are central to human experience and human subjectivity. But they are also representations of external bodies. This emerges from the just-quoted definition of images as "the affections of the human body whose ideas present external bodies as present to us" (E2p17s). For Spinoza, *all* affections are images, because the idea of any affection of the body "involves" the nature of the external body that affects it (E2p16) and thereby represents it. Recall from section 5 that for x to involve y is for it to be the case that x cannot be conceived without y . So in conceiving my affection a , I am also conceiving the external body that caused a – call it c . That is, in forming my idea of a , or $I(a)$, I also form $I(c)$ representing c . (I use ' $I(x)$ ' to represent the idea that has x for its object, whose objective reality is what $I(x)$ represents.)

A major question arises here. $I(a)$ is an idea that represents a bodily affection – a certain brain process, most likely. But even though my mind "perceives" a (by E2p12), it seems wrong to say that I am conscious of it. I'm not aware of my brain at all in ordinary situations. Instead, I am conscious of c , *via* a . What explains this disparity in what I am conscious of?

While this initially seems like a major problem for Spinoza to explain, I actually think it rests on a fundamental misunderstanding. The response is that I *am* conscious of $I(a)$, but not conscious of it *as* a brain process. I misidentify its object as not a but c . This gives Spinoza a ready handle to explain what perceptual subjectivity consists in for him: it consists in such misrepresentations. More precisely, it consists in my being conscious of $I(a)$ insofar as it represents c , but not insofar as it represents the nature of my body. At the same time, my idea of c will be deeply influenced by the nature of my body; which properties of c I am aware of will depend on how my body is capable of being influenced by c (see E2p16c2). And often my idea of c will include elements that are not properties of c at all. (These points will be crucial to elaborating Spinoza's account of phenomenal quality in the next section.)

Finally, just a few more words about consciousness of appetites. The first time a version of the term *conscius* occurs in a proposition in the *Ethics*, it is in E3p9, in which Spinoza states that we

are conscious of our *conatus*, or our striving to persevere in our being. (Earlier, in E1app, Spinoza had already claimed that humans are “*conscii* of their volitions and their appetite,” in particular their appetite for “their own advantage”; G II, 78.) In the scholium to E3p9, Spinoza says that *conatus* can take different forms depending on how it is conceived:

When this striving is related only to the mind, it is called will; but when it is related to the mind and body together, it is called appetite. [...] Between appetite and desire there is no difference, except that desire is generally related to men insofar as they are conscious of their appetites. So *desire* can be defined as *appetite together with consciousness of the appetite*. (E3p9s; his italics)

Before, I argued that Spinoza holds that we are fundamentally conscious of our affections. In E3p9s, he says that we are conscious of our appetites. Far from contradicting each other, these statements actually support each other. They can be reconciled as follows. In line with his general view that a thing's power is its essence (e.g., E1p34, E1p36), Spinoza states that a thing's essence just is its *conatus* (E3p7). Any mode is causally active just in virtue of its essence itself. However, human beings are essentially finite modes and as such are necessarily subject to affection by other modes (E4p4). This means that while humans are always “active” in the sense that their actions are partially determined by their *conatus* – hence by their essence itself – they are also always “passive” because their actions are also partly determined by external causes (see E3def2).

In E3p9dem, after recalling E3p7's claim that a thing's *conatus* is its essence, Spinoza refers back to E2p23, the claim that we are only conscious of ourselves through our affections. Because affections are determinations of our *conatus*, hence of our essence, in being conscious of our affections we are conscious of our essence, hence of ourselves. In other words, consciousness of affections and consciousness of our appetites comes to the same thing for Spinoza, because affections are determinations of the *conatus*.

The picture that emerges is that Spinoza's accounts of subjectivity and consciousness come together in his theory of affections and appetites. Fundamentally, humans know themselves through the affections that their bodies undergo. Only with some effort can they step back from the “order of the affections” to reconstruct, to the extent that they are able, the “order

of the intellect” that can provide them with adequate knowledge.¹⁵⁶ The order of the intellect is the order in which God understands the world. Although they are not experienced this way, the two orders are necessarily parallel, like the mind and the body. (This is because the orders are not orders of numerically distinct ideas, but of two ways of conceiving of one and the same idea.)

8. *Spinoza as a representationalist: Content and confusion*

The preceding section has finally put us in a position to determine Spinoza's view of phenomenal consciousness. I will argue that Spinoza is a representationalist about phenomenal qualities.¹⁵⁷ Representationalism has a claim to being the default interpretation of Spinoza's theory of consciousness, because it is plausible that Spinoza has a *fully* representationalist theory of mind – in other words, that he takes *every* feature of the mind to be ultimately derived from its representational features.¹⁵⁸ Still, an explanation of how this accounts for Spinoza's views about consciousness needs to be given. In particular, representationalism needs to be reconciled with the difference in perspectives that is opened up by Spinoza's claim that any idea can be considered both insofar as it is in God's mind and insofar as it is in a finite mind.

First it is important to get clear on what it means to be a representationalist about phenomenal consciousness. In its broadest form, representationalism is the view that the phenomenal quality of an experience is determined by what the experience represents: by its representational content. The determination involved can be understood in various ways. *Reductive* representationalism holds that phenomenal quality is nothing over and above representational content. This can in turn be understood either in *epistemic* or *ontological* terms: it is either the case that phenomenal quality can be *explained* in terms of representational content, or that it simply *is* representational content. Epistemically reductive representationalism naturally implies

¹⁵⁶ To explain how this occurs, it would be necessary to discuss Spinoza's epistemology, particularly his theory of the common notions, which are adequate ideas formed on the basis of ideas of affections (E2p37-39). This should then be extended into a discussion of how the mind can become conscious of “itself, and of God, and of things” (E5p39s). Because I am especially interested in sensible qualities, I restrict myself here to ideas of affections. However, in brief, my suggestion would be that in forming more adequate ideas, the mind takes up a firmer position in the order of the intellect at the expense of the order of the affections, becoming more conscious of its adequate ideas of itself, God and of things and less of its inadequate ideas. See LeBuffe (2010b) for an account of this transition that I largely agree with.

¹⁵⁷ Perler (2018, 250-4) argues for a closely related view, that Spinoza is a representationalist about the emotions.

¹⁵⁸ See Della Rocca (2008b, 118-22).

ontologically reductive representationalism, but the reverse is not true. *Non-reductive* representationalism, by contrast, holds that the determination involved is weaker, such that the phenomenal quality is not identical to representational content but does depend on it (ontologically non-reductive), or that it is not entirely explainable in terms of representational content (epistemically non-reductive). Again, it is possible to combine epistemic non-reductionism with ontological reductionism.¹⁵⁹

In the terms of this schema, it is fairly clear that Spinoza is an ontologically reductionist representationalist. It is much less clear if he would opt for an epistemic reduction of phenomenal character as well. To determine his position, it is necessary to consider his theory of affection and particularly the notion of confused ideas. In this section, I will present an interpretation of this theory that supports my contention that Spinoza's representationalism is ontologically reductive; in the next section I will argue that it is epistemically non-reductive.

How confusion is understood is going to have decisive repercussions for the characterization of Spinoza's theory of consciousness as a whole. Spinoza himself is notoriously unclear on what confusion consists in. It is clearly related to the falsity and inadequacy of ideas – such that an idea is confused iff it is false iff it is inadequate (see Della Rocca 1996, 58, 109) – but Spinoza never provides a definition of confusion. My goal in this section is to set out an understanding of this notion that supports the identification of the confusion of ideas with their phenomenal quality. The importance of the understanding of representational content to any construal of representationalism justifies a somewhat detailed discussion. I will discuss representation and confusion in turn.

In the context of Spinoza's theory of ideas, it is often said that various features of ideas are 'mind-relative.'¹⁶⁰ It is said both that the *content* and the *confusion* of ideas is mind-relative: that is, it is said that one and the same idea can have one content, that is not confused, in God's mind and a different content, that *is* confused, in a human mind. Returning to I(a), the idea of my affection, this idea is supposed to represent *a* non-confusedly in God's mind, while it represents *c* confusedly in my mind. The main thing I want to emphasize here is that any such

¹⁵⁹ This characterization of the various forms of representationalism is my own, but based on widespread uses of the notions of epistemic and ontological reduction. It is inspired by Chalmers' (2004, 349-53) classification, which uses the same categories but defines them differently. See also Lycan (2019, §2.4).

