UNIL | Université de Lausanne
Faculté des lettres

FACULTÉ DES LETTRES

SECTION DE PHILOSOPHIE

Interacting Minds in the Physical World

THÈSE DE DOCTORAT

présentée à la

Faculté des lettres
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès lettres

par

Alin Christoph Georg Cucu

Directeur de thèse :

Prof. Dr. Michael Esfeld

LAUSANNE 2022

UNIL | Université de Lausanne
Faculté des lettres

IMPRIMATUR

Le Décanat de la Faculté des lettres, sur le rapport d'une commission composée de :

Directeur de thèse :

Monsieur Michael Esfeld — Professeur, Faculté des lettres, UNIL

Membres du jury :

Monsieur Jeffrey Koperski — Professeur, Saginaw Valley State University, USA

Monsieur Daniel von Wachter — Professeur de philosophie, Internationale Akademie für Philosophie, Liechtenstein

autorise l'impression de la thèse de doctorat de

MONSIEUR ALIN CHRISTOPH GEORG CUCU

intitulée

Interacting Minds in the Physical World

sans se prononcer sur les opinions du candidat / de la candidate.

La Faculté des lettres, conformément à son règlement, ne décerne aucune mention.

Lausanne, le 15 novembre 2022

Léonard Burnand
Doyen de la Faculté des lettres

Year: 2022

Interacting Minds in the Physical World

Alin Christoph Georg Cucu

Abstract

Mental causation, idea that it is us – via our minds – who cause bodily actions is as commonsensical as it is indispensable for our understanding of ourselves as rational agents. Somewhat less uncontroversial, but nonetheless widespread (at least among ordinary people) is the idea that the mind is non-physical, following the intuition that what is physical can neither act nor think nor judge morally. Taken together, and cast into a metaphysical thesis, the two intuitions yield interactive dualism: the view that human persons and their minds are non-physical but can nonetheless interact with their bodies, most notably through their brains.

This thesis has two main objectives: first, to defend interactive dualism against objections, and second, to show how it can blend in with a physical world in which laws of nature hold. The first part (chapter 1) consists in a brief motivation of interactive dualism as opposed to non-interactive dualism. I argue that non-interactive, epiphenomenalist dualism sacrifices so many crucial aspects of our human existence that interactive dualism is highly to be preferred to it, barring even stronger counterarguments against the latter.

It is those putative counterarguments I address in parts II and III. Part II takes on the philosophical objections from the causal closure of the physical and from causal heterogeneity. The former takes the success of physics and physiology as basis for the doctrine of the causal closure of the physical (CCP). I argue against this that there is not only no convincing argument for CCP, but also that any science-based belief in CCP cannot be epistemically justified. As regards the causal heterogeneity objection, it is 'weighed and found wanting', because it relies on unwarranted assumptions about causation.

In part III, I examine the objection from energy conservation. It roughly says that if interactive dualism were true, then energy would not be conserved, which physics taught us cannot be the case. My reply is that the underlying conception of energy (and momentum) conservation is wrong-headed and not the one that actual physicists use. Instead of being categorical and global, conservation laws are conditional and local, thereby making natural room for mental interaction. Some dualists, however, have sought to make interactive dualism conservation-friendly, notably by invoking quantum physics; I show that these attempts are unnecessary and create more problems than they solve. Finally, in chapter 9, I turn the tables on non-interactionists by investigating current neurophysiological literature on volitional actions, which, though not addressing the question directly, still encourages the interactive dualist picture more than a non-interactionist one.

Part IV is about the interplay between interactive dualism and the laws of nature. The ultimate goal is, if possible, to come up with a theory of the laws of nature that explains the lawlike behavior of nature and at the same time makes room for interaction. I begin by pointing out that the laws of nature are or at least should be what physicalists worry about (chapter 10). I then proceed with a historical survey on the development of the notion of laws of nature (chapter 11) that sheds light on its theistic origin. This is followed by an inquiry into the question how a law of nature could possibly be broken (chapter 12) and a survey of the extant metaphysical theories of the laws of nature with special regard to their receptivity to interaction (chapter 13). Finally, in chapter 14 I develop a theory of the laws of nature both faithful to their divine origin and the possibility of mental interaction: dispositionalist divine decretalism, a synthesis of dispositionalism and Jeffrey Koperski's divine decretalism.

Résumé

La causalité mentale, idée selon laquelle c'est nous – via notre esprit – qui causons les actions corporelles, est aussi banale qu'indispensable à notre compréhension de nous-mêmes en tant qu'agents rationnels. Un peu moins incontestée, mais néanmoins répandue (au moins parmi les gens ordinaires) est l'idée que l'esprit est non-physique, suivant l'intuition que ce qui est physique ne peut ni agir, ni penser, ni juger moralement. Prises ensemble, et exprimées dans une thèse métaphysique, ces deux intuitions donnent le dualisme interactif : l'opinion selon laquelle les personnes humaines et leur esprit sont non-physiques mais peuvent néanmoins interagir avec leur corps, plus particulièrement par le biais de leur cerveau.

Cette thèse a deux objectifs principaux : premièrement, défendre le dualisme interactif contre les objections, et deuxièmement, montrer comment il peut s'harmoniser avec un monde physique dans lequel les lois de la nature tiennent. La première partie (chapitre 1) consiste en une brève motivation du dualisme interactif par rapport au dualisme non-interactif. Je soutiens que le dualisme non interactif, épiphénoméniste, sacrifie tant d'aspects cruciaux de notre existence humaine que le dualisme interactif doit lui être hautement préféré, sauf contre-arguments encore plus forts à son encontre.

Ce sont ces contre-arguments putatifs que j'aborde dans les parties II et III. La partie II s'attaque aux objections philosophiques de la fermeture causale de la physique et de l'hétérogénéité causale. La première prend le succès de la physique et de la physiologie comme base de la doctrine de la fermeture causale du physique (Causal Closure of the Physical, CCP). J'argumente contre cela qu'il n'y a non seulement aucun argument convaincant pour CCP, mais aussi que toute croyance scientifique en CCP ne peut être justifiée épistémiquement. Quant à l'objection de l'hétérogénéité causale, elle est "pesée et jugée insuffisante", car elle repose sur des présuppositions injustifiées concernant la causalité.

Dans la troisième partie, j'examine l'objection de la conservation de l'énergie. Elle dit en gros que si le dualisme interactif était vrai, alors l'énergie ne serait pas conservée, ce qui ne peut pas être le cas, comme la physique nous l'a appris. Ma réponse est que la conception sous-jacente de la conservation de l'énergie (et de la quantité de mouvement) est erronée et n'est pas celle qu'utilisent les physiciens. Au lieu d'être catégoriques et globales, les lois de conservation sont conditionnelles et locales, laissant ainsi une place naturelle à l'interaction mentale. Quelques dualistes ont toutefois cherché à rendre le dualisme interactif favorable à la conservation, notamment en invoquant la physique quantique ; je montre que ces tentatives sont inutiles et créent plus de problèmes qu'elles n'en résolvent.

Enfin, au chapitre 9, je retourne la situation contre les non-interactionnistes en étudiant la littérature neurophysiologique actuelle sur les actions volitives, qui, bien que n'abordant pas directement la question, encourage tout de même davantage l'image dualiste interactive que celle du non-interactionniste.

La quatrième partie porte sur l'interaction entre le dualisme interactif et les lois de la nature. Le but ultime est, si possible, d'aboutir à une théorie des lois de la nature qui explique le comportement de la nature conforme aux lois et qui, en même temps, fait de la place à l'interaction. Je commence par souligner que les lois de la nature sont ou du moins devraient être ce qui préoccupe les physicalistes (chapitre 10). Je procède ensuite à une enquête historique sur le développement de la notion de « lois de la nature » (chapitre 11) qui met en lumière son origine théiste. Cette étude est suivie d'une enquête sur la question de savoir comment une loi de la nature pourrait être violée (chapitre 12) et d'une étude des théories métaphysiques existantes des lois de la nature, en particulier en ce qui concerne leur réceptivité à l'interaction (chapitre 13). Enfin, au chapitre 14, je développe une théorie des lois de la nature à la fois fidèle à leur origine divine et à la possibilité d'une interaction mentale : le décrétalisme divin dispositionnaliste (dispositionalist divine decretalism), une synthèse du dispositionnalisme et du décrétalisme divin (divine decretalism) de Jeffrey Koperski.

Acknowledgments

I would like to thank first of all Michael Esfeld. He took me under his wings when I was in need of supervision and of an institutional home. Without his philosophical guidance, often by means of healthy controversy, this thesis would probably have been stuck in the mire of idiosyncratic dabbling with secondary themes.

I also want to thank Daniel von Wachter, who gave me the extraordinary chance as a lateral entrant to leap directly to a philosophical doctorate. Without this courageous and generous offer, I would probably still be a securely paid but deeply unhappy high school teacher. He also helped and supported me to do my first steps in philosophical writing.

J. Brian Pitts has tremendously helped me get my head around one of the centerpieces of this thesis' punchline, the Noether theorem, in connection with the construal of conservation principles as local. I cannot thank him enough for this invaluable help.

On a broader level, Richard Leibrandt has been an amazing inspiration to my philosophical thinking. His unvarnished but astute questions prompted me to think through a number of big meta-philosophical questions. I hope we will soon do a paper together.

Jeffrey Koperski and Eddy Keming Chen were prepared to enter into an extended e-mail exchange with me over their respective accounts of the laws of nature. I am grateful for that.

Chuck White has been a dear friend to me, despite our physical distance, and something like my academic cheerleader. I owe him a lot in terms of the motivational power he incited in me. Benjamin Graf likewise encouraged me along the way, affirming my vocation as a thinker, despite his grief over seeing me move away precisely because of philosophy.

Finally, I'd like to thank all those at conferences, workshops or the "Recherches actuelles en philosophie des sciences" seminar at the University of Lausanne. Every little comment or question you had the courage to put out has had some positive effect on the quality of this thesis.

Table of Contents

Introduction

> Babies are natural-born dualists. (Bloom 2004, Preface)

> Interactionism is the view that mind and body – or mental events and physical events – causally influence each other. That this is so is one of our common-sense beliefs, because it appears to be a feature of everyday experience. (Robinson 2020, 3.1)

> What sort of action would constitute intervention? And why would it be a bad thing? Perhaps my questions can be put like this: "What is intervention, that Thou must be mindful of it?" (Plantinga 2008, 369)

This thesis defends the metaphysical position of interactive dualism, which is roughly the view that the mind and the body are distinct entities that interact with each other. But although interactive dualism can be framed and examined as a distinctly metaphysical theory, it is also in some sense a default belief we have about ourselves, where 'belief' may not even be the appropriate term: according to psychologist Paul Bloom, "babies are natural-born dualists", but arguably babies cannot hold beliefs, at least not in the canonical way. The *gestalt* of interactive dualism seems to be something so deeply entrenched and enmeshed in human nature that it appears more like a pre-rational *condition* for acting, thinking and feeling like a human being. Only later in life does it surge to the surface of cognition and take the form of a more or less elaborate metaphysical belief. That belief is, roughly speaking, that (i) our selves are immaterial and distinct, if not separable from our bodies (those selves often conceived of and termed as souls) and that (ii) this immaterial self can interact with the body, and through the body with the physical world. I shall refer to this ontology of human beings as *interactive psychophysical dualism* (or to its non-interactive variant as just *psychophysical dualism*). (The rough sketch just given will be elaborated below).
It is thus wrong-headed to conceive of interactive dualism as a theoretical construct solely designed to explain phenomena like our impression of acting freely, and of our selves being distinct from our bodies. Interactive dualism is the default belief of the majority of ordinary people, prior to exposition to scientific or philosophical education. But of course that implicit belief can be made explicit, and spelled out in arbitrary metaphysical detail. That metaphysical explication began contemporaneous with the dawn of Western philosophy in ancient Greece, notably with Plato[1], and had eminent defenders in virtually all epochs (see (Goetz and Taliaferro 2011)). Thus even in terms of academic history, interactive dualism has a long-standing tradition. It is also true that this ontology of the human person was not shared by all philosophers, and that from the very beginning[2]. The rejection was particularly violent and vitriolic around the Second World War and for some timer after, in the age of verificationism and the identity theory of the mind. For example, Gilbert Ryle scornfully dismissed the notion of an immaterial self as a "ghost in the machine" (Ryle 1973, 5). However, in the last four decades, psychophysical dualism has made a stunning comeback in the shape of the works of such eminent philosophers such as John Foster (Foster 1991), Richard Swinburne (Swinburne 1996; 1997; 2013), Alvin Plantinga (Plantinga 2007; 2008), Charles Taliaferro (Taliaferro 1994), E.J. Lowe (Lowe 1992; 1996; 2006), Uwe Meixner (Meixner 2004; 2008;

[1] It should be heeded, however, that Plato was in general by far not as interested in spelling out a precise metaphysics as was Aristotle.

[2] For example, the ancient Greek atomists believed that everything, even human souls, consists of atoms (albeit not in a modern, scientific sense), which puts souls in the same ontological category as bodies.

2010; 2019), J.P. Moreland (Moreland and Rae 2000; Moreland 2018b; 2018a) and to a lesser degree David Chalmers (Chalmers 1996)[3].

Despite these recent developments, interactive psychophysical dualism is still, by and large, a view whose virtually innate acceptance stands in stark contrast to its widespread rejection at least in Western academia. Without attempting to give an explanation – the following represents no more than a description of the situation – things seem to be as follows: interactive dualism is part of a non-verbal, implicit (in fact, innate) pre-theoretical worldview which is impervious to metaphysical and scientific objections, but becomes vulnerable to them once it is made verbal and explicit (see fig. 0.1).

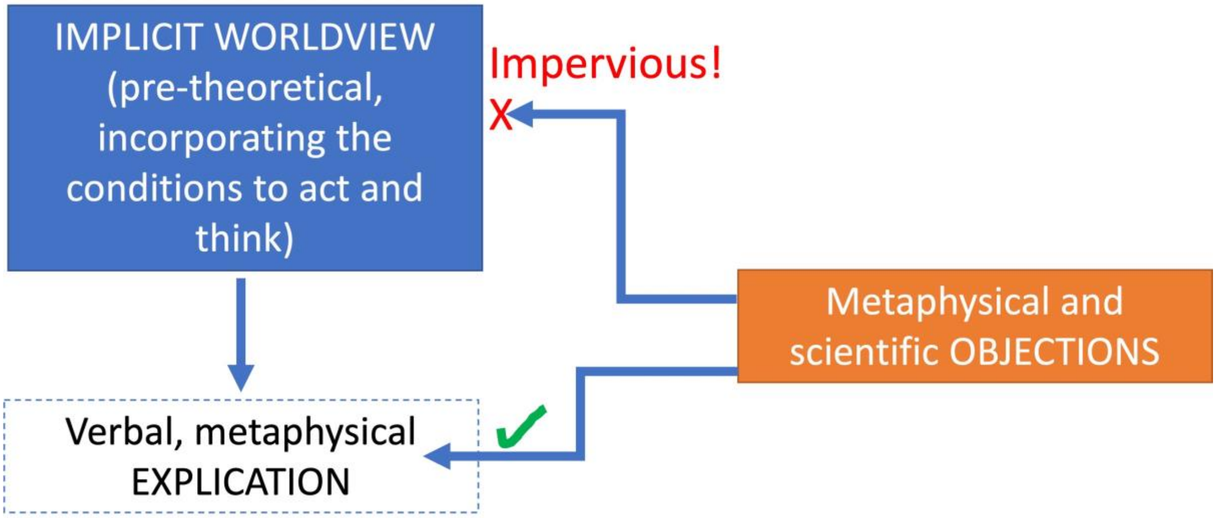

Figure 0.1

Let's stake out the terrain covered by 'psychophysical dualism' in more detail. First, the qualification 'psychophysical' is clearly important, for there are some sorts of dualism which are not even ontological dualisms, let alone dualisms about mind and matter. There is, most importantly, explanatory dualism (Maxwell 2000), which holds that there are two sorts of complementary explanations for phenomena involving persons (namely, scientific/mechanistic and personal explanation) without committing itself to a duality of types of entities or properties. Psychophysical dualism, by contrast, is a dualism of properties or even entities, not just of explanations (although the latter is entailed by the former). Psychophysical dualism comes in two main variants: property dualism and substance dualism. Property dualism is, roughly, the view that some physical entities have *sui generis* mental *properties* which are not physical in any sense – not identical to or constituted by physical properties, and explainable by them at best in the sense that the former supervene on the latter. Substance dualism is the view that, besides matter, there are at least partly mental *substances*, where a substance is, roughly speaking, a bearer of properties. To be more precise, the relevant type of mental substance, namely a human mind, I understand to be

> an essentially characterized particular that (1) has (and is the principle of unity for its) properties but is not had by or predicable of something more basic than it; (2) is an enduring continuant… (Moreland 2018, 103)

[3] Chalmers defends a dualism of properties where no immaterial self in terms of a substance appears.