¹⁶⁰ This terminology originates with Della Rocca (1996, ch. 3) but is also used by other authors, notably Garrett (2008).

statement has to be understood very carefully and that even then, the terminology raises more problems than it answers.

According to the view that content is mind-relative, the content of one and the same idea can change depending on which mind it is related to. This idea conflicts with how the recent literature understands Spinoza's notion of representational content. It is now common to identify an idea's representational content with the *objective reality* of its object.¹⁶¹ So the content of $I(a)$ – that which it represents – is the objective reality of a . For Spinoza (as for Descartes, from whom he inherits this terminology), objective reality is to be contrasted with *formal* reality, or reality independent of being represented. For Spinoza, any finite thing possesses both kinds of reality, in virtue of both being produced by God's power of acting (E1p16) and being represented in his intellect as a result of his infinite power of thinking (E2p7c).¹⁶²

Now, this understanding of representational content sits uneasily with the mind-relativity of content, because it is plausible that Spinoza believes that ideas can be individuated in terms of their content. That is, in line with Spinoza's understanding of essence as relating to what is necessary and sufficient for both "positing" and "conceiving" a thing (E2def2), the objective reality of x is necessary and sufficient for $I(x)$. It is necessary, because $I(x)$ cannot be conceived without conceiving x ; and sufficient, because in representing x , we thereby form $I(x)$.¹⁶³

To be sure, there is a complication here due to the fact that one and the same x can have objective reality in several distinct ideas. But in any idea other than $I(x)$, while it will represent x , it will presumably also represent something other than x , or it will represent x only partially. We can say that ideas are individuated by their *complete* content.

If ideas are individuated by their complete contents, then complete contents cannot be mind-relative. Because if the complete content of $I(x)$ is different in God's mind than it is in mine, then God and I are in fact forming different ideas. Moreover, consider the essence of $I(x)$ itself. If the content of $I(x)$ is mind-relative, then it seems that the essence of $I(x)$ is itself relative in some way – that its essence is not intrinsic to it. This is a strange result, because a thing's

¹⁶¹ See Garrett (2018c) and Hübner (2019, 2022).

¹⁶² For discussion, see Chapter 2, section 5 and references therein.

¹⁶³ For a fuller version of this argument, see Chapter 2, section 5; cf. Hübner (2022, 54-5).

essence is intrinsic to it by definition.¹⁶⁴ It is much more natural to say that an idea's content is not mind-relative.

Now, I am not sure that the defenders of mind-relativity really intend to hold that ideas' complete contents are mind-relative. They are perhaps better understood as saying that $I(x)$ is present with its *complete* content in God's mind, whereas it only has *part* of its content in mine. I believe this is a much more promising way of understanding the distinction between God's perspective on $I(x)$ and my own (though I still have some reservations). However, it also underscores how misleading the terminology of mind-relativity is. For my limited understanding of $I(x)$ does not alter its content – its complete content remains intrinsic to it, despite my limited grasp of this content. If this is the view, it could be more accurately rephrased in terms of a distinction between an idea's *intrinsic* and its *extrinsic* content, such that its intrinsic content is both what individuates it and what is present in God's mind, whereas its extrinsic content is present in my mind. Then, extrinsic content may be said to be a mind-relative feature of ideas, but intrinsic content is not.

I still have misgivings about this restatement of the view. This is because of the notion of 'extrinsic content'. It is important to realize that extrinsic content is not a *kind* of content. An idea does not have its intrinsic content and then, in addition, its extrinsic content. Instead, it has only its intrinsic content. Its 'extrinsic content' is due to the way (part of) its intrinsic content is related to something that the idea does not immediately represent: for example, $I(a)$ is related to c , which is external to a and which $I(a)$ only represents to the extent that it 'involves' c . Describing this as 'extrinsic content' however has the unfortunate consequence of seemingly reifying it.

My preferred alternative is to specify the distinction between how God and I grasp $I(x)$ not in terms of the content $I(x)$ has in our minds, or even in terms of its 'intrinsic' and its 'extrinsic' content, but in terms of how $I(x)$ is *understood* by God and me. Here I fall back on the distinction between two epistemic perspectives that is introduced in E2p17s, as discussed in the previous section. On this view, $I(a)$ has the *same* content in God's mind as it does in mine. However, I

¹⁶⁴ Perhaps it could be said that the totality of an idea's mind-relative manifestations constitutes its essence. But this move has the unfortunate consequence of making $I(x)$'s essence dependent on its representation by other finite minds. Say $I(x)$ is the idea of a geometric entity such as a triangle. On the current view, the essence of $I(x)$ depends on every thought of this triangle had by anyone ever. This is hard to square with Spinoza's general claim that essences are 'prior' to existents, and that all essences are caused by God (E1p25). It also appears to conflict with his parallelism, in that the existence of $I(x)$ exhibits dependence on other modes of its attribute, which x itself presumably does not.

understand $I(a)$ in the 'subjective' manner of E2p17s. As a result of this, I relate $I(a)$ primarily to c and take it to represent c instead of a . This is an illusion about $I(a)$'s representational content that results from the distinctive, affection-based way in which I apprehend $I(a)$. (However, it is not a complete error, as $I(a)$ *does* represent c , albeit indirectly – to the extent that c is involved in a . That is, c has objective reality in $I(a)$.) God is not subject to this illusion.

An analogy might make this clearer. Suppose I am using a microscope to look at a tiny worm in the blood, invisible to the naked eye. The microscope projects an image of the worm onto its ocular lens, which I see. I accordingly form an affective idea of this lens image, $I(l)$. However, I am not interested in and, in fact, not perceptually aware of the way the microscope constitutes this image. Although I may know in the abstract that the microscope can distort my perception of the worm in certain ways, I will not normally be immediately aware of this while looking at the lens image (especially if I am untrained in the use of microscopes). I cannot *see* these distortions in the way that I can see l . As a result, instead of taking $I(l)$ for what it is – a representation of a lens image – I will take it for a representation of the worm, w . This is not a complete error, as $I(l)$ *does* represent w , if indirectly – to the extent that w is involved in l . However, my situation here, with its necessary epistemic limitation due to the limits of my natural sight, would not occur for an omniscient entity like God, who could form $I(w)$ directly and who would immediately understand exactly how both w and the microscope together constitute l , and hence could deduce $I(l)$ from the ideas of w and the microscope. God would accordingly grasp the complete content of $I(l)$, whereas I can only grasp part of it.

This way of understanding content also helps to clarify in what ways $I(a)$ is and is not confused. I follow LeBuffe's "simpler" account of confusion here, which says that for Spinoza, "an idea is confused just because it represents its object or objects in a fragmentary and incomplete way."¹⁶⁵ "Object or objects" must be understood broadly, to refer to anything that an idea can represent. In other words, c is an object of $I(a)$ in this sense.

Spinoza suggests an account of confusion like this in E2p28dem, when he writes that "these ideas of the affections, insofar as they are related only to the human mind, are like conclusions

¹⁶⁵ LeBuffe (2010a, 58). He also describes a more complex account (advocated by Della Rocca 1996, 59–61, among others), according to which confused ideas are the result of the body's capacity for forming distinct images being exceeded. Spinoza discusses this process in E2p40s1. However, this scholium is primarily an explanation of how the mind forms ideas of the "transcendental terms" *being*, *thing* and *something* and universal ideas like *horse* and *human being*. It is not clear to me that every confused idea must be either transcendental or universal, so I'm not sure that Spinoza's discussion here is meant as an explanation of all confused ideas. At any rate, the 'simpler' account suffices for my purposes here.

without premises, i.e. (as is known through itself), they are confused ideas". $I(a)$ is like a conclusion without premises because the mind arrives at it without being able to deduce it from its causes. The 'premises' from which $I(a)$ could be deduced – $I(b)$ and $I(c)$ – are, after all, not in the human mind; in fact, the mind forms the latter two ideas on the basis of its affective ideas, instead of the other way around. Spinoza also refers to ideas understood in the order of affections as "confused and mutilated" (e.g., E2p29c), again emphasizing the faulty order in which these ideas are formed by the human mind.

It is important to note that, understood in this way, an idea's confusion is intrinsic to it. A different way to formulate the criterion of confusion is that confused ideas are those that do not entail adequate representations of their causes. For Spinoza, $I(a)$ 'involves' its causes, but only to a limited extent (E2p16). Recall that if x involves y , then x cannot be conceived without y ; this in turn means that the idea of x implies the idea of y (but not necessarily an adequate idea of y).¹⁶⁶ If that is right, then affective ideas are confused by definition for Spinoza, because they all represent causes that are external to the affected body (cf. the definition of imagination in E2p17s, cited above).

There are two things to note here. First, this distinguishes affective ideas from the adequate ideas that the mind is able to form, which do imply adequate representations of their causes. Primarily, these are ideas of God and of the attributes, which, as conceived through themselves, trivially entail conceptions of their causes – i.e., of themselves. Second, the affective idea is confused even in God's mind. Even in God's mind, $I(a)$ does not imply adequate representations of its causes. This is irrelevant for God, because God already possesses the ideas of a 's causes and so also has adequate knowledge of $I(a)$. This adequate knowledge, however, "is in God, not insofar as he is considered to be affected with the human mind, but insofar as he is considered to be affected with other ideas" (E2p28dem). It is in him, that is, insofar as he has the ideas from which $I(a)$ can be deduced, but not insofar as he has the idea of the human mind that contains $I(a)$. The reason for this, I want to suggest, is that $I(a)$ just does not entail adequate representations of these causes, no matter how it is considered. Again, the analogy with the ocular lens image may help here. The claim made here is that no consideration of the content of $I(l)$ by itself, no matter how detailed, will enable the construction of adequate ideas of the

¹⁶⁶ As Gabbey writes: "Spinoza uses *involvere* to mean 'to contain necessarily', that is 'to imply' or 'to entail', though implication seems closer than entailment to the notion of necessary containment. Moreover, *involvere* was used interchangeably with *implicare* by scholastic writers." (Gabbey 2008, 47n10; cited from Melamed 2012, 88.)

worm and the microscope. It will only lead to an idea of w to the extent that w 's nature is involved in $I(l)$. In this respect, then, $I(l)$ and $I(a)$ are intrinsically confused – their confusion is a property of their representational content.

Let's now return to representationalism. The upshot of the analysis just given is that it allows us to ascribe a robustly ontologically reductive form of representationalism to Spinoza. We can identify an idea's phenomenal character with its confusion. This confusion in turn is an intrinsic property of the idea's representational content.

9. Can the phenomenal quality of an idea be explained?

Although Spinoza is an ontologically reductive representationalist about phenomenal quality, I believe his representationalism is epistemically non-reductive. Spinoza's theory of phenomenal quality here repeats a feature that we encountered above in relation to the conceivability argument: Spinoza allows for two ways of conceiving one and the same thing, that are conceptually independent (neither entails the other) but are nevertheless necessarily connected. This is true of thought and extension for him. It is also true for the confused representational content that determines phenomenal quality, which can be conceived either according to the order of the affections or that of the intellect. In this case, however, there is the additional twist that only one of these ways of conceiving can be said to be adequate. The way of conceiving $I(a)$ that is accompanied by phenomenal quality is the confused way of conceiving it; but in forming an adequate conception of it, this phenomenal quality evaporates.

In the case of ideas of sensible qualities, a confused understanding of them grasps them according to the order of the affections. In doing so, it relates the content to its (unknown) causes resulting in inadequate representations of these causes; this is what experience of sensible qualities exists in. An adequate understanding of the same idea understands it according to the order of the intellect. This understanding will still be able to see that $I(a)$ is intrinsically confused, but it will not itself experience this confusion as it instead understands $I(a)$ through an understanding of its actual causes. In doing so, it will not experience the phenomenal quality that normally accompanies the sensible quality idea. This adequate understanding will represent the physical properties of the external body c and how these properties interact with my body b . It will represent physical qualities without being accompanied by a feeling of the color, warmth, of whatever other sensation this affection of my body produces.

Therefore, what it is like to experience these qualities will not be explainable in terms of the representational content of $I(a)$. Because an adequate understanding of this content does not entail anything about the phenomenal quality. But insofar as it is confused, it does not meet the minimum epistemic standard of being an explanation – which for Spinoza is that of belonging to the second or third kind of knowledge (see E2p40s2). Because a scientific explanation of experience will not contain confused ideas, it will not be able to explain what it is like to experience sensible qualities. It will be able to explain the sensible qualities, but – much like Bacon and Descartes, or even Boyle – it will explain them in physical terms without addressing their phenomenal quality.

10. Conclusion: Spinoza's pregnant silence on the sensible qualities

After the foregoing exposition of Spinoza's theory of phenomenal quality and his broader stance on conceivability arguments for dualism, it is now possible to return to the question with which this chapter opened: what explains Spinoza's relative silence about the sensible qualities?

Sometimes, a philosopher's silence about a topic is a sign that they have nothing to say. However, it is clear that this is not the case for Spinoza's reticence in relation to the sensible qualities. The reconstruction of his views has shown that he has a sophisticated and well-developed theory for thinking about consciousness and subjectivity, in terms of which he could have easily fitted the sensible qualities. So why didn't he? Why does he never mention sensible qualities again in the *Ethics* after E1app? If he did have something to say but did not say it, the natural way to explain this is that it was a strategic decision: Spinoza did not think the sensible qualities posed as big of a problem, or were as central to understanding human consciousness and subjectivity, as certain other phenomena he could address.

The arguments and reconstructions of this chapter have suggested two issues related to the sensible qualities that Spinoza thought were more important than the sensible qualities themselves. The first of these is Descartes's separability argument for dualism. In this argument, Spinoza diagnosed an illegitimate move from independent conceivability to the possibility of separate existence. Spinoza critically pointed out the explanatory gaps that resulted from this move. And in his own metaphysics, he developed an alternative way of accommodating explanatory gaps by way of construing independently conceived entities in parallel to each other.

He applied this structure not just at the very basis of his own metaphysics, but also as a way to accommodate phenomenal quality as something conceivable both adequately and inadequately, with its phenomenal quality accessible only in the latter case.

The other issue is subjectivity as such, and particularly its relation to the ethical question as to how it is possible to alter and improve one's subjective experience on the path towards greater virtue. Because this question relates especially to the experience of the affects, Spinoza emphasizes 'affective quality' when he mentions peculiarly subjective phenomena in E1app. He leaves unaddressed the fact that sensible qualities have a phenomenal character that goes beyond their affective character and does not produce an explanation for this phenomenal character. But this is because there *is* no explanation that does justice to the phenomenal character of the sensible qualities, to why things that are red are experienced with *this* specific phenomenal quality instead of another. While there is a sufficient reason for this correlation, there is no intelligible connection between the experienced quality and what it represents.

Conclusion

Over the course of this dissertation, many different kinds of parallelism have been introduced. So by way of conclusion, let me first briefly review them and explain how they connect to causal-explanatory mind-body parallelism as it was introduced in Chapter 1, and then end with some remarks about how these different parallelisms cohere into a general strategy for approaching mind-body problems.

This overview of parallelisms reflects only the ones that were encountered in the dissertation and is of course not exhaustive. Many other parallelisms could be constructed. Additionally, there are other historical versions of parallelism. Although I have limited myself to Spinozist forms of parallelism in this dissertation, discussing Leibniz's versions of the view would have brought up interesting other varieties of parallelism, such as a parallelism between efficient and final causes and a different way to develop a metaphysical parallelism based on the idea of expression. I hope to explore these in the future.

As for the parallelisms that have been encountered over the previous chapters, they differ in the way the parallel items are related. Spinoza's representational parallelism holds between an adequate idea and its object. His metaphysical parallelism holds between modes of corresponding attributes, or alternatively between modes that 'express' the same thing in different ways. While his mind-body parallelism tends to be seen as a metaphysical parallelism only, I argued in Chapter 2 that it incorporates representational parallelism as well.

Chapters 4 and 5 introduced a parallelism between the "order of affections" and the "order of the intellect," to pick up Spinoza's terms in E2p18s. The order of the intellect is the order of adequate ideas that are representationally parallel to their objects. Now, insofar as the affections are adequately conceived, their ideas are also representationally parallel to the bodily affections. But of course, this is not how they are normally conceived. Humans (and other finite entities)

form ideas of the affections according to the order of the affections, which is the order in which they were originally affected by the bodies that these ideas confusedly represent, supplemented by the various mental operations such as association and memory that determine our ordinary ways of thinking about the world. The order of affections diverges from the order of the intellect and in this way represents Spinoza's attempt to account for error, for subjectivity and for passivity, three things that, while not inherently desirable, nevertheless are constitutive of a distinctively finite perspective on reality.