Thus, substance dualism entails property dualism, but not vice versa. Traditionally, substance dualists have taken what I called 'mental substances' to be identical to persons: for something to be a person is for it to be a mental substance or soul (important exceptions include John Locke and, controversially, Aquinas and most Thomists). This tenet is captured in above definition by the "enduring continuant" feature. A human person endures through time – remains identical over time – because it remains the selfsame mental substance. This is a point worth reflecting on. Isn't the most crucial feature of a mental life (in the broadest sense, not restricted to but including reason, thereby excluding animals) not that it is the mental life of a *person*? Can we even conceive a mental life not belonging to a person? Or, conversely, conceive a person without a mental life? The ontologically more parsimonious property dualism sidesteps these vital questions. I believe that the crucial distinction is between persons as free and rational agents and nature as that which is subject to laws of nature; however, I take the metaphysical *locus* of personhood to obviously be mental substances not physical matter. I shall go on mostly using the term 'mental substance', because for the present debates it is the metaphysical properties of mental substances that is of interest rather than the personhood enshrined in those substances.

It is interesting at this point that discontent with both property as well as substance dualism has recently led some philosophers to embrace what could be called 'image dualism'[4] for lack of a better term. This sort of dualism admits persons as ontologically irreducible, albeit without adding anything (in terms of substances or properties) to a fundamentally physical, material ontology. Something like an image dualism can be found in Kant (who probably is the founding father of this philosophical approach), was famously defended in the 20th century by Sellars (Sellars 1962) and nowadays, for example, by Michael Esfeld (Esfeld 2020). Roughly, image dualism has it that although the world is wholly material, there are two different perspectives on it: the objective 'point of view from nowhere' (i.e. the scientific perspective) and the subjective 'first-person point of view' (i.e. the perspective persons take). Persons are held to be no addition to the material inventory of the world (or, 'no further facts'), but are constituted by the perspective some material objects (human beings) take on the world. This perspective is essentially experiential and normative (at any rate in the case of persons). Though touted by its proponents as a parsimonious way to account for persons while staying faithful to naturalism, image dualism seems more like a description than an explanation (when it talks about perspectives) and simply reasserting the problem instead of solving it (when claiming that persons are no further ontological entities). I therefore believe that in the final analysis, image dualism will collapse into either property dualism or substance dualism, but for our purposes here we can set aside this issue. At any rate, to anticipate the core issue of this thesis: even image dualism remains an 'ontological free lunch' only on some views of the laws of nature. If mental causation ascribed to such 'thin' persons is the truthmaker for things behaving differently than the laws of nature would have it, then the image dualist who wishes to salvage mental causation faces the same problems as an ontological dualist.

The interactive variant of psychophysical dualism, at any rate, continues to be treated with great caution. It is important at this point to realize that 'interactive' dualism is virtually always defended as the interactive version of *substance* dualism, or at least a dualism according to which the human self is a mental *entity*

[4] Taking cue from (Sellars 1962).

(substance)[5], as opposed to an ontology which merely admits of mental *properties* (property dualism). I do not know of any philosopher who defends a genuinely interactive kind of property dualism, and I too ultimately have the substance-based version of psychophysical dualism in mind. The tight connection between substance dualism and interaction is further corroborated by the fact that interaction is one of the main reasons why substance dualism is considered so problematic, as it requires the interaction of the mental substance with the physical world; by contrast, property dualism with its supervenience thesis feels much more at home with non-interaction, i.e. epiphenomenalism (Chalmers 1996, 134ff.). Because of this, and because of the incompatibility of supervenience (a regular ingredient of property dualist accounts) with interaction, I wish to be understood as defending interactive substance dualism, although the concomitant ontological questions around substancehood and the mind-body relation beyond interaction shall not be considered here.

The objective of this thesis is to defend the explicated metaphysical view against objections, while keeping in mind that it is part of something more ancient and larger than philosophical theorizing, such that any metaphysical and scientific defense should pass the test of reality as a whole. As will become clear in due course, the main metaphysical and scientific objections to *interactive* dualism all converge on the question if and how interaction is possible in a world in which laws of nature hold. That, in turn, raises to no small degree the question what laws of nature actually are. The main concern of this thesis will thus be the question "How can interactive dualism and the laws of nature both be true?"

The first chapter (part I) is dedicated to motivating interactive dualism within a theoretical framework. Here, a more precise formulation of mental interaction will make its appearance. I then argue that without some sort of mental interaction, no robust (libertarian) freedom is possible, and evolutionary theory loses all of its explanatory power with respect to mind. The chapter is concluded by some reflections on how we are justified in taking the freedom we feel we have at face value.

Part II is dedicated to answering metaphysical objections against interactive dualism. In chapter 2, I'll discuss the argument that since the physical world is causally closed, non-physical minds cannot act on it. My conclusion will that the so-called causal closure of the physical (CCP) is a principle quite distinct from and not entailed by physics, such that one could maintain physics as a full-fledged science and yet allow for mental interaction. Chapter 3 tackles the objection that it is unintelligible how something non-physical can act on something physical (dubbed the 'causal heterogeneity objection'). My answer is that either the underlying notion of causality is unduly restricted, or whatever is demanded of mental causation cannot be provided (at least not uncontroversially) by physical causation either. I conclude the chapter by attempting to give a positive metaphysical account of dualistic mental causation in particular and of causality in general.

Part III deals with the most important scientific (or scientifically tinged) objection, the objection from conservation laws. Chapter 4 lays out the objection, or rather the way it should be formulated. In chapter 5, I'll treat some popular but inadequate dualistic replies to the conservation objection, while chapter 6 deals with such attempts in the realm of quantum mechanics. An important lesson from all this will be drawn in chapter 7, which presents the simple but solid 'conditionality response' to the charge that

[5] If 'substance dualism' is taken to denote a view according to which both the self and the body are substances, then my view does not count as substance dualism. However, as I will not elaborate on my specific version of dualism here, I take the term 'substance dualism' to sufficiently demarcate views that consider the human self a mental substance from property dualism.

interactive dualism violates conservation principles. With this in mind, chapter 8 explores whether quantum mechanics can *facilitate* interactive dualism, even if it cannot make it energy- and momentum-conservative. Finally, chapter 9 investigates the separate, wholly scientific objection that no empirical evidence for interactive dualism exists. I argue that the reason this for this is mainly a lack of targeted research; if mental interactions in the brain occur, they are most likely very gentle and so their discovery would require directed and minute search. Such empirical data not being available, I examine the known mechanisms of neuronal stimulation (and some hypothetical ones) and conclude that none comes into question to account for volitional actions, such that a mental first cause is the most likely explanation.

The final part IV will deal with the basic issue, underlying the CCP objection as well as the bulk of the conservation laws objection, of how, if at all, interactive dualism and the laws of nature go together. In chapter 10, I explain that how the laws of nature are (or should be) the most fundamental issue with respect to interaction. Prompted by the assumption that this could shed some light on why the fronts are so hardened in the debate between non-interactionists and interactionists, chapter 11 gives a historical survey of the notion of the laws of nature, concluding that the modern use of the term is inextricably linked to theism, and to divine voluntarism in particular. In chapter 12 I examine the question how a law of nature can be broken or violated, which is one main worry underlying the hostility against mental interaction. I conclude that not even on necessitarian theories is it inevitable to construe interaction as law-breaking. Chapter 13 further examines the most important extant, non-theistic theories of the laws of nature regarding their capacity to both account for the stability and predictability of nature as well as to make room for interaction. My result is that the biggest problems of those theories lie, surprisingly, in accounting for the stability of nature, rather than in accommodating interaction. Still, dispositionalism and Chen and Goldstein's MinP theory emerge as the most promising accounts of the investigation. In chapter 14, finally, I present my own, theistic account of the laws of nature. First, I discuss Jeffrey Koperski's 'divine decretalism', in some sense a theistic version of MinP and a legitimate heir to the 17th century voluntarist theories. Despite its strengths, divine decretalism remains too vague with respect to generative causation. My proposal is to fill the gap with dispositions, resulting in what I call 'dispositionalist divine decretalism'. I conclude with a theological assessment of my account.

I. Motivation

1. Why Interactive Dualism?

> [I]f it isn't literally true that my wanting is causally responsible for my reaching, and my itching is causally responsible for my scratching, and my believing is causally responsible for my saying (. . .), if none of that is literally true, then practically everything I believe about anything is false and it's the end of the world. (Fodor 1983)

As insinuated in the introduction, the question "Why *interactive* dualism?" should be understood here not so much as an apology of interactive dualism but rather as an inquiry into what we lose by abandoning it. It is informative to reflect a bit on this notion of 'abandoning' interactive dualism before we move on. Since interactive dualism is a two-dimensional view – dualistic and interactionistic – what could be abandoned is either the 'dualism' part or the 'interaction' part. Of course, if psychophysical dualism is given up for some ontological monism (the most immediate candidate being physicalism), then it makes no longer sense to speak of 'interaction', for there are no two distinct types of entities that could interact. We might instead broaden the notion of interaction to 'causal efficacy'. This enables us to get in with a vexed question that even haunts *physicalists* who reject dualism. Among those people is Jerry Fodor (see introductory quote). Whether the causal inefficacy of the mind would be the "end of the world", as Fodor claims (or quips?) is uncertain, but, as Jaegwon Kim (another physicalist) comments, "it surely seems like the end of a world that includes Fodor and the rest of us as agents and cognizers" (1998, 32). It is therefore unsurprising that one of Fodor's as well as Kim's (and other physicalists') main projects was to insure the causal efficacy of the mind without abandoning their physicalism or, more importantly (since there lies physicalism's heart), the doctrine of the causal closure of the physical (which I will deal with in chapter 2), a project whose success even Kim himself was skeptical of (Kim 1998, 118-120), because it remains far from clear how the mental can have genuine *sui generis* causal efficacy in a physical world purely governed by physical causes (the tenet of CCP). The upshot is that the causal efficacy of the mind is something one should by no means abandon quickly, and which apparently cannot even be guaranteed by embracing physicalism.

In much the same way in which a negative of an image inverts its colors, epiphenomenalist dualism (in short epiphenomenalism) is physicalism's negative. Epiphenomenalism deliberately drops the causal efficacy (interaction) part and keeps dualism, the motivation being to not violate CCP. According to epiphenomenalism, our non-physical consciousness registers what impinges on our sense organs, including our own actions, but it is not causally active in bringing about these actions. According to epiphenomenalism, consciousness is just a non-physical *epiphenomenon*. That this view is rather unpalatable is immediately obvious; however, the fact that some philosophers (first and foremost David Chalmers) are prepared to sacrifice the mind's causal potency for dualism reveals as how crucial they must consider the dualistic picture of the mind. As indicated, I will not argue here why a dualistic ontology of the mind is required, but rather assume this and move on to showing what happens when one drops interactionism. In other words, the target view in the following sections should be understood as being epiphenomenalism[6].

[6] A third way dualists can take apart from interactionism and epiphenomenalism is psychophysical parallelism as defended first and foremost by Malebranche (Malebranche 1980) and Leibniz (Leibniz 1973). On Malebranche's occasionalist view, it is God who brings about the physical events at whose realization the mind had been aiming; Leibniz bolstered God's influence even further by claiming that He

1.1 What it means for our mental events to make a difference in the physical world

Let us first state more precisely what is meant by a non-physical mind being causally active in the physical world. Interactive dualism is, as a first approximation, the thesis that at least some non-physical mental events cause some physical events. For example, my effective intention (or volition, endeavoring, trying, choice event[7]) – at any rate, a mental event – to go downstairs and grab a cup of coffee causes my limbs to move in such a way that I actually move downstairs and prepare a cup of coffee. Interactive dualism is committed to the 'common sense' picture according to which mental events can cause physical events.

More can be said to render the view more precise and to delineate it from epiphenomenalism. In particular, how can we make clear that it was really a mental event M_1 that caused a physical event P_2 and not a prior physical event P_1, given that both M_1 and P_1 were present prior to P_2? We have to say, of course, that P_2 would not have occurred had M_1 not occurred (but only P_1). Let's therefore try the following counterfactual:

> **(ID-CF) For any physical event P that constitutes or is part of a volitional action whose content is contained in a volitionally structured mental event M which predates the action, it is the case that had M not occurred, P would not have occurred.**

Some qualifications are in order here. First, though other kinds of body movements may in principle also be of interest, I shall focus on willful (volitional) actions, because the putative causal role of the mental is clearest in this type of cases. Second, a "volitionally structured mental event" I understand a complex mental event that has the teleological nature of being directed at the realization of a certain state of affairs (the action), and is thus clearly differentiated from other types mental events such as beliefs or emotions. The somewhat bulky term also leaves open the possibility that a pure volition is part of but not identical to a mental event of the type under discussion here. Third, the mental event must predate the action, otherwise the typical order of causality is broken and M could be construed as just 'tracking', but not causing, the action. Fourth, for this counterfactual to truly render what is meant by interactive dualism, one must reject the supervenience of the mental on the physical. Consider figure 1.1:

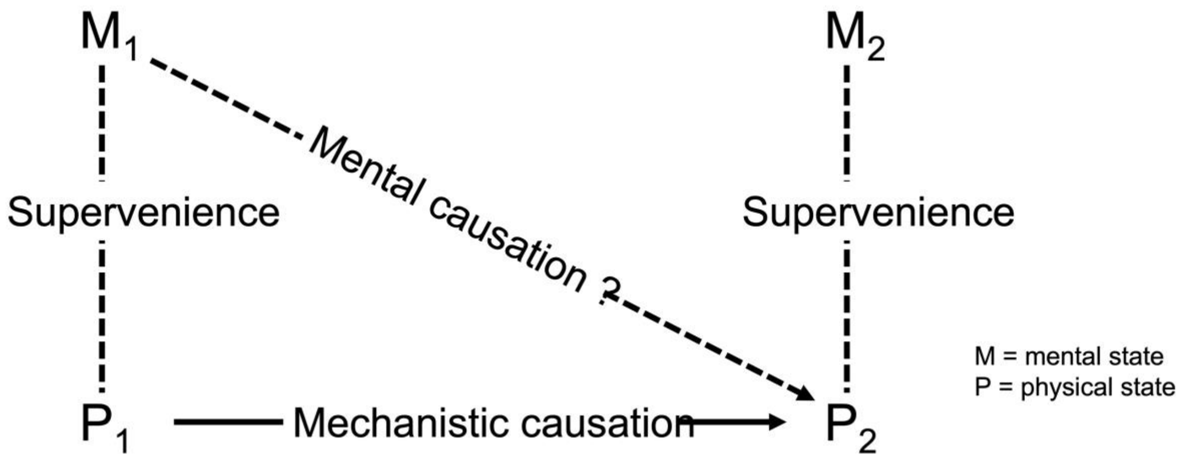

Figure 1.1: Mechanistic mental causation

---

established, at creation, a perfect harmony between the mental events of human beings (which He foreknew) and physical events (which He determined). None of those views counts as interactive in the sense of interest here, though.

[7] These are different terms used, in this order, by (Hasker 1999; Chisholm 2014; Swinburne 2013; Von Wachter 2003; 2009a)

On this picture, even if the mental events are construed dualistically, their causal efficacy is at best highly dubitable. The reason is that because of supervenience, the mental events M_1 and M_2 occur simultaneously with their physical base events P_1 and P_2, respectively; but P_1 is a sufficient cause for P_2, and so no causal contribution seems to remain for M_1, and this even in case CCP is rejected (hence the dashed diagonal arrow). Barring overdetermination, on a supervenience view ID-CF can very well be true without the mental being causally efficacious in the physical world whatsoever.

With these qualifications, I think ID-CF gives us all we need for interactive dualism and can even be endorsed by Humeans willing to subscribe to some kind of psychophysical correlations (or perhaps even laws).

1.2 Mental causation and freedom

> Our voluntary service he requires,
> Not our necessitated. Such with him
> Finds no acceptance, nor can find; for how
> Can hearts not free be tried whether they serve
> Willing or no, who will but what they must
> By destiny, and can no other choose? (Milton, *Paradise Lost*[8], 305)

A German folk song has it that "the thoughts are free". Indeed, no "dungeon can hold them"; we are free to think whatever we want, even while being bound physically. Under normal circumstances, we can even cast thoughts into action. That we are at least to some extent free agents is a central pillar of our self-conception as human beings, and something on which at least Western societies heavily build. In particular, moral responsibility seems to make no sense without freedom. Also, Christianity, the (historically) main cultural influence in the West, affirms freedom of choice as one constitutive feature of human nature, inimitably captured in the above quote by John Milton.

Freedom of action, though, is inseparably bound to interactive dualism, or so I shall argue.

My pivotal point is Peter van Inwagen's famous consequence argument:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our present acts) are not up to us. (Van Inwagen 1983, 16)

Van Inwagen argues that if determinism is true, we are not free. He does that via the proxy of the laws of nature and the fixedness of the past. Freedom is here understood in the classical libertarian sense: an act is free if (i) it is up to the agent to perform it and (ii) the agent had a genuine possibility to act otherwise[9], where "genuine" is understood not just as the absence of logical contradiction (or the presence of logical possibility), but as robust metaphysical possibility (a more detailed explication of libertarian freedom follows below).

---

[8] (Milton 2005)

[9] Frankfurt's (Frankfurt 1969) counterexamples are ultimately unpersuasive (see Swinburne 2013, 203ff.). Freedom may yet be minimally constructible with requirement (i). That position was held, for example, by Henry of Ghent in the Middle Ages (Porro 2014).