Because the orders diverge, these distinctively finite features may appear to escape from nature's lawful order, as dualists have often taken them to do. However, as I showed, Spinoza is able to preserve a parallelism between the two orders: a parallelism between their respective contributions to any action. This parallelism is unlike any others discussed in the dissertation, in three ways. First, the correlation involved is not direct but inverse. That is to say, intellectual ideas are causally productive in any finite action to the extent that affective ideas are causally impotent, and vice versa. Spinoza exploits this feature in his theory of freedom, saying that we are more free to the extent that we are more internally determined to act. Second, unlike any of Spinoza's other parallelisms, the things related here are not "one and the same" but distinct, being a thing's essence and its affections. (However, they do both contribute to one and the same volition.) Third, there is no problem in this case of making sense of causation between the parallel items. They do not belong to different attributes. In fact, their causal interaction is presupposed – it is part of what we want to understand in understanding the effect of the affections on our actions. Finally, although this does not set it apart from causal and metaphysical parallelism strictly speaking, the affect-essence parallelism is not an instance of mind-body parallelism. This is because it involves a conflict within a single attribute (although this conflict plays out in parallel ways both in the body and the mind).

In Chapter 4, I argued that this model is parallelist insofar as it obeys the two requirements of the states in the two causal orders being isomorphic to each other and of being mutually irreducible. These are requirements commonly made of parallelism and it seems fair to say that a view that obeys them is parallelist. But an attentive reader will have noticed that causal-explanatory parallelism makes an additional demand of the parallel items: they also stand in causal chains of *sufficient* and not merely partial causes. The affection-essence parallelism fails to meet this requirement, because in it the affection and the essence are a sufficient cause only when

put together, not apart from each other. So this is not an instance of causal-explanatory parallelism.

As a result, we have two kinds of parallelism. The ‘strong’ or ‘strict’ form of parallelism is a causal-explanatory parallelism. The ‘weak’ or ‘loose’ form of parallelism, like affection-essence parallelism, meets the requirements of isomorphism and irreducibility, but does not require causal or explanatory sufficiency of the parallel chains. What explains the distinction between these two? One seemingly plausible suggestion is that only strict parallelism is a parallelism across an explanatory gap. Because the explanatory gap supports an epistemic form of non-reductionism, a cross-gap parallelism that did not require causal and explanatory sufficiency would find itself faced with the task of coordinating the independently conceivable contributions of the items on each side. This is exactly the situation that parallelism is designed to avoid, however. The same problem does not occur for a loose parallelism. In the case of the affect-essence parallelism, there is no explanatory gap between the causal contribution of the affect and the essence – both belong to the same attribute, after all.

However, there appears to be a complication here. In the terms of the distinction just introduced, the parallelism between representational content and phenomenal quality discussed in Chapter 5 is a ‘loose’ parallelism, insofar as there is no complete explanation of any phenomenal quality in terms of other phenomenal qualities. (This is because there is no complete explanation of any phenomenal quality *qua* phenomenal quality at all.) The parallelism here is also a same-attribute parallelism. Nevertheless it is a cross-gap parallelism: there is an explanatory gap between the representational content of an idea and its phenomenal quality. Hence being cross-gap cannot be the only thing that distinguishes loose and strict parallelism from each other. Having noticed this, however, it is easy to see what does allow for a clean distinction: it is the additional requirement, besides being cross-gap, of the parallel items having complete explanations. If Spinoza had thought that phenomenal qualities could be entirely explained in qualitative terms, he would have developed a strict parallelism to explain them. But he didn’t think that, so he worked out a loose parallelism instead.

These distinctions between strict and loose and cross-gap and no-gap parallelisms are not just helpful for classifying the parallelisms encountered in this dissertation. They also provide some guiding questions to ask when developing parallelist views of other contested phenomena. The questions to be asked are the following:

1. *Is there an explanatory gap between the phenomenon to be accounted for and what it is parallel to?* In many cases there will be. However, in the case of agency, there does not seem to be an explanatory gap between a free action and its corresponding physical processes; or at least this possibility was not pursued in Chapter 4.¹⁶⁷
2. *Does the phenomenon to be accounted for have a complete explanation?* Although for Spinoza there is a sufficient reason why any experience has the phenomenal quality it has, this phenomenal quality cannot be explained insofar as it is qualitative, but only insofar as it is a representational content.
3. *What is the shared “order and connection” between the parallel items?* In the case of agency, the order and connection is due to actions being the result of parallel contributions of the agent’s *conatus* and that of the affecting body. In the case of representational parallelism, it is due to Spinoza’s commitment to *Causation* \leftrightarrow *Conception*. In the case of metaphysical parallelism, it is due to his claim that mind and body are expressions of one and the same thing.

If the answer to (1) is no, or the answer to (1) is yes but that to (2) is no, the resulting parallelism will be loose; if the answer to both (1) and (2) is yes it will be strict.

More than anything else, the success of parallelist theories of contested phenomena will depend on finding convincing answers to (3). It is by no means obvious that a shared order and connection will always be available. And it is possible that cases will turn up in which there demonstrably is no shared order and connection. However, the way to find these things out is to pursue further parallelist theories. It is by no means obvious either that agency or phenomenal quality would exhibit parallelism, yet if the arguments of this dissertation have been correct, they do (at least on Spinoza’s understanding of them).

¹⁶⁷ If there were such a gap, it would presumably manifest in the independent conceivability of a particular physical process corresponding to an action without its corresponding action, and vice versa. A theme in Anscombe’s writings on intentional action is that such an action is explained by answering a certain type of ‘why’ question, and that this question cannot be answered just by citing physical causes (at least not under a physical description). There might accordingly be something like an explanatory gap here. See Anscombe (1963, 1979).

Acknowledgements

I'm very grateful to my supervisor, Mike Griffin, for his support and criticism, both of which were (and are) much appreciated. Mike gave me the freedom I needed to explore this in some ways strange project, but he was also always ready to give feedback on my writing and ideas. His insistence on clarity of expression has done a lot to improve the dissertation (not enough, he might think).

During my PhD research I was lucky to be able to visit at Humboldt University in Berlin and (virtually) at the University of Groningen. I'm thankful to my hosts, Dominik Perler and Andrea Sangiacomo, for supporting me. During my stay in Berlin, Dominik read three draft chapters of the dissertation; his substantial feedback improved them a lot. I also presented a fourth chapter in his research colloquium, where it too received helpful feedback. Andrea was instrumental in helping me write the first chapter, especially when it came to giving it its final structure. I also presented this chapter at the history of philosophy colloquium of the Groningen department. Thanks are due to the audience, and particularly to Laura Georgescu, for helpful questions and suggestions.

Several people have read and commented on things that I've written for or related to this dissertation. For this I would like to thank Dávid Bartha, Andrea Christofidou, Kati Farkas, Philip Goff, Zhiwei Gu, Ferenc Huoranzski, Marta Santuccio and Huaming Xu. For conversations about my ideas, I would like to thank Daniel Garber, Julia Jorati, Martin Lin, Tim Crane, Ursula Renz, Howard Robinson and Jorge Secada.

For conversations and friendship, thank you (in no particular order) to Maarten, Rob, Kerem, Valentina, Zhiwei, Joshua, Dilara, Huaming, Maşuk, Nevim, Garrett, Damian, Jamie, Marta, László and Csilla and Péter. Thank you to Dávid and Szilvi for helping to make my stay in locked-down Berlin better than it would have otherwise been.

ACKNOWLEDGEMENTS

Special thanks are due to Forrest Schreick and Sam Taylor. Forrest has read and commented on several of the chapters, but apart from that he has been a continually encouraging presence as we both make our way through the thickets of neo-rationalist metaphysics of mind and its early modern inspirations. I have known Sam since we started our MA together at Utrecht, now a decade ago. Since then we have exchanged and discussed pretty much every text we have written (including most of the chapters of this dissertation). Our conversations have had a large impact on the philosopher I have ended up (being in the process of) becoming. I hope that they will still continue for a long time.

Finally, I would like to thank my parents for always supporting what I do.

Abbreviations and Conventions

EDITIONS:

- AT Descartes, René. 1964. *Oeuvres de Descartes*. Edited by Charles Ernest Adam and Paul Tannery. 11 vols. Paris: J. Vrin. Cited by volume and page number.
- CSM Descartes, René. 1984. *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. 2 vols. Cambridge: Cambridge University Press. Cited by volume and page number.
- G Spinoza, Benedictus de. 1925. *Opera*. Edited by Carl Gebhardt. 4 vols. Heidelberg: Carl Winters. Cited by volume and page number.

SPINOZA'S WORKS:

- CM *Metaphysical Thoughts (Cogitata Metaphysica)*. Cited by part and chapter number.
- CPP *Descartes' Principles of Philosophy (Renati Des Cartes Principiorum Philosophiae)*.
- E *Ethics*
- Ep *Letters*
- KV *Short Treatise on God, Man, and His Well-Being (Korte Verhandeling van God, de Mensch en deszelfs Welstand)*. Cited by part and chapter number.
- NS *Nagelate Schriften* (Dutch translation of OP; both 1677)
- OP *Opera Posthuma*
- TIE *Treatise on the Emendation of the Intellect (Tractatus de Intellectus Emendatione)*. Cited by paragraph number.
- TP *Political Treatise (Tractatus Politicus)*. Cited by chapter and paragraph number.

References to the *Ethics* and the CPP have the following structure: [Title][Part number][Element][Optional element number][Optional sub-element][Optional sub-element number] (etc.).

I use the following abbreviations for elements and sub-elements:

a	axiom
app	appendix
c	corollary
def	definition
DefAff	definition of the affects
dem	demonstration
e	enunciation
p	proposition
post	postulate
pref	preface
s	scholium

So E1p8s2 is the second scholium to proposition 8 of Part One of the *Ethics*.

I adapt all references to Spinoza's works to follow these conventions, including in cited material.

I use Curley's edition for translations (Spinoza 1985-2016). I sometimes silently remove capitalizations in my citations. Other divergences from Curley's text are noted when they occur. When citing from longer stretches of text, I include references to the corresponding page numbers in G (these are incorporated into the margins of Curley's edition).

Bibliography

- Achinstein, Peter. 1975. "Causation, Transparency, and Emphasis." *Canadian Journal of Philosophy* 5 (1): 1–23.
- Alter, Torin, and Yujin Nagasawa, eds. 2015. *Consciousness in the Physical World: Perspectives on Russellian Monism*. Oxford University Press.
- Anscombe, G. E. M. 1963. *Intention*. 2nd ed. Oxford: Blackwell.
- . 1979. "Under a Description." *Noûs* 13 (2): 219–33.
- . 2005. "The Causation of Action." In *Human Life, Action and Ethics*, edited by Mary Geach and Luke Gormally, 82–101. Exeter: Imprint Academic.
- Aquila, Richard E. 1978. "The Identity of Thought and Object in Spinoza." *Journal of the History of Philosophy* 16 (3): 271–88.
- Armstrong, Aurelia. 2009. "Autonomy and the Relational Individual: Spinoza and Feminism." In *Feminist Interpretations of Benedict Spinoza*, edited by Moira Gatens, 43–63. University Park: Pennsylvania State University Press.
- Armstrong, Aurelia, Keith Green, and Andrea Sangiacomo, eds. 2019. *Spinoza and Relational Autonomy: Being with Others*. Edinburgh: Edinburgh University Press.
- Árnadóttir, Steinvör Thöll, and Tim Crane. 2013. "There Is No Exclusion Problem." In Gibb, Lowe, and Ingthorsson, 2013, 248–66.
- Ayers, Michael. 1998. "Ideas and Objective Being." In *The Cambridge History of Seventeenth-Century Philosophy*, edited by Daniel Garber and Michael Ayers, 2:1062–1107. Cambridge: Cambridge University Press.
- Bennett, Jonathan. 1984. *A Study of Spinoza's Ethics*. Indianapolis: Hackett.

- . 1993. “Eight Questions About Spinoza.” In *Spinoza on Knowledge and the Human Mind*, edited by Yirmiyahu Yovel, 11–25. Leiden: Brill.
- Bennett, Karen. 2003. “Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It.” *Noûs* 37 (3): 471–497.
- . 2007. “Mental Causation.” *Philosophy Compass* 2 (2): 316–337.
- . 2008. “Exclusion Again.” In *Being Reduced: New Essays on Reduction, Explanation, and Causation*, edited by Jakob Hohwy and Jesper Kallestrup, 280–305. Oxford: Oxford University Press.
- Block, Ned. 1995. “On a Confusion about a Function of Consciousness.” *Brain and Behavioral Sciences* 18 (2): 227–47.
- Broad, C. D. 1925. *The Mind and Its Place in Nature*. London: Kegan Paul.
- Brüntrup, Godehard, and Ludwig Jaskolla, eds. 2016. *Panpsychism: Contemporary Perspectives*. New York: Oxford University Press.
- Carriero, John. 2011. “Conatus and Perfection in Spinoza.” *Midwest Studies in Philosophy* 35: 69–92.
- . 2019. “Spinoza on the Primary Affects.” In Naaman-Zauderer, 2019, 82–107.
- Chalmers, David J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- . 2003. “Consciousness and Its Place in Nature.” In *Blackwell Guide to the Philosophy of Mind*, edited by Stephen P. Stich and Ted A. Warfield, 102–42. Malden, MA: Blackwell. Reprinted in Chalmers, 2010, 103–39.
- . 2004. “The Representational Character of Experience.” In *The Future for Philosophy*, edited by Brian Leiter, 153–81. Oxford: Clarendon Press. Reprinted in Chalmers, 2010, 339–71.
- . 2010. *The Character of Consciousness*. New York: Oxford University Press.
- . 2015. “Panpsychism and Panprotopsychism.” In Alter and Nagasawa, 2015, 246–76.
- Clifford, William Kingdon. 1879. “Body and Mind.” In *Lectures and Essays*, 2:31–70. London: Macmillan and Co.
- Coleman, Sam. 2018. “The Evolution of Nagel’s Panpsychism.” *Klasis* 41: 180–202.
- Crane, Tim. 2019. “A Short History of Philosophical Theories of Consciousness in the 20th Century.” In *Philosophy of Mind in the Twentieth and Twenty-First Centuries*, edited by Amy Kind, 78–103. Oxon: Routledge.

- Crane, Tim, and Sarah Patterson. 2000. "Introduction." In *History of the Mind-Body Problem*, edited by Tim Crane and Sarah Patterson, 1–12. London: Routledge.
- Curley, E. M. 1969. *Spinoza's Metaphysics: An Essay in Interpretation*. Cambridge, MA: Harvard University Press.
- . 1988. *Behind the Geometrical Method: A Reading of Spinoza's Ethics*. Princeton: Princeton University Press.
- Curley, Edwin, and Gregory Walski. 1999. "Spinoza's Necessitarianism Reconsidered." In *New Essays on the Rationalists*, edited by Rocco J. Gennaro and Charles Huenemann, 241–62. New York: Oxford University Press.
- Dasgupta, Shamik. 2016. "Metaphysical Rationalism." *Noûs* 50 (2): 379–418. <https://doi.org/10.1111/nous.12082>.
- Davidson, Donald. 1963. "Actions, Reasons, and Causes." *The Journal of Philosophy* 60 (23): 685–700. Reprinted in Davidson, 2001, 3–19.
- . 1970. "Mental Events." In *Experience and Theory*, edited by L. Foster and J. Swanson, 79–101. London: Duckworth. Reprinted in Davidson, 2001, 207–25.
- . 2001. *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press.
- . 2005. "Spinoza's Causal Theory of the Affects." In *Truth, Language, and History*, by Donald Davidson, 295–313. Oxford: Clarendon Press.
- Davies, Catherine Glyn. 1990. *Conscience as Consciousness: The Idea of Self-Awareness in French Philosophical Writing from Descartes to Diderot*. Oxford: The Voltaire Foundation.
- Delahunty, R. J. 1985. *Spinoza*. London: Routledge & Kegan Paul.
- Deleuze, Gilles. 1988. *Spinoza: Practical Philosophy*. Translated by Robert Hurley. San Francisco: City Lights Books.
- . 1990. *Expressionism in Philosophy: Spinoza*. Translated by Martin Joughin. New York: Zone Books.
- Della Rocca, Michael. 1996. *Representation and the Mind-Body Problem in Spinoza*. Oxford: Oxford University Press.
- . 2002. "Spinoza's Substance Monism." In *Spinoza: Metaphysical Themes*, edited by Olli Koistinen and J. I. Biro, 11–37. New York: Oxford University Press.
- . 2003a. "A Rationalist Manifesto: Spinoza and the Principle of Sufficient Reason." *Philosophical Topics* 31 (1/2): 75–93.
- . 2003b. "The Power of an Idea: Spinoza's Critique of Pure Will." *Noûs* 37 (2): 200–231.