Michael Esfeld formalizes the consequence argument as follows (Esfeld 2020, 95-6):

1. **If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past.**
2. **It is not up to us what went on before we were born.**
3. **It is not up to us what the laws of nature are.**
4. **From (1) to (3): The consequences of these things (including our present acts) are not up to us.**
5. **If our present acts are not up to us, we do not have free will.**
6. **Conclusion: Determinism implies that we do not have free will.**

Now, as Esfeld correctly notes, "what is at issue here is not determinism, but universal laws of nature, be they deterministic or not". Even if the laws were probabilistic, then "the objective probabilities for our acts are the consequences of the laws of nature and events in the remote past" (ibd., 96); but for an act to occur with an objective probability is for it not to be free, at least not in the relevant sense. (The tension between libertarian freedom and objective probabilities will be further investigated in chapter 8 in connection with a quantum-mechanical approach to mental causation.) The consequence argument is valid: if its premises are true, then the conclusion follows with logical necessity. Is it also a sound argument, i.e. are its premises true? Premise 2 is widely accepted, although Esfeld himself has a way out of the dilemma that has it that premise 2 is false (see Esfeld 2020, 102ff.); other than that, it is uncontroversially considered as true, and I shall follow this view. Something similar goes for premise 3. Most people, including most philosophers would accept it as true; however, the Humean view of the laws of nature (see section 13.1.1) allows in principle for human agents to contribute to the determination of what the laws of nature are. For example, Jenann Ismael defends such a view (Ismael 2016, 111).For present purposes, and given the discontent with Humeanism – I myself defend a non-Humean view of the laws of nature – we may follow the mainstream view and consider premise 3 as true.

Remains premise 1. It can be objected that even deterministic laws do not fix the initial conditions of the universe, and thus leave some leeway (the view that even the initial conditions are fixed is called 'super-determinism'). However, that doesn't make any difference, for it is no more up to us what the initial conditions of the universe are than what the laws of nature or the events of the remote past are. Interestingly, it is exactly here that Esfeld dissents: according to him, it is up to us what the initial conditions are. This is possible on his view called Super-Humeanism. Super-Humeanism differs from Humeanism in that the laws of nature do not supervene on a mosaic of instantiations of intrinsic, natural properties (like mass, charge etc.), but on matter points without any intrinsic properties. One consequence of this is that all parameters of physical systems except for spatial location are functional descriptions of the motion of matter points. In other words, physical objects do not intrinsically have mass or charge; rather, they consist of matter points whose characteristic motion pattern we describe in part as having (a particular) mass etc. Now if the only fundamental ontological reality is the motion of matter, then also the initial conditions at the beginning of the universe are a function of the motion of matter, and if human actions can alter the motion pattern, those alterations can either be understood as alterations of the laws or of the initial conditions. Esfeld prefers the latter option: human actions can fix the initial conditions *ex post* without backwards causation, because "their determinate values are not yet there at the initial particle positions and their change. They come out of the particle motion" (Esfeld 2020, 105). Esfeld's solution is undoubtedly a very interesting and sophisticated one, albeit one that forces one to embrace Super-Humeanism. That move, however, exacts a steep price, one that many, including me, are not willing to pay (for more on criticism of Humeanism see section 13.2.1). I will therefore consider premise 1 likewise as true. All in all, every view (deterministic or whatever) which maintains that (a) the past is fixed and (b) that our actions are wholly governed by laws of nature entails that we are not free.

I will now argue that epiphenomenalist dualism (epiphenomenalism) is such a view. If true, we could *salva veritate* replace "determinism" in above reconstruction with "epiphenomenalism". In fact, epiphenomenalism bears the denial of freedom in its very conceptual genes. One of epiphenomenalism's central tenets is the doctrine of the causal closure of the physical (CCP), which says that every physical effect has a (sufficient) physical cause. From this follows that human actions have sufficient antecedent physical causes (barring overdetermination), which in turn have physical causes and so on; and all those physical causes and effects are under the sway of the laws of nature, which renders premises 1-3 true. The rest of the argument flows logically from said premises plus the libertarian conception of freedom (premise 5). Epiphenomenalism thus rules out human freedom.

There is yet another, more concrete way of conceptualizing the problem (see Moreland and Rae 2000, ch. 4). Suppose physical causation is like a sequence of domino bricks. Each brick represents one physical event and is (except for the first one) both effect (of the causality of the preceding brick/physical event) and cause (of the effect of the subsequent brick/physical event). In this picture, the events in our bodies are 'bricks' just like all other physical events. Now if we put a carton with two openings over some of the bricks, called the carton "agent" and declared that whatever happens inside the carton belongs to the "agent" or is under his control, would that make any difference to the way the bricks behave? Clearly not. All one has done is, in a way, a move analogous to the functionalist 'encircling' of functional brain events as mental events. The laws of nature govern how the bricks fall, and the way things unfold before reaching the "agent" are not under his control either.

So, what is ontologically necessary to ground robust libertarian freedom? It is time to spell out in more detail the picture underlying libertarian freedom. What one needs is agent causation understood as (1) the agent being sufficiently distinct from and (2) in control of his body. J.P. Moreland explicates the conditions of libertarianism as follows:

> Person p exercises libertarian agency and freely does some intentional act e (…) just in case
>
> (i) p is a substance that has the active power to bring about e
> (ii) p exerts active power as a first-mover (an "originator") to bring about e
> (iii) p has the categorical ability to refrain from exerting active power to bring about e
> (iv) p acts for the sake of a reason, which serves as the final cause or teleological goal for which p acts.
> (Moreland & Rae 2000, ch. 4)

Criteria (i)-(iii) would constitute a pure voluntary act; it is in combination with (iv) that an intentional voluntary act arises. Criterion (iv) adduces teleology, which cannot be had on an (event-)causal theory of action such as epiphenomenalism offers.

It is quite clear that an agent with these abilities must be something non-physical, at any rate something not under the governance of the laws of nature. Note the strong requirements of "active power" and "first mover": these are features for which we simply have no other ontological 'locus' than agents understood as substances, requirements that, however, animals might fulfill as well; it is the invocation of reasons as teleological elements that requires personhood. These three criteria (i, ii, iv) alone rule out event-causal accounts as well as 'ontologically weak' agent-causal accounts that seek to avoid calling the agent a substance, and/or endowing him with the power to initiate causal chains (such as the accounts of Randolph Clarke (R. Clarke 2000; 2011) or Robert Kane (Kane 1996; 2011).

Interactive substance dualism is an obvious way to spell out above criteria ontologically; perhaps an emergentist agent-causal position like Timothy O'Connor's (see for example O'Connor 2011) is also possible, but since we deal with interactive dualism here, I focus on a dualist metaphysics. The upshot is that with the choice of such a view of human persons, one turns Inwagen's argument (*mutatis mutandis*) 'upside down', in the following manner:

1. **If our present acts are not up to us, we do not have libertarian free will.**
2. **We have libertarian free will.**
3. **Therefore, our present acts are up to us. (1., 2., *modus tollens*)**
4. **If epiphenomenalism is true, then our acts are the consequences of the laws of nature and events in the remote past.**
5. **It is not up to us what went on before we were born.**
6. **It is not up to us what the laws of nature are.**
7. **If epiphenomenalism is true, then our present acts are not up to us. (from 4., 5., 6.)**
8. **Conclusion: Epiphenomenalism is false (from 3., 7., *modus tollens*)**

Of course, the denial of epiphenomenalism entails the denial of the causal closure of the physical. I will argue in chapter 2 why we can safely do that.

1.3 Mental causation and evolution

Perhaps somewhat surprisingly, another argument in favor of interactive dualism comes from evolutionary theory. According to evolutionary theory, only those features of living things prevail (via the selection of corresponding genes) that make for more reproductive fitness. Now clearly the features in question must make a difference in the physical world to be preserved by natural selection (otherwise, they are 'invisible' to natural selection). This requires a dualism of the interactive sort, and thus a non-supervenient dualism (see section 1.1). On supervenient dualism there are serious problems with making mental events count as causally efficacious. But mind should be able to cause physical events directly, for that is the only sure way it can be selected by evolutionary mechanisms[10].

To see why, consider the following simplified evolutionary step from a non-conscious being (NCB) to a conscious being (CB). We assume supervenient epiphenomenalist dualism for the conscious being. It is important to note that NCB has nothing whatsoever worth calling 'mind'; since it is not even phenomenally conscious, it fails to have all the other features of a mental life. (I am here assuming for the moment that

[10] Jeffrey Koperski remarks about this that mind could be "genetically tied to some other mutation that did confer an evolutionary advantage (pleiotropy)" (personal communication). To evaluate the viability of pleiotropy about the mental goes beyond the scope of this thesis; the more natural assumption, especially given our first-hand experience, is that mind could be selected because it was causally efficacious.

consciousness is the basis for mind; that foundation might also be intentionality (see (Brüntrup 2014)), in which case the situation could be adapted *mutatis mutandis*). Now the step from a totally non-mental, purely physical being (NCB) to a being with some degree of mind (CB) must be, in accord with evolutionary theory, a physical one, and a relatively small one: a more or less slight change in DNA sequence, brought about in a random fashion. What ensures that this modified genome is passed on to the next generation, such that the offspring of CB also exhibits mind? The mutation must prove conducive to survival, or more precisely, to reproductive fitness (survival at least until the time of reproduction). Otherwise, it is quite likely that CB will die before having reproduced; or, at any rate, that its progeny will do so, thereby bringing the line of mind to a quick end. Now epiphenomenalist mind *ex hypothesi* makes no difference to anything physical, therefore no difference to the reproductive fitness of CB. But epiphenomenalist mind might be worth than neutral: it might be disadvantageous. Arguably, the operation of consciousness exacts a price in terms of increased energy consumption, which, if not counterbalanced by a property favorable to survival, constitutes a drawback.

A related evolutionary argument in favor of interactive dualism is that only genuine mental causation (sometimes called 'downward causation') can ensure the link between the content of a mental state and the action caused by that state. If mental states were not causally active, any content could be linked to any action. Angus Menuge expresses this worry as follows:

> Why should appropriate beliefs and desires regularly precede actions if natural selection has no way to influence the relationship between the content of those attitudes and the physical movements constitutive of those actions? As far as natural selection is concerned, so long as adaptive behavior is produced, the content of these attitudes can be anything: provided one runs away from a devouring lion, it does not matter if one does so whilst thinking, "What a nice bunch of petunias!" (Menuge 2009, 96)

Clearly, this is not what we observe. The contents of our effective intentions faithfully produce the intended effect. But on an epiphenomenalist picture of mind this would be a breathtaking coincidence.

Things look quite different in case CB's mind is interactive. Then, CB's mental features do make a difference to the probability of CB's survival. Even pre-rational consciousness could have all sorts of effects, for example a sort of instinctive intensified search for a mate or a better distinction between edible and toxic plants and animals. As soon as reason comes in, of course, the survival advantages become obvious and undeniable. Additionally, interactive dualism makes possible an evolutionary explanation for the link between the content of action-directed mental states and their effects in terms of actions. On epiphenomenalism, by contrast, it at least remains utterly mysterious why beings with complex minds like us have been selected as opposed to beings without such minds. Further, it is an incredible coincidence that mental states track behavior. This, together with the other reasons given in the previous sections, constitutes a powerful case for why dualists should opt for an interactive, non-supervenient version of their metaphysical theory of mind.

1.4 Daily experience and the Principle of Credulity

> If there is any fact of experience that we are all familiar with, it's the simple fact that our own choices, decisions, reasonings, and cogitations seem to make a difference to our actual behaviour. (…) Another way to put this point is to say: it is just a plain empirical fact about our behaviour that it isn't predictable in the way that the behaviour of objects rolling down an inclined plane is predictable. And

> the reason it isn't predictable in that way is that we could often have done otherwise than we in fact did. Human freedom is just a fact of experience. (Searle 1984, 87-88)

The preceding arguments started from the assumption that we have freedom and reason in the robust sense explicated, and took from there to show that interactive dualism is a necessary condition for them. But what if those basic assumptions are false? We then would be subject to a very thorough and convincing illusion, for our impression to possess both reason and freedom is very strong indeed. Let us subsume this impression under the heading of 'interactive dualist seeming' (IDS). In this section, I will explore how plausible it is that this seeming is non-veridical.

That this is the way we conceive of ourselves hardly needs argument. It is, as Howard Robinson puts it "one of our common-sense beliefs, because it appears to be a feature of everyday experience" (Robinson 2020). More specifically, we experience ourselves

- sometimes as coerced and sometimes as acting freely and can tell the difference between the two;
- in akratic acts (acts resulting from weakness of will), while we are aware of the difference between a free action done for a reason judged as good and one judged as worse;
- as deliberating, that is pondering different courses of actions and ultimately acting out one.

The interesting question is whether such pre-theoretic experience/belief is enough to ground a full-blown metaphysical theory. In fact, it comes down to a question about the epistemic status of the way things seem to us to be. Note that what we are after is not so much the justification of beliefs[11] but their (probable) truth. The question thus is, what does the fact that something seems to us in a certain way say about the truth of the belief involved?

I propose to adopt what Richard Swinburne calls the 'Principle of Credulity' (PoC):

> [A]ny basic belief (…) is probably true (that is, it is more probable than not that the belief is true) on the believer's evidence that he believes it—in the absence of evidence in the form of other basic beliefs of that believer which makes it probable that he is mistaken. Put in another way, the principle claims that what seems to us to be so probably is so, that our apparent experiences are probably real experiences. (2013, 42)

A similar principle is defended by Michael Huemer under the name of 'Phenomenal Conservatism' (PC):

> (PC) If it seems to S as if P, then S thereby has at least prima facie justification for believing that P. (Huemer 2001, 99)

Common to both principles is that they are concerned with basic (or foundational) beliefs, that is, beliefs on which other beliefs build but which do not rest on other beliefs. While Huemer puts more emphasis on the justification of such beliefs, his PC has roughly the same consequences as Swinburne's PoC when it comes to the acceptance of a belief being true:

> According to phenomenal conservatism, the epistemological default position is to accept things as they appear. The appearances are presumed true, until proven false. (ibd., 100)

[11] As interesting and legitimate the question of justification is, in the present context the aim is to see whether a given belief is true and not so much whether it is rational to hold it.

In what follows, I will cite Swinburne's PoC as representative of the epistemological stance defended by both authors.

PoC holds paradigmatically for perceptual beliefs, but the involved concept of 'seeming' (or appearance) has a wider application. Memory-related beliefs are further examples, as are all phenomenal experiences not related to sense perception. Under this latter category falls the impression that it is me who mentally caused an intended action or thought. This notion requires some further explication. Until now I have been talking of mental causation mainly in terms of mental events causing other (mental or physical) events. But this is just a rather truncated expression, curtailed for analytical purposes, that doesn't do justice to what is really going on when we perform an action or a thought. Rather, it is *us* who cause[12] the thought or action; and at least often the caused thought or action is sustained by us until its consummation. In other words, we do not cause something that, once caused, breaks away from us and continues to exist on its own; normally, the thought or action exists only insofar as it is *my* thought or action, accomplished by me. The seeming that I (rather than physical events in my brain) cause my actions and thoughts is also a basic belief, directly accessible via introspection, and not inferred from other beliefs. Thus, it meets all criteria for being *prima facie* true.

Now, PoC commits people to considering counterevidence. Is such counterevidence available in case of interactive dualism? In a way, the bulk of this thesis deals with objections against interactive dualism. Here, I want to treat two specific arguments not examined elsewhere.

The first objection acknowledges that when we introspect our mental life while acting freely, we are not aware of any physical cause doing the work, but then argues that just being unaware of a cause does not provide conclusive evidence that there is no cause. At the same time, we have ample evidence that only physical forces are at work everywhere in our bodies (see Papineau's argument for physicalism in chs. 3 and 11). In conclusion, we should rather trust in science than in our introspection.

In reply it can be answered that we are not just *unaware of the lack* of something, but *positively aware* of something, namely our agency, and that in a way that suggests that interactive dualism is true. This is all the more significant as one can conceive of an introspective scenario on epiphenomenalism: here, we would be aware of events going on, albeit more as spectators than as actors. At this point, it is also instructive to note where we get our concept of libertarian freedom in the first place: it is not from philosophy textbooks, but from introspective experience. Thus, our experience is not interpreted in light of a metaphysical theory, but constitutes the very stuff to be explained by a metaphysical theory.

Stewart Goetz (Goetz 1997, 197-199) offers a further nuance to the IDS. He observes that when we make choices, we are *simultaneously* aware of our doing so. This simultaneous awareness is not possible on a physicalist account of agency. If we are but conduits for physical events, then, in order to perceive a (mental) event as the cause for an action, we would have to direct our attention to a *past* event (on the widespread assumption that the cause must precede the effect).

However, as insinuated above, there may be *scientific* counterevidence to the IDS. What immediately suggests itself is neurophysiological evidence, in particular the famous 'Libet experiments' (Libet et al. 1983; Libet 1985). I cannot discuss them in detail here nor do I have to, because others have done that already (Mele 2014; Von Wachter 2009, 267ff.; Swinburne 2013, 109ff.; von Wachter 2016; and see the

[12] Some agent causalists avoid calling the agent a cause (e.g. (Ginet 2007)), but this is really just a terminological issue.

comments of esp. Doty, Eccles, Jung, Näätänen, Ringo and Van Gulick in Libet 1985); however, given their widespread influence, it would be odd to omit them altogether. Inspired by the discovery of the readiness potential (RP) by Deecke and Kornhuber (Deecke and Kornhuber 1978), Libet wanted to know more about the temporal connection between the readiness potential, the intention to do an action and the action in terms of muscle contraction. The RP is an electric activity in the supplementary motor are (SMA) of the cortex. Libet had his subjects sit in a chair, an EEG device connected to their skulls and an EMG device connected to their arm. They were told to move their index at will within a given time frame, and to read the time of their intention to move off a giant, rapidly moving clock in front of them. Libet found that RP onset consistently preceded what the test subjects called an 'urge to act' by about 500 ms, while the muscle contraction occurred roughly 200 ms after the urge. This was a remarkable result: a brain event seemed to precede a mental event in a relevant manner, suggesting that the former caused the latter! If this could be generalized, it would mean that our impression of us causing our actions is wrong; it is in fact physical brain events that cause them.