- . 2008a. “Causation Without Intelligibility and Causation Without God in Descartes.” In *A Companion to Descartes*, edited by Janet Broughton and John Carriero, 235–50. Malden, MA: Blackwell.
- . 2008b. *Spinoza*. London: Routledge.
- . 2010. “PSR.” *Philosophers’ Imprint* 10 (7): 1–13.
- . 2015. “Interpreting Spinoza: The Real Is the Rational.” *Journal of the History of Philosophy* 53 (3): 523–35.
- , ed. 2018. *The Oxford Handbook of Spinoza*. New York: Oxford University Press.
- Donagan, Alan. 1988. *Spinoza*. Chicago: University of Chicago Press.
- Douglas, Alexander. 2018. “Quatenus and Spinoza’s Monism.” *Journal of the History of Philosophy* 56 (2): 261–80. <https://doi.org/10.1353/hph.2018.0023>.
- Fechner, Gustav Theodor. 1851. *Zend-Avesta*. Leipzig: Voss.
- . 1860. *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.
- Floridi, Luciano. 2011. *The Philosophy of Information*. Oxford: Oxford University Press.
- Fodor, Jerry A. 1989. “Making Mind Matter More.” *Philosophical Topics* 17 (1): 59–79.
- Frost, Kim. 2020. “What Could a Two-Way Power Be?” *Topoi* 39 (5): 1141–53.
- Gabbey, Alan. 2008. “Spinoza, Infinite Modes and the Infinitive Mood.” *Studia Spinozana* 16: 41–66.
- Garber, Daniel. 2015a. “Some Additional Words.” *Journal of the History of Philosophy* 53 (3): 537–39.
- . 2015b. “Superheroes in the History of Philosophy: Spinoza, Super-Rationalist.” *Journal of the History of Philosophy* 53 (3): 507–21.
- . 2021. “Spinoza’s Non-Theory of Non-Consciousness.” In Melamed, 2021, 304–27.
- Garrett, Don. 1991. “Spinoza’s Necessitarianism.” In *God and Nature: Spinoza’s Metaphysics*, edited by Yirmiyahu Yovel, 191–218. Leiden: Brill.
- . 2002. “Spinoza’s Conatus Argument.” In *Spinoza: Metaphysical Themes*, edited by Olli Koistinen and J. I. Biro, 127–58. New York: Oxford University Press.
- . 2008. “Representation and Consciousness in Spinoza’s Naturalistic Theory of the Imagination.” In *Interpreting Spinoza: Critical Essays*, edited by Charles Huenemann, 4–25. Cambridge: Cambridge University Press.
- . 2009. “Descartes, Spinoza, and Locke on Extended Thinking Beings.” In *Topics in Early Modern Philosophy of Mind*, edited by Jon Miller, 85–104. Dordrecht: Springer.

- . 2017. “The Indiscernibility of Identicals and the Transitivity of Identity in Spinoza’s Logic of the Attributes.” In *Spinoza’s Ethics: A Critical Guide*, edited by Yitzhak Y. Melamed, 12–42. Cambridge: Cambridge University Press.
- . 2018a. *Nature and Necessity in Spinoza’s Philosophy*. New York: Oxford University Press.
- . 2018b. “Postscript: Consciousness Revisited.” In Garrett, 2018a, 415–23.
- . 2018c. “Representation, Misrepresentation, and Error in Spinoza’s Philosophy of Mind.” In Della Rocca, 2018, 190–203.
- . 2021a. “Spinoza’s Ethical Theory.” In Garrett, 2021b, 282–308.
- , ed. 2021b. *The Cambridge Companion to Spinoza*. 2nd ed. Cambridge: Cambridge University Press.
- Gibb, Sophie. 2015. “The Causal Closure Principle.” *The Philosophical Quarterly* 65 (261): 626–47.
- Gibb, Sophie, E. J. Lowe, and R. D. Ingthorsson, eds. 2013. *Mental Causation and Ontology*. Oxford: Oxford University Press.
- Gibbons, John. 2006. “Mental Causation without Downward Causation.” *The Philosophical Review* 115 (1): 79–103. <https://doi.org/10.1215/00318108-2005-003>.
- Goff, Philip. 2017. *Consciousness and Fundamental Reality*. New York: Oxford University Press.
- Griffin, Michael V. 2008. “Necessitarianism in Spinoza and Leibniz.” In *Interpreting Spinoza: Critical Essays*, edited by Charlie Huenemann, 71–93. Cambridge: Cambridge University Press.
- Griffith, Meghan. 2017. “Agent Causation.” In *The Routledge Companion to Free Will*, edited by Kevin Timpe, Meghan Griffith, and Neil Levy, 72–85. New York: Routledge.
- Gueroult, Martial. 1968. *Spinoza I: Dieu (Ethique, I)*. Hildesheim: Olms.
- . 1974. *Spinoza II: L’âme (Ethique, II)*. Hildesheim: Olms.
- Hampshire, Stuart. 2005. *Spinoza and Spinozism*. Oxford: Clarendon Press.
- Harnad, Stevan. 1990. “The Symbol Grounding Problem.” *Physica D*: 335–46.
- Hegel, Georg Wilhelm Friedrich. 1990. *Lectures on the History of Philosophy: The Lectures of 1825-6*. Edited by Robert F. Brown. Translated by R. F. Brown and J. M. Stewart, with the assistance of H. S. Harris. 3 vols. Berkeley and Los Angeles, CA: University of California Press.
- Heidelberger, Michael. 2004. *Nature from within: Gustav Theodor Fechner and His Psychophysical Worldview*. Translated by Cynthia Kloor. Pittsburgh: University of Pittsburgh Press.
- Heil, John. 2020. *Philosophy of Mind: A Contemporary Introduction*. 4th ed. New York: Routledge.

- Heinämaa, Sara, Vili Lähteenmäki, and Pauliina Remes, eds. 2007a. *Consciousness: From Perception to Reflection in the History of Philosophy*. Dordrecht: Springer.
- . 2007b. “Introduction.” In Heinämaa, Lähteenmäki, and Remes, 2007a, 1–26.
- Hübner, Karolina. 2018. “Spinoza’s Unorthodox Metaphysics of the Will.” In Della Rocca, 2018, 343–69.
- . 2019. “Spinoza on Intentionality, Materialism, and Mind-Body Relations.” *Philosophers’ Imprint* 19 (43): 1–23.
- . 2022. “Representation and Mind-Body Identity in Spinoza’s Philosophy.” *Journal of the History of Philosophy* 60 (1): 47–77.
- Hume, David. 1978. *An Enquiry Concerning Human Understanding*. Edited by P. H. Nidditch. Oxford: Clarendon Press.
- Hutto, Daniel D., and Erik Myin. 2013. *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, Mass.: MIT Press.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Clarendon Press.
- James, Susan. 2020. *Spinoza on Learning to Live Together*. Oxford: Oxford University Press.
- Jaquet, Chantal. 2011. “Ordo et Connexio.” In *The Continuum Companion to Spinoza*, edited by Wiep van Bunge, Henri Krop, Piet Steenbakkers, and Jeroen van de Ven, 278. London: Continuum.
- . 2018. *Affects, Actions and Passions in Spinoza: The Unity of Body and Mind*. Translated by Tatiana Reznichenko. Edinburgh: Edinburgh University Press.
- . 2021. “The Mind-Body Union.” In Melamed, 2021, 296–303.
- Jarrett, Charles. 1978. “The Logical Structure of Spinoza’s ‘Ethics’, Part I.” *Synthese* 37 (1): 15–65.
- . 1982. “Materialism.” *Philosophy Research Archives* 8: 457–97.
- . 1991. “Spinoza’s Denial of Mind-Body Interaction and the Explanation of Human Action.” *Southern Journal of Philosophy* 29 (4): 465–85.
- Jonas, Hans. 1965. “Spinoza and the Theory of Organism.” *Journal of the History of Philosophy* 3 (1): 43–57. <https://doi.org/10.1353/hph.2008.1418>.
- Jorati, Julia. 2017. *Leibniz on Causation and Agency*. Cambridge: Cambridge University Press.
- Jorgensen, Larry M. 2020. “Seventeenth-Century Theories of Consciousness.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2020. <https://plato.stanford.edu/archives/spr2020/entries/consciousness-17th/>.