According to one line of criticism of this interpretation of the Libet findings (see Mele 2014), it is precisely this generalization that cannot be done. To conclude from one very artificial and highly simplified laboratory setting to complex situations in everyday life is simply invalid, those critics say. Indeed, even if one accepts that one's brain determines the time of a finger flexion, it does not follow that it also determines complex moral and prudential decisions, like which profession to pursue. Still, the very possibility even of such trifles as a finger flexion being 'done to me by my brain' seems worrying. But there are further reasons why the anti-IDS interpretation is flawed. A second one is the conflation of 'urge' and 'intention' (see Von Wachter 2007, 2016). The subjects reported of the former, not of the latter; and the two are clearly distinct, at least if one understands 'intention' as the mental event that already constitutes the onset of the action, not just an inclination to do the action. I can have an urge to scratch myself without forming an (effective) intention to do so (for example because I know that scratching a chickenpox pustule makes things worse), and so Libet's subjects could have their urge caused by the RP without the effective intention being so caused. Libet, wary of the anti-freedom interpretation of his results, even made further experiments to test this latter hypothesis. After all, if the intention is fully under the control of the subject, then they can refrain from performing the action even after RP onset. Indeed, Libet found evidence that this is so (Libet 1985) and called this power of the human mind a 'veto'.

All in all, the Libet experiments do not yield evidence against IDS, because their results can neatly be explained within an interactive dualist framework. I conclude that by PoC, the commonsense belief that it is us who mentally cause actions and thoughts is still intact.

II. Philosophical Objections against Interactive Dualism

2. The CCP Objection

This part deals with two major objections against interactive dualism. One is that it is impossible for something mental to causally affect something physical, given the radically different natures of the mental and the physical; the other is that a causal influence of the mental on the physical world is excluded because the physical realm is causally closed. I will deal with the former objection in the next chapter; in this chapter, I will address the latter.

The doctrine that the physical realm is causally closed figures under various names in the literature. I will here use the abbreviation with which I am most familiar, namely CCP (Causal Closure of the Physical). CCP constitutes undoubtedly the most widespread, and, in the eyes of physicalists, the strongest objection to interactive dualism, which can be seen from the following quotes:

> Over the latter half of the last century English-speaking philosophy became increasingly committed to naturalistic doctrines. Much of this naturalistic turn can be attributed to the widespread acceptance of the thesis that the physical realm is causally closed. (Papineau 2009, 53)

> [I]t was the empirical evidence for causal closure that persuaded philosophers to be physicalists. Once mid-century physiological research had established that all physical effects had physical causes, even in bodies and brains, philosophers quickly figured out that general physicalism followed. (Montero and Papineau 2016, 188)

2.1 Specifying the doctrine of the causal closure of the physical

A useful expression of CCP is given by Jaegwon Kim:

> If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical. (Kim 1998, 40)

One often also finds the shorter version that "Every physical event that has a cause, has a physical cause". However, CCP stands in need of further specification to do the work physicalists want it to do: the physical causes in question must be *sufficient* causes, causes whose sole presence suffices for the effect to occur. Thus David Papineau:

> Every physical effect has a sufficient physical cause. (Papineau 2009, 53)

Without this clause, dualists could well agree to the above formulations and add that some physical causes are just *necessary* conditions (causes that *need* to be present but whose sole presence does not suffice for the effect to happen) for their effects and so need to be complemented by mental causes. One might still argue that even the causal sufficiency addition is not enough, though. Couldn't a physical and a mental cause both be sufficient conditions for a physical effect to happen? This is the idea of causal overdetermination. In fact, there are authors who defend such a view (e.g. (Mills 1996)). I will, however, go along with the majority of philosophers who reject it and thus grant physicalism the full range of premises it needs to run its argument against interactive dualism (the rejection of causal overdetermination is one of them, as we will see shortly).

2.2 Arguments for CCP and Replies

Until now CCP has been presented merely as a metaphysical assertion. What reasons are there to believe it? I shall investigate what I believe to be the four strongest ones and reply to each in turn.

2.2.1 The argument from conservative basic forces

The first argument is an inductive argument. It starts from the assumption that all 'macroscopic' forces (those whose effects we directly observe; in effect compound forces) physics has dealt with so far can be reduced to conservative basic forces (and are thus conservative themselves). Conservative forces are forces that conserve both energy and momentum. From this, so the argument goes, one can inductively generalize to all other macroscopic forces in nature to the effect that all those forces are also constituted by the conservative basic forces and are therefore conservative. The first to extend this principle of conservative forces to all areas of physics including physiology was Herrmann von Helmholtz (1821-1894)[13]. While many of Helmholtz's contemporaries still believed in 'vital forces' specific to living beings and not (necessarily) conservative, the reducibility of macroscopic to conservative forces is nowadays almost unanimously accepted at least in physiology[14].

Now, is the reducibility to conservative forces really decisive for ruling out mental interaction and thus establishing CCP? No, it isn't. It were only so if mental interactions were necessarily non-conservative. David Papineau himself, premier defender of CCP, admits that "there is no outright inconsistency between the conservation of energy and vital forces, and many late nineteenth-century figures were quite explicit, not to say enthusiastic, about accepting both." (Papineau 2001, 25). One possibility Papineau grants how mental forces could act on the brain is if they are "governed by deterministic force laws" (ibd.). What he means by that is that those forces compensate losses in kinetic energy by buildups in potential energy, and vice versa – in other words, that they are conservative. Indeed, such law-governed mental interactions are being considered by Robin Collins (Collins 2011a). I will deal with Collins's and other dualistic approaches to making mental interaction compatible with physical law in more detail in chapter 5; for now it suffices to state that conservation laws are in principle consistent with non-physical forces acting on physical matter, as long as those forces are law-governed and not spontaneous, i.e. not describable by any natural law.

There is, however, a yet deeper point. What Papineau does not seem to be aware of is the fact that physical theory can accommodate even spontaneous forces. The key to understanding this is the Noether theorem (Noether 1918, see also Cucu and Pitts 2019). The Noether theorem will play a central role in the defense of interactive dualism against the objection from conservation principles (see chapters 4 and 7, where the Noether theorem will be treated in more detail). Roughly, the Noether theorem establishes a biconditionality between certain symmetries in a physical system and the conservation of energy or momentum:

Symmetry ≡ conservation

---

[13] See (Papineau 2001, 22ff.); (Coleman and Coleman 1977, 150-154).

[14] Though not necessarily in physics generally. There are some issues in General Relativity and Cosmology that cast doubt on whether energy is always conserved, quite apart from all non-physical interactions (Pitts 2004; 2019).

In other words, if the symmetries hold, then conservation holds, and vice versa; conversely, in keeping with the truth table of biconditionals, if either the left or the right hand side is false, then the other side of the biconditional must also be false (i.e., if the symmetry doesn't hold, then conservation doesn't hold). One can show via relatively simple mathematics (Cucu and Pitts 2019) that any sort of spontaneous interaction entails non-conservation and thus a break of symmetry or a break of symmetry and thus non-conservation. A concrete case study to this effect is carried out in section 5.2.

All in all, the *a priori* argument from conservative forces fails to establish CCP, because it cannot provide a conceptual exclusion of *sui generis* mental causes; such a conclusion could only be reached either if mental causes of physical effects do not, for an independent reason, exist *in any possible world* (see the 'causal heterogeneity objection' in chapter 3, which of course renders the reference to conservative forces otiose) or if they do not exist *in our actual* world, which requires empirical investigation. In fact, Papineau's second main argument pertains to the latter point. It is to this that we will turn presently. Before that, however, I'd like to reply to an argument closely related to the one we've just discussed.

2.2.2 The argument from the in-principle completeness of physics

Suppose we knew all the basic constituents and forces of the universe and how they interact. Would this enable us, at least in principle, to give a full, sufficient, causal explanation of all physical phenomena, as required by CCP? Physicalists typically believe it does. The doctrine involved is often called the 'completeness of physics'. Since however, as everyone agrees, present physics is incomplete, we better speak of the 'in-principle completeness of physics' (IPCP), understood as

> …the possibility of a complete and comprehensive physical theory of all physical phenomena. (Kim 1998, 40)

Now, does IPCP entitle us to conclude CCP? Only under one condition, namely that the explanations given in IPCP (or, rather, the causes specified by those explanations) are both sufficient *and* necessary. Of course, they need to be sufficient, for otherwise they would always have to be complemented by non-physical causes; but they also must be necessary, for otherwise non-physical causes could be alternative sufficient (albeit non-necessary) explanations for physical effects. An immediate problem with this is that the necessity condition is just a different rendering of the metaphysical posit that mental causes are *a priori* ruled out. On what basis could this be assumed? There seems to be none. Even if physics were complete such that it has an explanation ready for every physical effect, it would not follow that none of these effects could ever be brought about by a non-physical cause (see Von Wachter 2019). We could close the case at this point; however, I wish to demonstrate the utter failure of IPCP to establish CCP. Let's therefore grant, for the sake of argument, that IPCP entails CCP.

Even with this concession there are severe problems. First, we do not even know whether a complete physics is even possible; maybe such a hope is illusory. But even if it is possible, the main difficulty persists: namely that we do not know what the proper extension of 'the physical' is. We obviously cannot equate it with everything that figures in present physical theories, for present physics is incomplete; nor can we identify it with some future physics, because we cannot rule out the possibility that some future (complete) physics will include mental properties and laws. In the latter case, mental properties and laws will be re-christened as physical, and CCP reduces to the uncontroversial thesis that everything is causally connected to other things. In principle, it might even incorporate immaterial beings like souls, angels and God[15]! In that case, physicalism (as it is propagated today) will have been refuted by physics, or else trivialized by simply absorbing all the entities formerly considered as problematic.

Papineau seeks to avoid this result by defining 'physics' purely in terms of completeness, more precisely as "the science of whatever properties are needed for a complete set of laws covering such effects as stones falling, darts hitting boards etc." (Papineau 1990, 70). Let's call this construal of physics PHYSICS to distinguish it from the contemporary science (henceforth 'physics'). On PHYSICS, CCP is analytically true. However, PHYSICS does not rule out mental properties and laws from being PHYSICAL (not even if they emerge from a physical basis), as long as they contribute to the explanation of physical events. For example, if for a complete explanation of my mowing the lawn the complex desires, beliefs and intentions I had are indispensable (desire: leave nature to itself; belief: my wife is not pleased with the look of a wild meadow; intention: peace at home), then it follows that those mental events are PHYSICAL. And we are back with the original problem of triviality.

Papineau replies that it is unlikely that any mental categories will feature in PHYSICS:

> It is one thing to hold that the categories of current physics are going to be superseded. It is quite another to hold that they are going to be superseded (*inter alia*) by mental categories. (ibd.)

It is unlikely indeed that what we now understand as physical will one day be *superseded* by something mental. However, in future PHYSICS, nothing of this sort needs to happen. It could simply be that mental properties and laws *come alongside* properties and laws we uncontroversially view as physical now. Thus, the dilemma persists: physics does not entail CCP because it is incomplete; and PHYSICS renders it trivial.

Tim Crane (Crane 1991, 35-36) suggests that the physicalist should opt for a third way. He calls this third option 'ideal-physics':

> The idea is this: take all the phenomena about whose *physical* status there is no dispute. Suppose there to be a theory, resembling current physics, that gives all the laws and properties needed to account for all these phenomena. Then the claim that this theory is complete is the claim that its laws and properties suffice to account for, or 'fix', *all other* phenomena, including mental and behavioural phenomena. (ibd., 36)

Ideal-physics, finally, does entail CCP in a non-trivial way. However, it is essentially a supervenience claim, as Crane himself notes, and so stands and falls with supervenience. It should also be noted that ideal-physic*s* does not entail (reductive) physicalism. Rather, it is compatible with supervenient property dualism. Nicholas Maxwell, in defending a mere explanatory dualism, nicely sums it up:

> [T]he fact that physics is silent about such features . . . provides no grounds whatsoever for holding that such features don't exist, or are inherently unintelligible if they do exist. (Maxwell 2000, 51)

Thus, admittedly, the defender of CCP may be able to find a definition of the physical that renders CCP non-trivial. But that solution depends on the controversial supervenience relation, and dualism cannot be ruled out that way. Thus, IPCP fails to be a persuasive argument for believing in CCP rather than in a physical world open to interaction.

[15] Qualification: only to the extent to which these immaterial persons are subject to laws/regularities.

2.2.3 The empirical argument for CCP

The empirical argument for CCP basically consists in the claim that there is no empirical evidence for mental interactions. As such, it is clearly distinct and independent from the previous two arguments which heavily rely on *a priori* assumptions about the structure of the physical universe, and in fact more interesting: for even if non-physical causes were in principle possible, their empirically confirmed absence would strongly speak against interactive dualism. However, Papineau reminds us that the empirical argument must be seen as operating against the background of the *a priori* argument from fundamental forces:

> The earlier argument suggested that at least most natural phenomena, if not all, can be explained by a few fundamental physical forces. This focused the issue of what kind of evidence would demonstrate the existence of extra mental or vital forces. For once we know which other forces exist, then we will know which anomalous accelerations would indicate the presence of special mental or vital forces. Against this background, the argument from physiology is then simply that detailed modern research has failed to uncover any such anomalous physical processes. (2001, 31)

In summary, once one knows the operations of hitherto known forces, then anomalies in systems (like brains) where only the familiar forces are expected to be at work show that some other, mental forces must be operating; conversely, failure to detect such anomalies implies the absence of such forces.

The first important inquiry to be made here is what kinds of empirical investigations have been conducted to ascertain the absence of anomalous forces. Papineau remains astonishingly vague in this respect. The only concrete instances of pertinent experiments he cites are from the 19th century (referencing Coleman and Coleman 1971), and only one of them figures specifically in his paper (Max Rubner's respiration calorimeter experiments from 1889). But these experiments, as Papineau himself admits, "failed to provide compelling evidence against vital or mental forces" (2001, 31). Consequently, he locates the main driving force for the final departure from non-physical forces in the rapid advances made in molecular biology from the 1950s onward, summing them up in a rather sweeping way:

> During the first half of the century the catalytic role and protein constitution of enzymes were recognized, basic biochemical cycles were identified, and the structure of proteins analysed, culminating in the discovery of DNA. In the same period, neurophysiological research mapped the body's neuronal network and analysed the electrical mechanisms responsible for neuronal activity. Together, these developments made it difficult to go on maintaining that special forces operate inside living bodies. If there were such forces, they could be expected to display some manifestation of their presence. But detailed physiological investigation failed to uncover evidence of anything except familiar physical forces. (ibd.)

Robb and Heil (Robb and Heil 2019) concur:

> Were *Completeness* false owing to the routine intervention of souls, physiology would have discovered anomalies within the human body, events that have no physical explanation.

As compelling as this may sound *prima facie*, it would certainly be useful to be presented with at least one or two exemplary instances of such revealing experiments. After all, molecular biologists are not normally concerned with measuring forces in and between tiny molecules; this is rather the domain of biophysicists and in part of physiologists (e.g. when exploring the properties of the cell membranes of neurons). But since a good deal of microphysical work with respect to cells has been done, and done systematically and over a long period of time, we can grant Papineau's observation. The question is just whether it really helps him establish CCP.

A first critical question in order here is this: does the research Papineau refers to really cover the relevant ground? Jeffrey Koperski thinks it does not:

> When it comes to CCoP, the truth is that dualism and all the other nonphysicalist programs were never refuted as much as they were ignored (...) The physiological research that Papineau touts is all about the nonmental, either uncovering detailed mechanisms or understanding how one subsystem is integrated into a whole. Any question about the mind is carefully bracketed to the side, virtually guaranteeing that mental causes will not be detected (...) Perhaps medical research should remain methodologically agnostic about the mental. But if scientists have tacitly agreed to not investigate that question, no amount of scientific progress would count as evidence against an irreducible mental realm. To claim that the matter has been resolved by the evidence begs the question. (Koperski 2020,121-122)

I believe Koperski is right. It is one thing to give a full explanation of brain processes modulo the volition of the human agent; it is another to give a full physical explanation of volitional actions. It does not suffice to explain away mental causes in muscle contraction, insulin shedding or heart regulation; one must also explain them away in the neuronal firings that arguably initiate volitional actions. So, has a full physical explanation of volitional actions been given? Many believe the Libet experiments have done exactly that. (By the way, it is curious that Papineau doesn't mention them, although they were published more than one and a half decade prior to his paper. Did he not consider the evidence unambiguous enough after all?). I argued in section 1.4 that the Libet experiments fail to necessitate the non-interactionist conclusion. Moreover, there are voices more eminent than mine – not all of them dualist – who call for a more cautious interpretation of the Libet experiments (see the references in section 1.4), one where mental causes might still find their distinctive role.

There are yet stronger reasons to disbelieve Papineau's claim. One is the phenomenon of neuroplasticity. Neuroplasticity means that our brains are susceptible to reorganization and rebuilding of neurons *even in adulthood* and *due to conscious mental efforts*. There is a significant literature on the topic; I shall list some of the empirical findings in chapter 9.

I'd like to even go one step further by showing that we have strong reasons to believe that volitional actions *cannot* take place without mental causes. Here I can just sketch my reply; a full treatment will occupy the whole of chapter 9. The first thing to note is that, *contra* Papineau, we do not have sufficiently detailed knowledge of brain (micro-)processes to exclude (or to confirm) anomalies putatively caused by mental interaction. So, all one can do presently is to focus on what we know about molecular processes in the brain and ask: is a complete physical explanation of volitional action even *possible,* given these basic mechanisms? (An approach Papineau, advocate of force reducibility, should like). This may sound like an odd question, given the completely physical makeup of the brain, but if it is a silly question, then a quick and definite answer it should (and could) receive. Surprisingly, it is precisely a study of the brain processes involved in volitional action, and in particular of their molecular goings-on, that seriously raises the question whether a wholly physical account of volitional actions is even *possible* or at least *plausible*. Again, this conclusion will be expounded in detail in chapter 9.

Probably, detractors of interaction will still feel uneasy, even if my above arguments are regarded as having some purchase, arguably for some meta-philosophical reason: what of the regularity or even stability of nature, if interactions are allowed which are not themselves governed by the laws of nature? This is a point well-taken, which I shall consider at length in part IV of this thesis.

Let's go back again one step and grant that the findings of physiology suggest that there are no non-physical forces at work. Even in this case, several replies in the defense of interactive dualism are possible.
One first rejoinder is mentioned by Papineau himself:

> It may still be that such [mental] forces are activated inside cells, but operate in such a way as to "pay back" all the energy they "borrow", and vice versa. (Papineau 2001, 31)

Given that, at least in the context of cells, we can only ever measure energy changes summarily, as opposed to individually for every molecule (or atom? Where would we stop anyway?), this proposal is initially plausible. It has one downside, however. In order to uphold energy conservation, the mind would have to have physical properties to the extent that one can meaningfully assign energetical states to it. Again, this idea is most seriously elaborated on by Collins (2011). However, Collins's theory comes at a steep price, and I think that price for a dubious gain in physical credibility much too high. In section 5.4 I will explain why.
Let's turn to another option, one not considered by Papineau (although he is aware of the physical basis for it). A non-physical mind could interfere with the brain by changing the trajectories of particles without changing their energy content, such that no anomalies in terms of energy expenditure could be detected. This proposal comes in at least two variants, one treating the particles as classical physical objects, and the other treating them as quantum objects. In the former case, one pays the preservation of energy conservation with a non-conservation of linear momentum. This is due to the change in direction unaccounted for by anything physical. Of course, one could take up the approach sketched above of ascribing physical properties to the mind (here, the property of having a linear momentum) – to which the same objection applies. But as I have insinuated, such a move is simply unnecessary, for violations of momentum conservation are conceptually integrated by the Noether theorem and thus become a matter of conditionality.
The quantum version of this hypothesis consists in the idea that the non-physical mind can exploit quantum indeterminacies to bring about changes without violating the conservation of either energy or momentum (the most prominent example for such a proposal being (Beck and Eccles 1992); see also (Beck 1996; 2008)). This approach faces several problems which I am going to treat in detail in chapter 6.
Lastly, a third response is to simply say that although failures of energy and/or momentum conservation occur (due to mental 'forces'), they are so gentle as to be undetectable (see for example (Pitts 2019a)). This is in fact more than a cheap *ad hoc* move. As a number of authors suggest (Beck and Eccles 1992; Beck 2008; Georgiev and Glazebrook 2014) very minute changes in the microphysical properties of brain particles may suffice to induce major changes (but see (Wilson 1999) for a dissenting voice). Whether or not these suggestions work is another question, one I shall try to answer in chapter 6. Until further notice, the 'gentle failure' approach is a viable option.

In summary, the empirical argument is the spearhead of physicalism, but it is far from clear that it can strike a fatal blow. It is physicalism's spearhead because, if indeed empirical science managed to explain *all* relevant brain processes without remainder, interactive dualism would be forced to retreat to approaches all of which have their grave problems (e.g. causal overdetermination or quantum-mechanical views, see chapters 6 and 8), or to surrender altogether. However, a closer look at empirical neuroscience – one elaborated in chapter 9 – suggests that Papineau's confidence is overstated, to put it mildly. At the very minimum, with the current state of knowledge, no knock-down argument against interactive dualism arises.

2.2.4 The argument from scientific realism

The last argument for CCP I'll be treating is the argument from scientific realism. Scientific realism is, roughly, a "positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences" (Chakravartty 2017). I take this to imply that at least the entities contained in scientific theories (especially theories of fundamental physics) should be believed in, i.e. considered as existent. This is in line with the way Michael Esfeld (whose argument for CCP I am going to critique) expresses scientific realism:

> [P]hysical theories as they stand tell the truth about the physical realm (or theories in the same vein as our current ones do so). (Esfeld 2022, 263)

A slightly different angle on the issue (relevant here and shedding an interesting light on scientific realism) is taken by the slogan-like expression "Scientific realism is the description of the universe from a point of nowhere and nowhen". Since such a description is independent of the perspective of any mental being, it entails the mind-independent existence of the elements figuring in scientific theories.
Proceeding from this basic assumption (which I share), the argument simply states that CCP follows from scientific realism. Thus Michael Esfeld (using the abbreviation PCC instead of CCP):

> What, then, is the argument for PCC? The argument is that if the physical laws are true, PCC is true as well. In other words, the argument is that scientific realism with respect to physical laws is committed to PCC. For this argument, it is irrelevant what the physical laws are—that is, whether they are the laws of classical mechanics, of the standard model of elementary particles in quantum field theory, of a future theory of quantum gravity, etc. Furthermore, it is irrelevant whether they are deterministic or probabilistic. The argument is that taking the laws of our physical theories (classical, relativistic, quantum) as guide to the true laws of nature, anyone who is a scientific realist is committed to PCC. (Esfeld 2019, 170)

This instantly strikes me as curious. I, for one, consider myself a scientific realist but no adherent of CCP, and I fail to see the logical tension between these two positions. Perhaps some more insight into the notion of 'law' Esfeld has in mind could help appreciate his tenet more. The following two citations reveal more about the conception of 'law' he uses:

> Laws are formulated in terms of differential equations such that the variables figuring on the right hand side of the equation indicate what determines the temporal evolution of the objects as they figure on the left hand side of the equation. (ibd., 171)

> What is relevant is that one cannot add to this structure of physical theories formulated in terms of differential equations that indicate what determines the temporal evolution of the objects under consideration a clause to the effect that something may intervene from the outside that stops the evolution of the objects as indicated by what figures on the right hand side of the differential equation in question. (ibd., 172)

So, for Esfeld, a law of physics can be expressed as a differential equation whose structure reflects the actual goings-on in nature: the temporal evolution of the objects on the left-hand side of the equation are determined by the variables and their mathematical relations on the right-hand side (determination here meaning 'logical entailment', as will become clear shortly). As far as it goes, this is entirely plausible, but without further metaphysical assumptions, nothing like CCP follows from this. Esfeld's claim that "one cannot add to this structure…a clause to the effect that something may intervene from the outside" apparently is such an additional assumption, which, however, needs argumentative support. His main argument in its favor is that (fundamental) laws of physics apply to the universe as a whole such that, if something (e.g., God), is to intervene, it must intervene at the level of the entire universe, and cannot interact locally (Esfeld, personal communication).

For the sake of a better overview of the critical points in Esfeld's argument, let's try to reconstruct its logical structure:

P1 The entities, quantities and relations figuring in laws of nature understood as differential equations exist mind-independently (scientific realism).

P2 If the entities, quantities and relations figuring in laws of nature understood as differential equations exist mind-independently, then the right-hand side of the equations indicates what determines the temporal evolution of the physical objects represented by the left-hand side of the equations.

P3 The right-hand side of the equations indicates what determines the temporal evolution of the physical objects represented by the left-hand side of the equations. (P1, P2, *modus ponens*)

P4 The differential equations (i.e., the laws) apply to the universe as a whole.

P5 No mental entities, quantities or relations figure in the differential equations of any known law of nature.

C Therefore, scientific realism implies that no mental cause influences the behavior of physical matter (CCP). (P3-P5)

Although Esfeld doesn't mention it explicitly (although he accepted it in personal conversation), I added P5 for the sake of completeness. P1 is accepted *ex hypothesi*, P3 follows validly from P1 and P2 and P5 is uncontroversially true. That leaves P2 and P4 as targets.

Let's start with the easier one, P4. I see three objections. First, it is not clear that one must equate laws of nature with differential equations. Some philosophers think otherwise (see e.g. (Koperski 2017; 2020; Von Wachter 2015)). I shall come back to this point in chapters 12 and 14. Second, even differential equations yield an outcome only if fed with initial conditions. True, the initial conditions of one application of a differential equation were the outcome of a previous one; but ultimately, at least the initial conditions of the universe cannot be specified by any law of nature, whether in the shape of a differential equation or not, as Big Bang theorists well know. Third, it is by no means clear that the differential equations "apply to the universe as a whole". To be more precise, the truth of this statement depends on its exact meaning. If Esfeld means that the equations hold at any given space-time point, then this seems relatively uncontroversial; however, an equation describing the behavior of the whole universe is something far more contentious. What Esfeld indeed has in mind here are differential, dynamical equations, applicable in principle to the whole universe (Esfeld, personal communication). One extreme form of such an equation would be a Schrödinger or Bohmian equation comprising all the fundamental particles in the universe. But we need not now resolve the question whether such a picture is viable; it can be granted that laws, understood in a 'structural' rather than dynamical sense hold for the whole universe, at least in a somewhat abstract way. Such laws, however, have the drawback (from Esfeld's perspective) that they do not yield any concrete histories of the universe without initial conditions, and, possibly, being translated into differential equations. So, P4 might even be granted with some qualifications, and yet end up as a 'toothless tiger'.

Now, Esfeld gets his result via the auxiliary non-intervention clause. But why should we accept it? Could it not be the case that although the equation describes the evolution of the physical universe, there are non-physical persons (or God), who could at any moment intervene at will? Esfeld is right that no persons or mental factors figure in physical equations. Might this not be precisely because the equations yield the evolution of physical systems only in case no such person interacts? As long as we cannot definitively rule out such a possibility, I see two options to do justice to both the interactionist as well as the scientific realist: either the universal equation will have to be complemented by a non-intervention clause, or one considers laws of nature as something with universal validity yet with no unqualified predictive power regarding temporal evolution.

Let's turn to P2. Its truth depends crucially on what is understood by 'determines'. Esfeld makes his view clear:

> [T]he propositions stating the laws of nature and the propositions describing the state of the world at an arbitrary time (i.e. the propositions describing the initial conditions) entail the propositions describing the state of the world at any other time. Hence, determinism in physics is—only— about entailment relations among propositions. There is nothing here that suggests that what there is in the state of the world at a given time literally produces or brings about the evolution towards the future. The question for a scientific realist then is what it is in the world that makes these propositions true, that is, in virtue of what in the ontology these entailment relations among propositions hold. (Esfeld 2019., 173)

Esfeld thus explicitly rejects any sort of causal determination. Instead, he views the relations between physical states as 'entailments', a term reminiscent of logical relations. On the bright side, the view is in line with actual physical practice: physicists seldom if ever speak of causality at all (cf. Russell's famous quote in (Russell 1912)). The gnawing question is just, how can anything like CCP be derived from such a view of determination? Esfeld rightly diagnoses the crucial question to be "in virtue of what in the ontology these entailment relations among propositions hold." What indeed? On a view that rejects causal production, only the factual motion of matter can serve as a truthmaker for those relations (as Humeans in general, and Esfeld as a Super-Humeanist, hold). But this leaves open the question in virtue of what those motions take place, especially regarding future motions (that are needed for a genuine entailment relation). Esfeld, as all good Humeans, will refuse to give any explanation, with reference to the basic assumptions of Humeanism. This strikes many, including me, as unsatisfactory, and it raises the question what justifies belief in a continued stability of the laws of nature in the future. After all, since matter just moved around in certain ways in the past, without any deeper explanation for those patterns, why can it not suddenly alter its patterns of motion in the future? There is nothing that could prevent such a global alteration, or, for that matter, local glitches (I shall come back to these problems in chapter 13). The fact that logical entailments are necessarily true does not salvage the 'Humean entailment'; the mathematical entailments involved do not bring about the physical states in any way, which by the way leaves us with the 'unreasonable effectiveness of mathematics for physics' (cf.(Wigner 1967b)) and still lacking a reason to believe that the physical dynamics must follow the mathematical structure.

One might reply that on stronger views of determination, P2 may come out true. Such a more robust metaphysics can be had on virtually all other views of the laws of nature (see chapter 13). But even if this is so, P4 fails to be compelling, and so the whole argument breaks down. Further, a stronger view of determination may cause problems of its own. To be stronger than Humean entailment, such a view will be committed to viewing determination as a sort of *causal production* or at least *causal constraint*, something Humeanism anathemizes and physics per se does not require. This is not decisive, however, for, on the other hand, nothing in science precludes ontologically robust causation. However, causal production *tout court* does not help establish CCP, because it still allows for mental causes, for example via the ceteris paribus clause that the laws describe the temporal evolution as long as no interaction occurs. This conception of determination is what Von Wachter calls 'weak determinism' (Von Wachter 2009b, 188). As Von Wachter puts it elsewhere:

> A [weakly] deterministic process is one that can stop only if something stops it. (2015, 44)

To my mind, this conception of determination gives all their due: to science, a grounding for laws of nature; to mental interaction and free will, the possibility to interact. However, it also pulls the carpet from under CCP and physicalism, because it clearly admits of non-physical causes.

What could give the physicalist what he needs is the introduction of a strong modal element; for example, he might hold that the causal processes in question happen with necessity either in the actual world or across nomologically or even metaphysically possible worlds. However, as Daniel von Wachter has shown (Von Wachter 2009b; 2012; 2019), such a strong modal determinism is nothing more than a metaphysical posit lacking empirical as well as philosophical support. Also, physicalists do in general not rely on strong modal determinism (Esfeld's (2019) criticism of Von Wachter 2019).

I am aware that there is still much to be said, in particular regarding the different conceptions of laws of nature as they turn on interactive dualism. The detailed discussion of these issues is reserved for chapters 12 and 13. For now, I conclude that the argument from scientific realism falls short of compelling support for CCP. Robert Bishop nicely sums up the gist of the arguments *contra* CCP given in this section:

> Physics does not imply its own closure. . . . Rather, CoP is a metaphysical doctrine. Indeed, physics tells us what happens when particular forces are taken into account, but nothing about what happens when influences unaccounted for by physics are present. (Bishop 2011, 606)

One can coherently be an interactive dualist and committed to scientific realism.

2.3 Why belief in CCP cannot be justified

So far, we've seen that the arguments in favor of CCP fail to be compelling (or, in the case of the argument from IPCP, merely establish its non-trivial conceptual coherence). In the absence of counterevidence, though, CCP still *may* be true. As indicated above, I will present counterevidence from neurophysiology in chapter 9. In the remainder of the present chapter, the argument against CCP will be another: namely that CCP is 'unbelievable', that is, that one cannot be justified in believing CCP.

There is a strong *a priori* bias against CCP from a common-sense point of view, because mental events are consistently cited in the explanation of human actions. That stands in stark contradiction to CCP, which denies the causal efficacy of the mental. For example, Lynn Rudder Baker, herself no Cartesian dualist, argues that if CCP threatens to undermine our ordinary (and scientific) explanatory practices – many of which cite the mental – it's CCP that has to go (Baker 1993). Koperski argues in the same vein:

> [L]et's not pretend that exhaustive knowledge of the interplay of nerves, muscles, and joints will ever be able to explain why my arm is moving. That event involves the mental: I *want* my coffee and *intend* to grasp the mug. (2020, 121-22; italics in original)

However, physicalist functionalists can very easily avoid having to pick their poison, because for them mental events are just physical events, and so their causal efficacy is a matter of course.

The following is argument is therefore directed at a different type of non-interactionism, namely dualist epiphenomenalism. Epiphenomenalism is exactly the sort of middle ground required: it does not beg the question against physicalism because it rejects interaction; and it does not beg the question against interactive dualism either because it considers mental properties as ontologically distinct from physical properties. In presenting the following argument I rely heavily on Richard Swinburne's version of it (Swinburne 2013, ch. 4; 2019).

The starting point is to treat CCP as an *empirical* doctrine that is entailed by some scientific theory. (The alternative is, of course, to consider it a *metaphysical* doctrine, but we've already seen that even scientifically informed versions of such *a priori* assumptions are unpersuasive, see section 2.2.1). Now, according to Swinburne (2019), one can only be justified in believing a scientific theory if

(i) one is justified in believing that it *predicts* certain events and
(ii) one is justified in believing that these events *occur*

where 'being justified in believing X' means, via the Principle of Credulity (see section 1.4), that it seems to one that X and that there is no counterevidence to X. Conditions (i) and (ii) can be summed up as 'being justified that a scientific theory *successfully* predicts certain events'. I take this to uncontroversially reflect the demands made to justified belief in a scientific theory: (i) alone is compatible with the theory simply being false; if a theory predicts certain events, but these events do not occur under the relevant circumstances, then one is not justified in believing in the theory after all. On the other hand, if (ii) were the only criterion met, there would not be any theory to begin with. Both conditions *together* correspond to the way scientific theories are in fact accepted. For example, the theory of semi-conservative DNA replication (Meselson and Stahl 1958) can be justifiably believed because it predicts that ^{15}N-DNA replicated with ^{14}N nucleotides consistently generates a band midway between ^{15}N- and ^{14}N-DNA after centrifugation; and this is exactly what happens when one carries out the experiment.