- Katsafanas, Paul. 2016. *The Nietzschean Self: Moral Psychology, Agency, and the Unconscious*. Oxford: Oxford University Press.
- Kim, Jaegwon. 1993. "The Nonreductivist's Troubles with Mental Causation." In *Mental Causation*, edited by John Heil and Alfred R. Mele, 189–210. New York: Oxford University Press.
- . 1998. *Mind in a Physical World*. Cambridge, Mass.: MIT Press.
- . 2011. *Philosophy of Mind*. 3rd ed. Boulder: Westview Press.
- Kisner, Matthew J. 2011. *Spinoza on Human Freedom: Reason, Autonomy and the Good Life*. Cambridge: Cambridge University Press.
- Kisner, Matthew J., and Andrew Youpa, eds. 2014. *Essays on Spinoza's Ethical Theory*. Oxford: Oxford University Press.
- Klein, Julie R. 2019. "Materializing Spinoza's Account of Human Freedom." In Naaman-Zauderer, 2019, 152–73.
- Koistinen, Olli. 1996. "Causality, Intensionality and Identity: Mind Body Interaction in Spinoza." *Ratio* 9 (1): 23–38.
- , ed. 2009. *The Cambridge Companion to Spinoza's Ethics*. Cambridge: Cambridge University Press.
- . 2018. "Spinoza on Mind." In Della Rocca, 2018, 273–94.
- Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lærke, Mogens. 2012. "Spinoza's Monism? What Monism?" In *Spinoza on Monism*, edited by Philip Goff, 244–61. Basingstoke: Palgrave Macmillan.
- . 2017. "Aspects of Spinoza's Theory of Essence: Formal Essence, Non-Existence, and Two Types of Actuality." In *The Actual and the Possible: Modality and Metaphysics in Modern Philosophy*, edited by Mark Sinclair, 11–44. Oxford: Oxford University Press.
- Lærke, Mogens, Justin E. H. Smith, and Eric Schliesser, eds. 2013. *Philosophy and Its History: Aims and Methods in the Study of Early Modern Philosophy*. New York: Oxford University Press.
- LeBuffe, Michael. 2010a. *From Bondage to Freedom: Spinoza on Human Excellence*. New York: Oxford University Press.
- . 2010b. "Theories about Consciousness in Spinoza's Ethics." *Philosophical Review* 119 (4): 531–63.
- Leibniz, G. W. 1989. *Philosophical Essays*. Edited and translated by Roger Ariew and Daniel Garber. Indianapolis: Hackett.
- Leiter, Brian. 2007. "Nietzsche's Theory of the Will." *Philosopher's Imprint* 7 (7): 1–15.

- Levine, Joseph. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64 (4): 354–61.
- Lewis, C. S. 1960. "Conscience and Conscious." In *Studies in Words*, 181–214. Cambridge: Cambridge University Press.
- Lin, Martin. 2009. "The Power of Reason in Spinoza." In Koistinen, 2009, 258–83.
- . 2012. "Rationalism and Necessitarianism." *Noûs* 46 (3): 418–48.
- . 2019. *Being and Reason: An Essay on Spinoza's Metaphysics*. New York: Oxford University Press.
- . 2020. "The Many Faces of Spinoza's Causal Axiom." In *Causation and Cognition: Perspectives on Early Modern Philosophy*, edited by Dominik Perler and Sebastian Bender, 212–32. New York: Routledge.
- . 2021. "Spinoza on the Metaphysics of Thought and Extension." In Garrett, 2021b, 113–40.
- Lloyd, Genevieve. 2017. "Spinoza and the Problem of Consciousness." In *Consciousness and the Great Philosophers*, edited by Stephen Leach and James Tartaglia, 82–88. Oxon: Routledge.
- Lowe, E. J. 2000. *An Introduction to the Philosophy of Mind*. Cambridge: Cambridge University Press.
- . 2003. "Physical Causal Closure and the Invisibility of Mental Causation." In Walter and Heckmann, 2003, 137–54.
- . 2013. "The Will as a Rational Free Power." In *Powers and Capacities in Philosophy: The New Aristotelianism*, edited by Ruth Groff and John Greco, 172–85. New York: Routledge.
- Lycan, William. 2019. "Representational Theories of Consciousness." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2019. <https://plato.stanford.edu/archives/fall2019/entries/consciousness-representational/>.
- Mach, Ernst. 1897. *Contributions to the Analysis of the Sensations*. Translated by C. M. Williams. La Salle: Open Court Publishing Company.
- Maimon, Solomon. 2018. *The Autobiography of Solomon Maimon*. Edited by Yitzhak Melamed and Abraham Socher. Translated by Paul Reitter. Princeton, NJ: Princeton University Press.
- Malcolm, Norman. 1968. "The Conceivability of Mechanism." *Philosophical Review* 77 (January): 45–72.
- Marshall, Colin R. 2009. "The Mind and the Body as 'One and the Same Thing' in Spinoza." *British Journal for the History of Philosophy* 17 (5): 897–919. <https://doi.org/10.1080/09608780903135030>.

- Marshall, Eugene. 2008. "Spinoza's Cognitive Affects and Their Feel." *British Journal for the History of Philosophy* 16 (1): 1–23. <https://doi.org/10.1080/09608780701789251>.
- . 2013. *The Spiritual Automaton: Spinoza's Science of the Mind*. Oxford University Press.
- Martin, Christopher. 2007. "Consciousness in Spinoza's Philosophy of Mind." *Southern Journal of Philosophy* 45 (2): 269–87.
- . 2010. "A New Challenge to the Necessitarian Reading of Spinoza." In *Oxford Studies in Early Modern Philosophy V*, edited by Daniel Garber and Steven Nadler, 25–70. Oxford: Clarendon Press.
- McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98 (391): 349–66.
- McKenna, Michael, and D. Justin Coates. 2020. "Compatibilism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/compatibilism/>.
- Melamed, Yitzhak Y. 2010. "Acosmism or Weak Individuals?: Hegel, Spinoza, and the Reality of the Finite." *Journal of the History of Philosophy* 48 (1): 77–92.
- . 2012. "Spinoza's Deification of Existence." In *Oxford Studies in Early Modern Philosophy VI*, edited by Daniel Garber and Donald Rutherford, 75–104. Oxford: Oxford University Press.
- . 2013a. "Charitable Interpretations and the Political Domestication of Spinoza, or, Benedict in the Land of the Secular Imagination." In Lærke, Smith, and Schliesser, 2013, 258–77.
- . 2013b. *Spinoza's Metaphysics: Substance and Thought*. Oxford: Oxford University Press.
- . 2017. "The Causes of Our Belief in Free Will: Spinoza on Necessary, 'Innate,' yet False Cognition." In *Spinoza's Ethics: A Critical Guide*, edited by Yitzhak Y. Melamed, 121–41. Cambridge: Cambridge University Press.
- . 2018. "The Building Blocks of Spinoza's Metaphysics: Substance, Attributes and Modes." In Della Rocca, 2018, 84–113.
- , ed. 2021. *A Companion to Spinoza*. Hoboken, NJ: Wiley Blackwell.
- Melamed, Yitzhak Y., and Martin Lin. 2021. "Principle of Sufficient Reason." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/sufficient-reason/>.
- Mercer, Christia. 2019. "The Contextualist Revolution in Early Modern Philosophy." *Journal of the History of Philosophy* 57 (3): 529–48.