The problem with CCP is that it rules out justified belief in any scientific theory that entails it. (As an aside, since the following arguments apply to scientific theories in general, it even rules out justified belief in *any scientific theory* whatsoever). To see why, we need to consider how epistemic justification along the lines of (i) and (ii) can come about.

Let's start with (i). One can be justified in believing that a scientific theory predicts certain events either by direct experience, memory or testimony (or, as Swinburne puts it, 'apparent' experience, 'apparent' memory or 'apparent' testimony, to allow for the possibility of error). Apparent direct *experience* justifies the belief that a scientific theory predicts certain events because the scientist 'sees' that the mathematics or the logical structure of the theory entail the events in question. But that is (fortunately) not the only way one can be justified in believing such predictions, for otherwise justification would be available only for those who know a theory first-hand, and who do not have to rely on memory or testimony at all to see that a theory makes certain predictions – in other words, they would have to hold the whole theory and its entailments at once in their minds, something well-nigh impossible. The scientist (or any other person) is also justified in believing the predictions if they apparently *remember* the theory making them; and so are they when they receive an apparent *testimony* that the theory makes the predictions by another person who is justified in believing the predictions via memory or direct experience. A fourth, though not independent, way of justification is logical consequence: if a theory T1 is logically entailed by another theory T2 which is believed justifiably, then one is justified in believing T1; but this of course presupposes being justified in believing T2.

It turns out that justification of (ii) occurs in the same ways as in the case of (i). To be justified in believing that the predicted events occurred, I either need to have an apparent direct (sense) experience of them, or the apparent memory of such an experience, or the apparent testimony of a person who either had direct experience of them or apparently remembers them.

A problem with justification arises in all cases where *prima facie* mental-to-physical causation (MP) – which is forbidden by CCP – is required. This is true of memory and testimony, but not of direct experience. Let me explain. For a better overview, recall that there are two conditions to be met for justified belief in a scientific theory (prediction and occurrence of certain events), and three epistemic ways mentioned above in which the conditions can be met, which makes six cases to be investigated. Table 3.1 displays those six cases in an orderly manner.

Condition to be fulfilled → Justification pathway ↓	Prediction	Occurrence
Direct experience	$P_1 \rightarrow M_1 \rightarrow M_4$ $M_3 \rightarrow M_4$	$P_1 \rightarrow M_1$
Memory	$P_1 \rightarrow M_1 \rightarrow M_4 \rightarrow M_2$ or $M_3 \rightarrow M_4 \rightarrow M_2$ (for very recent events) $P_1 \rightarrow M_1 \rightarrow P_2 \rightarrow M_2 (\rightarrow M_4)$ or $M_3 \rightarrow M_4 \rightarrow P_2 \rightarrow M_2$ (for temporally more distant events) Reconstructed version for temporally more distant events: $P_1 \rightarrow P_2 \rightarrow M_2 (\rightarrow M_4)$ (wildly implausible!) or $M_3 \rightarrow M_4 \rightarrow M_2$ (implausible)	$P_1 \rightarrow M_1 \rightarrow M_4 \rightarrow M_2$ (for very recent events) $P_1 \rightarrow M_1 \rightarrow P_2 \rightarrow M_2 (\rightarrow M_4)$ (for temporally more distant events) Reconstructed version for temporally more distant events: $P_1 \rightarrow P_2 \rightarrow M_2 (\rightarrow M_4)$ (plausible)

Testimony	$P_1 \rightarrow M_1 \rightarrow P_2 \rightarrow M_2 + M_5 \rightarrow P_3$ or $M_3 \rightarrow M_4 \rightarrow P_2 \rightarrow M_2 + M_5 \rightarrow P_3$ Reconstructed version for temporally more distant events: $P_1 \rightarrow P_2 \rightarrow P_3$ (wildly implausible!) No possibility to avoid MP causation in case of an initial M-event	$P_1 \rightarrow M_1 \rightarrow P_2 \rightarrow M_2 + M_5 \rightarrow P_3$ Reconstructed version for temporally more distant events: $P_1 \rightarrow P_2 \rightarrow P_3$ (wildly implausible!)

Table 3.1: Causal pathways of the two criteria for justified belief in a scientific theory. P_1 = external physical event; P_2 = memory trace in the brain; P_3 = communication (testimony) of memory content; M_1 = direct experience of external physical event; M_2 = memory experience (experience of past experience M_1); M_3 = grasping of scientific theory; M_4 = grasping of logical entailment of the theory; M_5 = intention to testify. Unproblematic cases in green, problematic ones in red.

I begin with direct experience. Experience justifies belief in the occurrence of some physical phenomena in the following way: a physical stimulus P_1 (emitted by a physical object) impinges on our senses and thus causes the mental event M_1 of the experience of the physical object. For example, the light emitted by a cell under the microscope causes in me the direct sense experience of the cell; in the absence of counterevidence (e.g., that someone has built a miniature video screen in the microscope's light path) I am justified in believing that I see a cell. Direct experience can also justify the belief that a theory makes certain predictions. This may involve a physical event P_1 (for example, seeing the mathematical formula representing a theory written on paper or a screen), which causes a direct experience M_1 of P_1; M_1 then in turn causes a direct grasp of the logical entailment of the theory (M_4). It may, however, go without any physical event whatsoever. Upon thinking about a metaphysical theory (M_3), for example, I may have the direct experience of grasping a logical entailment of the theory, and in the absence of counterevidence (e.g., that I overlooked something) I am justified in believing that the proposition in question does follow from the theory. Hence, all that is involved in direct experience is physical-to-mental causation (PM) or mental-to-mental causation (MM), none of which is prohibited by CCP.

Let's turn to memory. Based on the common distinction between short-term and long-term memory, we might here distinguish between very recent events (less than about five minutes ago) and more temporally distant events. In the first case, no problematic MP causation arises, for either an initial physical or mental event causes mental events that eventually lead to my remembering the occurrence of certain events, or to my remembering that a theory makes a certain prediction; no brain trace (a physical event) caused by a prior mental event is involved. This is different for long-term memory. On the plausible assumption that long-term memory involves a brain trace (here P_2) that is caused by the direct experience of the prediction or occurrence (M_1/M_4), MP causation has entered the scene. If, however, MP is ruled out by CCP, it becomes unclear how belief in my memory of the prediction or occurrence can be justified since there is no causal connection between the original mental event and my ensuing belief which is based on a brain trace.

A defender of CCP could find a way out by reconstructing the causal chain such that no MP causation is involved. This works neatly in case a physical event stands at the beginning of the causal chain, and for occurrences: it could directly cause the brain trace which later causes my memory experience. However, the reconstruction fails for predictions because one cannot legitimately speak of *understanding* that a theory makes certain predictions if all that happens are physical events in the brain. The same goes for the case in which the causal chain is initiated by a mental event (say, a 'Eureka'): since MP causation is forbidden, the initial mental event cannot leave a brain trace and must directly cause other mental events leading to the memory experience; and that is highly implausible for any long-term memory.

The last setting, testimony, is no less problematic than memory. If the causal chain begins with a physical event (physical expression of the content of a theory, or occurrence of a physical event), then one *prima facie* has *two* occurrences of MP causation: first, when the direct experience causes a brain trace; and second, when the memory of the prediction or occurrence plus the intention to testify cause the person to testify (and any form of testimony involves physical events). Again, a defender of CCP might want to reconstruct the causal chain to the effect that the initial physical event directly causes a brain trace which causes the testimony. This, of course, stretches the concept of 'testimony' almost beyond recognition. Can we legitimately speak of 'testimony' when all that happens is that the words coming out of my mouth (or the sentences I type, etc.) are nothing but the effect of some brain events? What is the difference between such 'testimony' and, say, fluid aphasia, whose symptom is that streams of unintended words flow out of a person's mouth? But there is another severe problem still. If the causal chain begins with a mental event (not uncommon in the case of a realization that a theory makes a certain prediction), how is MP causation to be avoided? The first mental event must not cause a brain trace; but even if one (implausibly) waives this condition for long-term memory, how will the causal stream of only mental events eventually issue in a (physical) testimony? Given that few of us report having received a testimony via telepathy, it seems that MP causation is unavoidable.

A very relevant – and delicate – application of these insights is the very research program that many take to provide evidence for the causal inefficacy of mental events (and thus for CCP), namely the program initiated by Benjamin Libet. I have already argued (section 1.4) that the conclusion that the subjects' intention to move their finger was not caused by their intention but wholly caused by a prior brain event is unwarranted. Here, I wish to point out that *no* conclusion drawn from the Libet experiments to the effect that CCP is true can be justified, because the experiments rely on the very negation of CCP. This is because the crucial information, the time of the conscious intention to move one's finger, can only be obtained if the subject's intention to relate the memory of this event causes their testimony. And as can be seen from the last line of table 3.1, the causal chain beginning with a mental event (here, the memory of the intention to move a finger) cannot avoid MP causation.

To sum up: on CCP, any scientific theory that entails CCP can only be justifiably believed if one of the following settings is possible:

- Belief in the theory making certain predictions can be held in the mind all at once, without reliance on memory or testimony. (This is implausible especially when the connection between theory and predictions is complex).
- A direct experience of the occurrence of certain events and/or belief in the theory making certain predictions can be stored in short-term memory, without reliance on long-term memory or testimony. (Again, implausible especially in the case of a complex connection between theory and predictions).
- Long-term memory of the belief that a theory makes certain predictions can be caused without there being an intermediate M-event of *grasping* that the theory makes those predictions. (This verges on absurdity).

- Testimony of both occurrences and predictions occurs as the result of a wholly physical causal chain. (Stretches the idea of 'testimony'; leaves out belief as a mental event; not applicable in the case of a mental event being the start of the causal chain).

Since all these options are clearly less plausible than the *prima facie*, common sense view that mental events cause physical events, I conclude that in order to justifiably believe a theory that entails CCP one must assume its falsehood, and so CCP is (in an epistemic, not a logical sense) self-refuting. But, again, this insight is startling only for epiphenomenalists; (reductive) physicalists, identifying mental event with physical events, have no issues with 'mental' causation to begin with.

3. The Heterogeneity Objection

> I ask you please to tell me how the soul of a human being (it being only a thinking substance) can determine the bodily spirits, in order to bring about voluntary actions. (Elisabeth in her first letter to Descartes, May 6, 1643)[16]

The above quote by Elisabeth of Bohemia contains all the stuff of which is made what I call the 'heterogeneity objection'[17] against interactive dualism. Elisabeth picks up on the Cartesian distinction between the mind ('soul') which has mental properties ('thinking substance'), and the material body with its physical properties, and is then left puzzled how the former can bring about any effect in the latter.
The basis of Elisabeth's (as well as many other people's) incredulity is the (supposed) radical difference of body and mind, together with the intuition that such stark disparity prevents mutual causal interaction. To put it more succinctly, in the words of Robb and Heil:

> If minds and bodies are so radically different (…) it is not easy to see how they could interact. (Robb and Heil 2019)

The heterogeneity objection has traditionally been levelled against substance dualism (as in Elisabeth's case), which, to remind, is the only feasible kind of interactive dualism and the one ultimately under discussion here. For the sake of completeness, it should be mentioned that some see the objection also as applying to any form of property dualism that allows mental properties to affect physical states *qua* mental properties, and not via subvenient physical properties (see also Robb and Heil 2019, § 4). As a matter of fact, this is of interest only for causal overdetermination accounts (where the question mark in fig. 1.1. is removed), but nonetheless surprisingly so (who would have thought that something as innocuous as a mental property overdetermining a physical effect could come into the firing line of Elisabethan skepticism?).
Now if radical difference between the mental and the physical is the problem, one could of course try to deny from the outset that there is such radical difference. However, such a denial *tout court* is not open to dualists. One of their basic claims *is* the fundamental difference between the mental and the physical, which gives rise to and justifies their dualistic ontology. Denying the difference would mean opening wide the gate to the city hard-won in endless battles against physicalist onslaught. On the other hand, dualists should be wary of claiming *too much* difference. Fundamental difference need not entail that there is *nothing* in common. In fact, such an assertion would make it unnecessarily hard to positively spell out the causal interaction of mind and body. I suspect that Descartes' famous (or infamous?) denial of extension to the soul is one such instance of 'throwing out the baby with the bathwater'.
Let us now see how the objection goes in more detail.

3.1 Heterogeneity as mere difference

On this version, difference *tout court* is the problem. Thus again Robb and Heil:

> If minds and bodies are so radically different (…), it is not easy to see how they could interact. (ibd.)

[16] (Shapiro 2007, 62)
[17] Robb and Heil (2019) use the term 'causal nexus objection'.

Robb and Heil point out that that radical difference should be understood as the *absence of any shared properties*: in other words, minds and bodies have no properties in common. Unfortunately, not all philosophers writing about the topic at hand spell out that clearly what they mean by 'radical difference'. We thus have to face a certain amount of uncertainty and, possibly, ambiguity. Juhani Pietarinen, for example, thinks that there is no reason to believe that radical difference entails non-interaction, albeit without explicating precisely what he means by "different nature":

> [T]he whole problem contained in such questions arises simply from a supposition that is false and cannot in any way be proved, namely that, if the soul and the body are two substances whose nature is different, this prevents them from being able to act on each other. (Pietarinen 2009, 134)

However, since Pietarinen is defending a broadly Cartesian approach to mental causation, we can plausibly assume he has in mind a Cartesian type of difference in nature, which is virtually a lack of shared properties. Now, the answer to the question whether radical difference entails non-interaction depends in no small degree on the availability of good paradigm cases where a 'radical difference' prevents interference of one thing with another. Perhaps abstract objects constitute such a case. Using Robb and Heil's criterion, one might indeed plausibly assume that abstract and concrete objects share no properties other than existence (which to denote as a property or predicate is controversial but also renders the interaction question trivial, because obviously a non-existing thing cannot interact with anything). But this raises the question of how we can have knowledge of abstract objects, given one assumes a causal theory of knowledge (which most if not all philosophers reasonably do). The possible answers are: either we do after all not have knowledge of abstract objects (outlandish); or what we took to be abstract objects turn out to be concrete objects (only slightly less outlandish); or abstract objects exert causal influence after all (the least outlandish option in my view). Of course this is a vexed and complex question, and I am not going to explore it here; all I wish to show is that it is far from clear that radically different objects cannot interact. At any rate, the difference between minds and physical systems is not as great as that between abstract objects and physical systems. Minds are, like physical systems, concrete objects. That's perhaps one non-trivial attribute the two share. And more can be found (see section 3.2).

Perhaps something like the suspicion of an incoherence lurks in the back of the objectors' minds. That incoherence cannot be self-evident, at any rate, because otherwise there would be no sane-minded, intelligent defenders of interactive dualism. A way of showing a non-evident incoherence is by constructing an analogous case of an evident incoherence and then establish an argument from analogy. I am not aware of any such argument in the literature (which is unsurprising, given the fact that evident cases of incoherent interaction are hard to come by, as indicated above). However, Descartes himself seems to have attempted the inverse in his letter to Elisabeth from 21 May 1643 (as well as in other places), namely constructing a sufficiently analogous, evident case of *coherent* interaction:

> [S]upposing that heaviness is a real quality, of which we have no other knowledge but that it has the power to move a body in which it is toward the center of the earth, we have no difficulty in conceiving how it moves the body, nor how it is joined to it; and we do not think that this happens through a real contact of one surface against another, for we experience in ourselves that we have a specific notion for conceiving that; and I think that we use this notion badly, in applying it to heaviness, which, as I hope to demonstrate in my Physics, is nothing really distinct from body. But I do think that it was given to us for conceiving the way in which the soul moves the body. (Shapiro 2007, 66)

He argues that the scholastic theory of how a massive body is drawn towards the center of the earth is, despite being false, *coherent*. The scholastic theory has it, as Descartes points out in the same letter, that certain qualities of bodies, like heat and heaviness, are *real*, that is "have existence distinct from that of the body, and by consequence, to be substances, even though we have named them qualities". On the scholastic view then, an immaterial but real 'substance' causes a physical motion. Descartes, though believing that such qualities are "nothing really distinct from body", was convinced that the conceivability of the scholastic theory could be a guide to "conceiving the way in which the soul moves the body". The analogy is clear: in both cases something immaterial moves something material. No incoherence can be detected in the gravitational case; therefore, insofar as the analogy is intact, none can be found in the mental causation case. I conclude that in the absence of a clear precedence and of incoherence, a mere, even if radical, difference between mind and matter does not warrant the conclusion that they cannot interact.

3.2 Heterogeneity as the lack of a causal nexus

The 'radical difference' version of the heterogeneity objection did not produce any obvious incoherence in the idea that an immaterial mind can interact with the body. Still, a coherent view can be false. If it can be shown that mind-body-interaction lacks a crucial feature for causality, the objection might gain considerable traction. Such a position is taken by the 'causal nexus' objection. The idea is encapsulated in the following causal nexus (CN) principle put forth by Robb and Heil (2019):

> (CN) Any causal relation requires a nexus, some interface by means of which cause and effect are connected.

The objection consists in arguing that mind-body-causation lacks such a nexus and hence violates CN. Elisabeth's worries about mental causation seem to be grounded in something like a suspected violation of CN:

> For it seems that all determination of movement happens through the impulsion of the thing moved, by the manner in which it is pushed by that which moves it, or else by the particular qualities and shape of the surface of the latter. Physical contact is required for the first two conditions, extension for the third. (Letter to Descartes from 6 May 1643; Shapiro 2007, 62)

She explicitly mentions physical contact and extension as conditions for causation of movement. Since Descartes had previously denied the mind spatiality altogether (and hence the possibility of physical contact), it is clear why Elisabeth has trouble with Cartesian mental causation. I take it that Elisabeth's worry lies at the heart of many people's intuition that something immaterial is too 'airy' or 'thin' or 'insubstantial' to cause anything in the physical world. Thus, the objection Elisabeth voices depends first on the assumption that causal relations must fulfill CN, and second on the subordinate condition that the feature that meets the requirements of CN in the case at hand is physical contact; from this it follows that non-spatial souls cannot causally influence bodies.

What can be said in reply? A first possibility is to question the demand that CN be met. One way to do this is to ask whether we have any robust notion of a causal interface in those physical cases which are uncontroversially instances of causation (say, two billiard balls bouncing off each other); for if we don't, then it seems unfair to ask for one in more controversial cases. *Contact* is of course the first candidate that leaps to mind for the billiard ball scenario. The first thing to note here is that 'contact' is not what common sense takes it to be. There is no straightforward equivalent to macroscopic solidity on the atomic and subatomic levels. The atoms the billiard balls ultimately consist of are not miniature versions of the billiard balls; instead, they partly consist of electron 'clouds' either sticking together (producing cohesion) or repelling each other, and a lot of 'empty space' between the electron shell and the atomic nucleus. Thus, given that macroscopic contact consists in the mutual attraction or overlap of electron clouds, the causal nexus for this kind of causal event cannot be contact, for contact is the *result* rather than the *precondition* for those events.

Of course, the causal nexus may be something other than contact. But even if we could find a better physical candidate, a regress is likely to be looming, since for every level we get CN analyzed in terms of effects produced at a lower level. If there is one fundamental physical level such that no level lower than it can be found, then the regress must stop there with a brute-fact, unanalyzable causal nexus, which raises the question why such an unanalyzable nexus could not be posited for mental interaction. On the other hand, if there is no such fundamental level, then the regress becomes infinite (something to be avoided if possible). An easy way out would be to simply construe causal nexus as *spatiotemporal overlap* without further analysis: a causal nexus is given just in case two (or more) objects overlap spatiotemporally. Still, on a Cartesian non-spatial picture of the mind, interaction between minds and the physical world is not available even with this unassuming condition, but more on that shortly.

There are two more general problems with this solution. The first is that there may be spatiotemporally overlapping objects which do not exert any causal influence on each other (at least this is conceivable). True, spatiotemporal overlap might be construed as just a *necessary,* not a *sufficient* condition for a causal nexus. But then the instances where objects are causally inert with respect to each other and the instances where they exert causal influence on each other must be differentiated by something, some 'extra ingredient' present only in causation. What that could be is unclear. This is admittedly not a compelling reason to abandon the identification of spatiotemporal overlap as the causal nexus. It might, however, invite one to rethink the Cartesian notion of the non-spatiality of minds: if minds were allowed to participate in spacetime, then they could be causally effective even if a causal nexus in terms of spatiotemporal overlap is required (see below).

The deeper and more universal problem, however, is that not all effects in physics that can conceivably be called 'causal' can be explained in terms of spatial overlap. Take gravitation: the earth exerts an influence on the moon without ever touching it. Or consider electromagnetic attraction/repulsion: charged objects attract or repel each other by a postulated electromagnetic field, but this field is not another physical object that somehow mediates physical contact between the two objects[18]. Also, so-called 'EPR' experiments with entangled photons have proved beyond doubt that those photons, however widely separated, still form a single physical system in accordance with the predictions of quantum mechanics; one consequence of this is that a measurement of one of the photons *instantly* alters the state of the other photon in exactly such a way as is to be expected were the two one single system (Aspect, Grangier, and Roger 1981; 1982; Aspect, Dalibard, and Roger 1982). The upshot is that the traditional notion of the locality of physical systems is deeply shaken, if not shattered[19].

---

[18] Note that this leads some to deny the existence of fields altogether, e.g. (Landry et al. 2017; Lazarovici 2018; Deckert, Esfeld, and Oldofredi 2020).

[19] See

I am saying this not to suggest that EPR entanglements could give us any conceptual guidance for how mental causation works, but rather to underscore the point that modern physics is far from providing a universal account of causal nexuses. And if we do not really know what we are talking about in the (presumably) clearer case, what justifies our insisting on the fulfillment of the ill-understood condition in the more obscure case? Besides, mental-to-mental causation may provide further reason to discard contact as the causal interface. Thoughts (on the dualist understanding), being non-spatial, definitely have no spatial contact with each other, and yet one thought can cause another.

Ultimately, the quest for a causal nexus may prove futile. If causation requires a causal nexus, then wherever causation takes place, there will consequently be a causal nexus, whatever it may be, in the physical-to-physical as well as the mental-to-physical case. It may thus be more fruitful to focus on whether mental causation actually takes place empirically (see chapter 9), or on other metaphysical aspects of mental causation or causation in general (which I will do in the remainder of this chapter).

However, before plunging into more positive aspects concerning the bridging of the 'heterogeneity gap', I'd like to dwell a little more on the aforementioned worry that the non-spatiality of the mind makes it hard to see how it could interact with the body. Independent of the question of a causal nexus, how much of a problem is that? On Cartesian substance dualism, the mind has traditionally been construed as non-spatial – more precisely, not describable by spatial attributes – to sharply distinguish it from the spatial body. But of course one can remain an interactive (substance) dualist without adhering to everything Descartes wrote. It seems clear that by understanding non-spatiality as 'not belonging to spacetime', then it indeed poses a grave problem for mental causation – at least insofar as we understand timelessness and non-spatiality to be genuine hindrances to causal interaction with the physical realm (but note the intricacies concerning abstract objects gestured towards in section 3.1). However, one could take a more moderate view and consider the mind spatially *non-extended*, but spatially *located*, like a mathematical point. Then, the causal problem does not immediately arise, for it is not obvious that mathematical points could not be part of the makeup of the physical world. In fact, André Deckert's and Michael Esfeld's very elaborate account of a parsimonious ontology of the material world (Landry et al. 2017) assumes precisely such points as fundamental. Could minds not be point-like in this sense, hence non-extended but still in spacetime? As a matter of fact, non-extended point particles are the only relativistically invariant objects in today's physics. This seems thus to be a conceptually viable approach.

There is, however, yet another way of solving the puzzle: namely to drop even the non-extendedness of the mind. Non-dualist William Lycan considers this in his "redress" of dualism (Lycan 2018). "Why not suppose that minds are located where it feels as if they are located, in the head behind the eyes?" he asks (ibd., 27). To the complaint that there is already the brain and other "physical stuff" in that place, and that no two things can occupy the exact same spatial region, he has a threefold reply: (1) the spatial exclusion rule holds for physical objects, not for immaterial ones like the mind; (2) physical stuff is mostly empty space (cf. the makeup of atoms); (3) there is at the very least no conceptual or conceivability problem, since "ghosts and disembodied spirits supposedly move about in space, and that does not cause readers/audiences any conceptual dissonance" (ibd.). I am not so sure about (2) (the presumed empty space may turn out to be filled with hitherto unknown physical objects), but apart from that, I agree with Lycan. In fact, even Descartes considered the attribution of matter and extension to the soul unproblematic, so long as it was understood as flowing from the body-soul union (Letter to Elisabeth from 28 June 1643; Shapiro 2007, 71). In the same vein, E.J. Lowe suggests as part of his 'non-Cartesian substance dualism' to attribute spatial and even physical properties to selves which are nevertheless distinct from the body (Lowe 2006).

Is this enough to satisfy even the staunchest skeptic? Again, Elisabeth of Bohemia is a good test case. She cannot believe that something that is not material can causally impact matter. For her, the mind would have to be both extended *and* material (something staunchly denied by dualists); in fact, she says it would be easier for her "to concede matter and extension to the soul than to concede the capacity to move a body and to be moved by it to an immaterial thing" (Letter to Descartes from 10 June 1643; Shapiro 2007, 68). She also reports that she cannot conceive of an imperceptible cause as other than a negation of matter (ibd.), and this apparently means for her a denial of all the features that could establish a causal connection. Relatedly, she expresses her perplexity about her senses showing her *that* the soul does move the body, without her knowing *how*. All in all, it seems that Elisabeth had more trouble with *imagining* dualistic mental causation than to pin-point a substantial metaphysical puzzle. Plausibly, a similar difficulty haunts many people nowadays. Can anything be done about it? Can we help bridge the imaginative gap?

For people like Elisabeth, who at least confessed belief in God, a toehold is readily available, and it is surprising or even puzzling that neither her nor Descartes ever bring it up in their conversation. It is that God, according to the Christian tradition, often interacted with the physical world in the past and will do so in the future; the Bible, the Old as well as the New Testament, are full of such accounts. But if God, being a pure spirit (in fact the paradigm example of a spirit, see John 4:24), can bring about physical effects, and this (apparently) caused no conceptual difficulty with the vast majority of Christians, then why exactly should it be an insurmountable problem to conceive of immaterial souls doing a similar thing? God's infinity, his un-createdness and omnipotence surely are relevant only to gauge the *scope* of his interventions, not the *possibility* of intervention itself. So, theists, of all people, should have no trouble whatsoever to conceive of the interaction of something immaterial with something material, even if they ultimately deny that God or souls causally intervene in nature (as do deists or psychophysical parallelists like Leibniz, who both deny God's interaction on grounds other than a lack of ability on God's part).

Where does this leave us? Upon metaphysical scrutiny, the causal heterogeneity objection proved tenuous: for lack of a clear precedence, it remains unclear how much of a difference there must be for two types of entities to lack the ability of causal interaction; on the other hand, the demand for a causal nexus turns out to be a two-edged sword, because it is not even possible for the most mundane exemplars of causation to specify what the causal nexus consists in. I conclude that ontological heterogeneity between mind and matter, as assumed on dualism, does not warrant skepticism concerning interactive dualism.

3.3 The SPL account of causation

So far, what I tried to show was that there is no convincing argument against the possibility that non-physical minds interact with the physical world. That leaves open *how* dualistic interaction takes place. What I wish to do now is present my account of the metaphysics of dualistic mental causation. I am doing this for two reasons: first, because it is always better, if possible, to have something to say about *how* something happens than to just say *that* it happens. Second, the metaphysics of the laws of nature is inseparably bound up with the metaphysics of causation. Since I will have a lot to say about the interplay between the laws of nature and interactive dualism (see part IV), it stands to reason to first clarify my position on causation, and what better context could there be than the causal heterogeneity objection?

The account I am defending is what Richard Swinburne calls the 'substance-powers-liabilities' (SPL) account, a modified version of dispositionalism. It is modified because dispositionalism typically considers causation in the inanimate physical realm, but the causal agency human beings possess must of course deviate in some important ways from mere physical causation (see also section 1.2). I will first draw a rough (but sufficiently fine-grained) sketch of dispositionalism and then explained how and why the SPL theory deviates from it.

Dispositionalism typically rests on a substance ontology. A further important tenet is that substances possess *intrinsic natures or essences,* which are kind-specific and at least partly consist in dispositions: electrons have their specific set of dispositions, as have protons, water molecules and metals (using here a liberal substance ontology). A disposition (or power) is characterized by what Bird calls the "stimulus-manifestation relation" (Bird 2007). Roughly speaking, a substance possesses disposition *D* iff in circumstances *C* and in the presence of stimulus *S* manifestation *M* occurs. On dispositionalism, dispositions are intrinsic causal properties of substances and thus the truthmakers of both causal connections and laws of nature.

An immediate problem for applying dispositionalism to dualistic mental causation is that modern dispositionalism is designed to account for lawful connections in (physical) nature: it is the strict regularity with which the stimulus-manifestation relations appear that makes them amenable to be the supervenience basis for laws of nature. Events caused by free agents (see section 1.2), however, are precisely not what should come under the sway of or what should be reconstructed according to laws of nature. Free agents perform actions because they *choose* to do so (not because some irresistible disposition made them perform them), and refrain from doing them because, again, they choose to, not because some external disposition thwarted the performance.

Swinburne therefore offers a modified SPL account for free agents:

> On the SPL account of cause substances cause effects because they have the powers they do; and if they have a liability necessarily to exercise some power, they will inevitably do so (or if they have a probabilistic liability to do, then—with the degree of probability involved—probably they will do so). But this account allows as metaphysically possible the existence of substances which have powers and liabilities of a kind peculiar to that substance, which would not necessarily be possessed by any substance which had all the same other properties as it; the powers and liabilities of such a substance could be a 'brute fact' about that substance, not deriving from its belonging to some universal kind. Further, since the SPL account treats 'exercising causal influence' as a basic category, not analysable in terms of any other category, it also allows as metaphysically possible the existence of a substance which has a power but no deterministic liability inevitably to exercise that power, or even a probabilistic liability to do so. (By a probabilistic liability to exercise a power I mean a natural probability of some particular value *p* that the substance will exercise the power.) Such a substance could exercise its power to cause some effect because it intends—that is, tries—to cause that effect, and not because it has a propensity to cause the effect. I now argue that human beings are substances of this kind. (Swinburne 2013, 132-33)

He basically considers human beings as substances with their own (mental) causal powers, albeit in such a way that those powers are exercised not because of an 'irresistible' propensity (as would be the case with inanimate matter), but via conscious intention. Such powers require an individuation not in terms of species membership, but in terms of belonging to an individual substance, according to Swinburne. The volition *V* to cause *E* (say, lift my arm) is then not an essential disposition of the class 'human being', but a disposition that only I at a certain moment have. The crucial move is, of course, to release human beings from the sway of the laws of nature – our behavior in terms of effects caused by dispositions must not be governed by dispositions of the kinds that can be found in inanimate matter, on pain of losing freedom.

The SPL account squares well with the theory of the laws of nature I am going to develop in chapter 14, because the underlying metaphysics is in both cases dispositional. But that is not the only reason for me to adopt the SPL account. Other approaches to causation have severe drawbacks, which reflect the problems of their correspondent nomological theories. A dispositional conception catches our deep-seated intuition that there is something in or about the cause that 'makes the effect happen': the 'oomph' of one billiard ball to push the other away, or, in more elaborate language, the "derivativeness of the effect from the cause" (E. Anscombe 2018, 62); let's call this conception of causality, to use a term from the first dispositionalist, Aristotle, *efficient causation*. Humean theories, for example, explicitly and deliberately deny efficient causation. To be sure, Humeanism is typically a theory about physical nature, but some Humeans have also developed accounts of mental causation, for example Michael Esfeld (Esfeld 2019; 2020; 2022). His 'Super-Humeanism' is particularly interesting because it accords libertarian free will to persons, without granting them dispositional powers. Esfeld goes so far as to affirm that human choices are free and not determined by a prior physical history (see his treatment of Van Inwagen's Consequence Argument in section 1.2), so that human choices are indeed the truthmakers of some particle movements, albeit without there being a disposition involved. But that leaves out the most interesting metaphysical part, and even denies that there is a need for clarification. True, on Super-Humeanism matter does not possess intrinsic qualities and thus the problem of the modification of its behavior by mental interaction becomes more of a formal than a substantial metaphysical challenge (see section 1.2); but since Esfeld considers persons as "irreducible" to physical matter, and even sometimes speaks of "mind points" as opposed to "matter points", he owes us an account of how persons can do something that mere matter apparently cannot. Further, even though it is no difficulty for Humeanism to accommodate a sudden change in the behavior of matter (for there is nothing grounding the behavior), if one adduces an ontologically distinct kind of being like persons, that very addition commits one to giving an account *how* that ontological kind brings about changes in matter. The most natural answer for a Humean would be to construe mental interaction just like the goings-on in inanimate nature, namely as mere *correlations* instead of *causation* along the lines described above. But that runs counter to the intuition I mentioned.

The same goes for a less radical approach, the universally quantified laws (UQL) account of causation. Hoffman and Rosenkrantz present it in their paper about dualistic interaction:

> It is extremely plausible that causal interaction can be understood in terms of universally quantified general laws. If causal interaction is understood in terms of such general laws, then there can be functional or correlational causal relationships which do not involve the production of motion by the transference of motion. (…) For instance, consider the Law of Universal Gravitation. According to this law, there is mutual gravitational attraction between any two pieces of matter. As a result, two such pieces of matter accelerate toward each other. It is a law that each acquires a motion, but it is not the case that motion is transferred from one to the other. (Hoffman and Rosenkrantz 1991, 198-99)

If extended to mental causation (which Hoffman and Rosenkrantz do have in view), the account has it that it is a universally quantified law that mental event-type M (say, a volition to raise one's arm) is followed by (or correlated with) physical event-type P (the firing of certain synapses in the supplementary motor field). Nothing more needs to be said about the connection between M and P. Now, the UQL account elegantly sidesteps any possible difficulties that could arise from the attempt to give a metaphysically rich account of causation by simply refusing to say whether or not there is any "deeper explanation" of the motion brought about (ibd., 199). Again, although the UQL account is in principle combinable with a range of metaphysical theories, its agnosticism is unsatisfactory. We rightly demand a deeper explanation, and the SPL account provides one.