- Miller, Jon. 2007. "The Status of Consciousness in Spinoza's Concept of Mind." In Heinämaa, Lähteenmäki, and Remes, 2007a, 203–20.
- Montero, Barbara. 2003. "Varieties of Causal Closure." In Walter and Heckmann, 2003, 173–87.
- Morrison, John. 2013. "The Relation between Conception and Causation in Spinoza's Metaphysics." *Philosophers' Imprint* 13 (3): 1–17.
- Naaman-Zauderer, Noa, ed. 2019. *Freedom, Action, and Motivation in Spinoza's Ethics*. New York: Routledge.
- Nadler, Steven. 2006. *Spinoza's Ethics: An Introduction*. Cambridge: Cambridge University Press.
- . 2008. "Spinoza and Consciousness." *Mind* 117 (467): 575–601.
- Nagel, Thomas. 1974. "What Is It like to Be a Bat?" *Philosophical Review* 83 (October): 435–50.
- . 1979. "Panpsychism." In *Mortal Questions*, by Thomas Nagel, 181–95. Cambridge: Cambridge University Press.
- . 1986. *The View from Nowhere*. New York: Oxford University Press.
- Newlands, Samuel. 2011. "Hegel's Idealist Reading of Spinoza." *Philosophy Compass* 6 (2): 100–108.
- Noorloos, Ruben. 2021. "‘Except God, No Substance Can Be Conceived’: Spinoza on Other Substances." *Analysis* 81 (4): 656–65.
- O'Connor, Timothy, and Christopher Franklin. 2020. "Free Will." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/freewill/>.
- Papineau, David. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry M. Loewer, 3–36. Cambridge: Cambridge University Press.
- . 2002. *Thinking About Consciousness*. Oxford: Clarendon Press.
- Pauen, Michael. 2011. "Spinoza and the Theory of Identity (2p1–13)." In *Spinoza's Ethics: A Collective Commentary*, edited by Renz Hampe, Ursula Renz, and Robert Schnepf, 81–98. Leiden: Brill.
- Pereboom, Derk. 2015. "The Phenomenology of Agency and Deterministic Agent Causation." In *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology: Essays in Honor of Charles Guignon*, edited by Hans Pedersen and Megan Altman, 277–94. Dordrecht: Springer. https://doi.org/10.1007/978-94-017-9442-8_17.
- Perler, Dominik. 2014. "Suárez on Consciousness." *Vivarium* 52 (3–4): 261–86. <https://doi.org/10.1163/15685349-12341277>.

- . 2018. *Feelings Transformed: Philosophical Theories of the Emotions, 1270-1670*. Translated by Tony Crawford. New York: Oxford University Press.
- Peterman, Alison. 2017. “Descartes and Spinoza: Two Approaches to Embodiment.” In *Embodiment: A History*, edited by Justin E. H. Smith, 215–40. New York: Oxford University Press.
- Pruss, Alexander R. 2006. *The Principle of Sufficient Reason: A Reassessment*. Cambridge: Cambridge University Press.
- Ravven, Heidi M. 2014. “Moral Agency without Free Will: Spinoza’s Naturalizing of Moral Psychology in a Maimonidean Key.” In *Spinoza and Medieval Jewish Philosophy*, edited by Steven Nadler, 128–51. Cambridge: Cambridge University Press.
- Renz, Ursula. 2018. *The Explainability of Experience: Realism and Subjectivity in Spinoza’s Theory of the Human Mind*. Translated by Mark Ilseemann. New York: Oxford University Press.
- Renz, Ursula, and Barnaby R. Hutchins. 2021. “Spinoza on Human and Divine Knowledge.” In Melamed, 2021, 251–64.
- Rice, Lee. 1990. “Reflexive Ideas in Spinoza.” *Journal of the History of Philosophy* 28 (2): 201–11.
- . 1999. “Paradoxes of Parallelism in Spinoza.” *Iyyun: The Jerusalem Philosophical Quarterly* 48 (January): 37–54.
- Richardson, R. C. 1982. “The ‘Scandal’ of Cartesian Interactionism.” *Mind* 91 (361): 20–37.
- Robb, David, and John Heil. 2021. “Mental Causation.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2021/entries/mental-causation/>.
- Robinson, Howard. 2020. “Dualism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2020/entries/dualism/>.
- Rozemond, Marleen. 1999. “Descartes on Mind-Body Interaction: What’s the Problem?” *Journal of the History of Philosophy* 37 (3): 435–67.
- Sangiaco, Andrea. 2015. “The Ontology of Determination: From Descartes to Spinoza.” *Science in Context* 28 (4): 515–43.
- Schmidt, Andreas. 2009. “Substance Monism and Identity Theory in Spinoza.” In Koistinen, 2009, 79–98.
- Seager, William E., ed. 2020. *The Routledge Handbook of Panpsychism*. New York: Routledge.
- Shapiro, Lisa, trans. 2007. *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. The Other Voice in Early Modern Europe. Chicago: University of Chicago Press.

- Shein, Noa. 2009. "The False Dichotomy between Objective and Subjective Interpretations of Spinoza's Theory of Attributes." *British Journal for the History of Philosophy* 17 (3): 505–32. <https://doi.org/10.1080/09608780902986631>.
- Shoemaker, Sydney. 2007. *Physical Realization*. Oxford: Oxford University Press.
- . 2013. "Physical Realization without Preemption." In Gibb, Lowe, and Ingthorsson, 2013, 35–57.
- Simmons, Alison. 2001. "Changing the Cartesian Mind: Leibniz on Sensation, Representation and Consciousness." *The Philosophical Review* 110 (1): 31–75.
- Skrbina, David. 2014. "Dualism, Dual-Aspectism, and the Mind." In *Contemporary Dualism: A Defense*, edited by Andrea Lavazza and Howard Robinson, 220–44. New York: Routledge.
- Spinoza, Benedictus de. 1985–2016. *The Collected Works of Spinoza*. Translated by E. M. Curley. 2 vols. Princeton, NJ: Princeton University Press.
- Steward, Helen. 2012. *A Metaphysics for Freedom*. Oxford: Oxford University Press.
- Stoljar, Daniel. 2018. Review of Goff (2017). *Notre Dame Philosophical Reviews*. <https://ndpr.nd.edu/reviews/consciousness-and-fundamental-reality/>.
- Strawson, Galen. 1986. *Freedom and Belief*. Oxford: Clarendon Press.
- Stubenberg, Leopold. 2018. "Neutral Monism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2018/entries/neutral-monism/>.
- Thiel, Udo. 2011. *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. New York: Oxford University Press.
- Tye, Michael. 2018. "Qualia." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2018 Edition. <https://plato.stanford.edu/archives/sum2018/entries/qualia/>.
- Van Cleve, James. 1990. "Mind-Dust or Magic? Panpsychism Versus Emergence." *Philosophical Perspectives* 4: 215–26.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
- . 2015. *Metaphysics*. 4th ed. Boulder: Westview Press.
- Velmans, Max. 2009. *Understanding Consciousness*. 2nd ed. Hove, East Sussex: Routledge.
- Viljanen, Valtteri. 2011. *Spinoza's Geometry of Power*. Cambridge: Cambridge University Press.
- Walter, Sven, and Heinz-Dieter Heckmann, eds. 2003. *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter: Imprint Academic.

- Wartofsky, Marx. 1973. "Action and Passion: Spinoza's Construction of a Scientific Psychology." In *Spinoza: A Collection of Critical Essays*, edited by Marjorie Grene, 329–53. Garden City, NY: Anchor Books.
- Wegener, Mai. 2009. "Der Psychophysische Parallelismus." *N.T.M.* 17 (3): 277–316. <https://doi.org/10.1007/s00048-009-0344-3>.
- Wilson, Margaret. 1996. "Spinoza's Theory of Knowledge." In *The Cambridge Companion to Spinoza*, edited by Don Garrett, 89–141. Cambridge: Cambridge University Press.
- . 1999a. *Ideas and Mechanism*. Princeton, NJ: Princeton University Press.
- . 1999b. "Objects, Ideas, and 'Minds': Comments on Spinoza's Theory of Mind." In Wilson, 1999a, 126–40.
- . 1999c. "Spinoza's Causal Axiom (*Ethics* I, Axiom 4)." In Wilson, 1999a, 141–65.
- Wolfson, Harry Austryn. 1934. *The Philosophy of Spinoza: Unfolding the Latent Processes of His Reasoning*. 2 vols. Cambridge, MA: Harvard University Press.
- Wundt, Wilhelm. 1894. "Über Psychische Causalität und das Prinzip des Psychophysischen Parallelismus." *Philosophische Studien* 10: 1–124.
- Yakira, Elhanan. 2015. *Spinoza and the Case for Philosophy*. Cambridge: Cambridge University Press.
- Zhong, Lei. 2014. "Sophisticated Exclusion and Sophisticated Causation." *Journal of Philosophy* 111 (7): 341–60. <https://doi.org/10.5840/jphil2014111724>.
- . 2017. "平行主義的復興 (Parallelism Revived)." Translated by Dong Xin. 《自然辯證法通訊》 39: 1–10.