Although one might legitimately stop here, I wish to go a little further. Robb and Heil, in considering the causal heterogeneity objection, formulate a possible bridging principle between mind and body:

> If something in a soul could become present in a body, this could bridge the immaterial and material. (2019)

Let's call this the 'becoming present' (BP) principle. I wish to combine this principle with my SPL analysis of mental interaction. The demand for BP is reasonable, not only for the mental-to-physical case. After all, a dispositionalist analysis yields that a billiard ball, in circumstances *C* (here, the ball rolling toward another ball) has the disposition *D* to bring about manifestation *M* (here, to push the other billiard ball into motion). But a physicist would give us some more details: for example, the first billiard ball must minimally have a certain momentum to push forward the other billiard ball[20], a momentum which partly or completely is 'transferred'[21] to the other ball. If one takes this transfer as a general feature of causation, one might speak of something of the cause (e.g., motion) *becoming present* in the effect. (This covers even cases where the cause doesn't seem to lose or transfer anything). Thus, BP seems to me to enshrine a reasonable minimal requirement for dualistic mental causation, indeed for any kind of causation. The question that then arises is what this 'something' that 'becomes present' is in the case of interactive dualism.
David Fair suggests it should be, as in the mere physical case, *energy*. Though originally conceived to account for physical-to-physical causation, he submits that his 'energy flow theory' of causation might apply to interactive dualism as well:

> Actually the theory that causation is a matter of energy flow might be compatible with certain forms of dualist interactionism; the reduction of the mental to neurophysiology is not essential. (Fair 1979, 237)

This makes sense insofar as at least many state changes in the brain require additions (or, for that matter, subtractions) of energy[22]; at any rate, if energy transfer from the mind to the body in fact occurred, this could at once solve the heterogeneity problem.
The difficulty with this proposal is that it commits the dualist to positing (in-principle measurable, or at least quantifiable!) energy for the mind, a solution that, as I already insinuated, is problematic (for details

---

[20] I am well aware that there are reasons not to consider quantities like momentum, mass or charge as 'bringing about something'. The main argument here is that only position should be taken as fundamental, since only position can ever be observed directly. All the other quantities should then be analyzed functionally in terms of position. (For an elaborate account of this see (Esfeld 2020; Landry et al. 2017)). The question whether such radical ontological parsimony is adequate will be taken up in part V in more detail. In particular, it requires an abandonment of powers metaphysics for some kind of Humeanism. It is therefore legitimate to continue speaking in terms of a powers ontology until the question has been thoroughly treated.

[21] The word is in quotation marks because it is not clear metaphysically that there truly is an entity called 'momentum' which disappears in one object and appears in the other. Still, mathematically as well as phenomenally, one ball loses something (motion) which the other gains. It is legitimate, therefore, to speak at least tentatively (or perhaps metaphorically) of a transfer.

[22] There are accounts that claim to explain mental causation without energy expenditure. They will be discussed in chapters 5 and 6.

see chapter 5). To anticipate the upshot: one should avoid commitment to the ascription of physical properties to the mind (except for spatiality), for this at once commits one to the strict mathematical corset of physical theories (representing the pertinent laws of nature), which in turn threatens the very freedom which is a central motivation for interactive dualism (see section 1.2). This, however, at once creates a dilemma. If a physical property like energy or motion is to become present in the effects brought about by the mind, then, by BP, those physical properties need to have been or be present in the mind, which is to be avoided. Another problem with the 'energy transfer' approach is that treating energy as a transferrable kind of 'stuff' does not sit well with the Noether theorem. Thus Maudlin, Okon and Sudarsky:

> The whole notion of energy as a substance is a fiction, a holdover from an earlier, and now obsolete, physics. (Maudlin, Okon, and Sudarsky 2020, 69)

On the other hand, if the 'something' in the mind becoming present in the body is to be wholly non-physical, it is not immediately clear what it could be.

My proposal is to solve the dilemma by conceiving of the transferred 'thing' (or, better, property, or mode) as something of which physical motion is only a special case. (Motion is what we ultimately want to get at; energy is just the precondition for motion). Let's call this wider notion MOTION.

Inserted into BP, we get

> If MOTION in the mind could become present in the body, this could bridge the immaterial and material.

Of course, this raises the question what MOTION in an immaterial mind and MOTION in physical objects (= motion) have in common. It is clear that MOTION must *not* be thought of as distance traveled per time unit, which Michael Esfeld rightly complains would be nonsense if applied to the mind (Esfeld, personal communication). What is conceivable, however, is the attribution of *metaphorical* motion to the mind. At least our language clearly suggests this as a possibility: we say we are 'moved' (emotionally, i.e. mentally) by a film or speech, or that someone is mentally 'agile'. Concerning volitional bodily actions, which are of most interest to interactive dualism, it is less obvious that our language has a concept for motion of the mind; but some reflection on the phenomenon of akrasia (weakness of will) shows that we can meaningfully conceive of mental motion (or the lack thereof). Consider yourself lying in bed in the morning after a bad night: you are awake and know you should get up. What hinders you is not some physical handicap (a broken leg, or strings fastening you to the bed), but your difficulties to muster enough *willpower* for finally greeting the day. We know that we must use our will (a mental endeavor) to get up; we must, so to speak, move our will in order to move our body. What is this but mental (self-)motion?

So far we've been speaking mainly of intuitions and language. Can we flesh out those initial thoughts philosophically? Timo Kajamies (Kajamies 2009) works out in detail two interlocked theses from Descartes' works about how motion can be present in the mind, and how this motion can become present in the body. His basic idea is that the property of the effect must be in the cause either 'formally' or 'eminently'. After some tweaking and honing, Kajamies presents his 'Causal Reality Principle 3' (CRP-3):

> Causal Reality Principle 3 (CRP-3)
>
> In order for x to impart F to y, x must either
> a) actually contain F at least to the same degree than y contains F
> or

b) contain F eminently, eminent containment understood in accordance with the Eminence Thesis 3 (ET-3).

ET-3 reads as follows:

Eminence Thesis 3 (ET-3)

Y's property Φ is eminently contained in X if and only if: Φ is not formally contained in X; X is an entity displaying a greater degree of relative independence than Φ (i.e., higher up the ontological hierarchy than Φ); and X contains formally, and at least to the same degree as Y, the same kind of active power that is the metaphysical basis of Φ.

Let's unpack this. By some property Φ being *formally* contained (or *formally* inherent) in some entity X, Kajamies means that Φ exists in X precisely in the way in which X is represented in our ideas – trianglehood in triangles, motion in bodies and so on. Part of the *eminent* inherence of some Φ in X is that Φ exists in X in a way different from how it is represented in our ideas. For example, motion exists eminently in God – he can move physical objects, but not as we or animals do, with the aid of our bodies. Another important notion here is 'relative independence' in connection with 'ontological hierarchy'. The only completely independent being, for Descartes as for theists in general, is God; he is on top of the ontological hierarchy. Next are created beings like angels, humans and animals (notwithstanding a ranking between them, but that is not the point now), and on a yet lower-level properties and modes which depend for their exemplification on created beings (Kajamies 2009, 167-68).

Finally, 'active power': this is the overarching topic of the anthology of which Kajamies' essay is a part, and it is understood throughout the book as a power that is "capable of being and acting 'in itself', producing effects completely spontaneously, with nothing external affecting it" (Pietarinen and Viljanen 2009, 1). 'Active power' means of course that we're confronted with a kind of powers account here. However, the addition 'active' is crucial. After all, on a powers metaphysics, physical objects have their powers, too, but those powers are crucially different from the mind's power in that they cannot move themselves. Thus, the used notion of 'active' is exactly what I claimed is necessary for free will (see section 1.2, especially Moreland's conditions for libertarian freedom) and reason and leads to a rejection of the causal closure principle (chapter 2).

If we replace x by 'the mind', y by 'the brain' and Φ respectively F by 'motion' in CRP-3 and ET-3, we get the following:

CRP-3*:

In order for the mind to impart motion to the brain, the mind must either a) actually contain motion at least to the same degree than the brain contains motion or

b) contain motion eminently, eminent containment understood in accordance with the Eminence Thesis 3 (ET-3).

ET-3*:

The brain's property motion is eminently contained in the mind if and only if: motion is not formally contained in the mind; the mind is an entity displaying a greater degree of relative independence than motion (i.e., higher up the ontological hierarchy than the

brain); and the mind contains formally, and at least to the same degree as the brain, the same kind of active power that is the metaphysical basis of motion.

As regards CRP-3*, option b) clearly applies. As far as ET-3* is concerned, it makes perfect sense the way it is spelled out: motion is not formally contained in the mind; the mind, being an entity (or substance), is higher up the ontological hierarchy than motion, which is a property or mode; and the mind possesses *ex hypothesi* at least the same degree of active power as the brain.

If one is prepared to accept a powers ontology in the first place (along the lines of Swinburne's SPL account), I submit that the above principles might be a good place to start, perhaps even a satisfactory account of dualistic mental interaction.

III. Scientific Objections to Interactive Dualism

4. The Objection from Conservation Principles

We now turn from metaphysical objections against interactive dualism to scientific ones, or at least ones that directly invoke scientific discoveries or principles. As we have seen, the arguments in favor of the causal closure principle (CCP) already include references to scientific theory-building principles and to empirical findings, but more in the sense that a defense of CCP takes recourse to them than that CCP naturally follows from them. The objections under consideration now, by contrast, are taken to be a direct consequence of well-established laws of nature, or (in the case of the empirical objection, see chapter 9) of well-established empirical findings. We begin with the most widespread objection, the objection from conservation principles.

It is widely believed among philosophers of mind that the conservation principles for energy and momentum (henceforth referred to as "conservation principles"[23]) pose a serious, if not fatal problem for interactive dualism. (The objection is typically predicated energy conservation, but the conservation of momentum is equally important, and revealingly neglected[24].) However, a thorough exposition of the objection is hard to find, and the proponents differ from each other as to what the argument actually consists in. This chapter serves the purpose of finding an adequate expression of the objection to begin with. Let's start with a selection of quotes that shows just how little care opponents of interactive dualism have given to specifying what precisely they take issue with (italics all mine unless indicated otherwise:

> Now, the principle of conservation of energy requires that the *total amount of energy in the universe remain constant*, even as it is continually transferred and transformed in and among the myriad systems of causal relations. If Descartes is right that a nonphysical mind can cause the body to move, for example, we decide to go to a concert and go, then physical energy must increase in and around our body, since we get up and go to the concert. In order, however, for physical energy to increase in any system, it has to have been transferred from some other physical system. But the mind, according to Descartes, is not a physical system and therefore it does not have any energy to transfer. The mind cannot account for the fact that our body ends up at the concert. (Flanagan 1991, 21)

> [T]he return signals, the directives from mind to brain (…) are not physical; they are not light waves or sound waves or cosmic rays or streams of subatomic particles. *No physical energy or mass is associated with them. How, then, do they get to make a difference to what happens in the brain cells* they must affect, if the mind is to have any influence over the body? *A fundamental principle of physics is that any change in the trajectory of any physical entity is an acceleration requiring the expenditure of energy*, and *where is this energy* [in mind-brain-interaction] *to come from*? It is this principle of the conservation of energy that accounts for the physical impossibility of ''perpetual motion machines'', and the same principle is apparently violated by dualism. This confrontation between quite standard physics and dualism has been endlessly discussed since Descartes's own day, and is widely regarded as the inescapable and fatal flaw of dualism. (Montero and Papineau 2016), 34-35)

---

[23] In most of the literature, those principles are referred to as "laws". There are good reasons, however, to distinguish them from textbook examples laws of nature, because they fulfill the role of framing principles. One might therefore address them as meta-laws, or, less bulky, as principles.

[24] The fact that momentum is seldom addressed is indicative of the same insufficient grasp of physics that makes those philosophers so impressed with the objection from energy conservation.

> How could disembodied consciousness cause anything? And how could the physical sequence of events in the material universe be disrupted by what is going on in the parallel mental universe? Causation in the material world works by energy transfer of some sort: transfer of motion, electrical energy, gravitational force. But pure consciousness could not give off energies of *these* kinds, so how does it cause changes in anything? How does it reach out and touch something? What mediates its alleged causal powers? There seems to be no answer to that question; we are simply told that this is so. (Mcginn 2000, 92, italics original)

> If we want to change anything within the system, we will either have to use the energy that is already within the system, or we will have to introduce energy from the outside. If we use the energy in the system, then the mind, since it is not within the body, can have no effect on the body. If we do not use the energy already in the system, then mass and energy are not conserved, or the system is not closed. However, if the mind is to effect a change in the body, then it must presumably introduce physical energy into the body. But according to our first proposition, *the mind is nonphysical, and so it cannot expend physical energy*. (Westphal 2016, 42; italics mine)

> We have good reason to believe that physical events constitute a closed causal system. Insofar as physical events have a causal ancestry, that ancestry can be described in purely physical terms. If this were not true, *if we had to import reference to non-physical events in order to account for physical goings-on, we would have violations of the conservation of energy and momentum*, and so forth. (Pollock 1989, 19; italics mine)

> Physics says that the amount of matter/energy in the universe is constant, but substance dualism seems to imply that there is *another kind of energy, mental energy or spiritual energy*, that is not fixed by physics. So if substance dualism is true then it seems that one of the most fundamental laws of physics, the law of conservation, must be false. (Searle 2004, 42)

> Dualism violates conservation of energy. If immaterial mind could move matter, then it would create energy; and if matter were to act on immaterial mind, then energy would disappear. In either case energy—a property of all—and only concrete things would fail to be conserved. And so *physics, chemistry, biology, and economics would collapse*. Faced with a choice between these "hard" sciences and primitive superstition, we opt for the former. (Bunge 1980, 17; italics mine)

Note the hodgepodge of aspects appearing across these quotations. There is, for example, the recurring theme that immaterial minds lack a crucial feature of physical objects, namely the ability to transfer energy (and momentum, for that matter), and so are unable to interact with the physical world (see Flanagan, Dennett, Westphal and McGinn). This argument, however, is a version of the heterogeneity objection (see chapter 3) rather than of the objection from conservation principles (henceforth 'OCP'). For clearly, if the mind is unable to cause anything in the physical world, how is it supposed to violate any conservation principles?

Then there is the rather unsurprising observation by Pollock that if the causal closure of the physical (CCP) does not hold, violations of the conservation principles by the mind will ensue. This at least properly addresses conservation principles, but the intended objection against interactive dualism is not clear. Is it that CCP must hold since otherwise, violations of the conservation principles would ensue, and such violations are not observed empirically? Or because we (allegedly) have empirical evidence that violations don't occur, and can therefore conclude that CCP remains intact, thereby excluding dualistic interaction? In the former case, the objection rests on a purely philosophical (not scientific, see section 2.2.1) principle; in the latter case, it relies on gapless empirical (particular) observations. In no case do we get an objection building on a genuinely scientific *principle* or *law*.

By far the most radical expression of (something like) OCP is the one proffered by Mario Bunge. His complaint that if immaterial minds could interact with physical objects, they would "create" energy or make it "disappear" is in fact not far off the mark, and closest to the worry physicalists have (or should have) about interactive dualism in connection with conservation principles. Whether the failure of energy conservation would be so widespread as Bunge claims, and thus whether "physics, chemistry, biology, and economics would collapse" is another matter, and we do not learn from him why we should believe this alarmist claim. We will come back to those questions in a moment.
The most apt version of the objection is the one offered by John Searle (also mentioned by Flanagan). He renders the principle of energy conservation as stating that the amount of energy in the universe is constant. This at last seems like a serious threat to interactive dualism: if true, it requires the dualist to either give up interaction altogether (for a non-physical mind could not add or subtract energy to/from the universe) or to make the mind physical as well (so as to participate in the mutual exchange of energy among things physical), or to construct interactive dualism in such a way that the non-physical mind interacts without altering the total amount of energy in the universe (something Searle doesn't consider). Unfortunately, Searle instantly blunts the force of the argument by talking of 'mental' energy, which would require a conversion into physical energy to be of any interest for the issue at hand.

I take it that two viable main lines of argument can be gleaned from above survey of extant versions of OCP: first, an empirical line, according to which violations of the conservation principles, which interactive dualism requires, simply do not happen as a matter of observational fact; and second, an *a priori* line that renders the conservation principles as asserting that the total amount of energy (and linear momentum) in the universe is constant, which in turn excludes interactive dualism, if indeed it entails addition and subtraction of energy/momentum to/from the universe. A thorough reply to the empirical line must wait until chapter 9. What can be said in response to the *a priori* line will be the concern of the remainder of this and the other chapters of part III. For a better overview, here's a reconstruction of Searle's argument as I understand it[25]:

P1 If non-physical minds influenced the physical universe, then the energy of the physical universe would not remain constant (*ex hypothesi*).
P2 The energy of the physical universe is constant (with nomological necessity).
C Therefore, it is false (with physical necessity) that non-physical minds influence the physical universe (P1, P2, *modus tollens*).

Clearly, an interactive dualist rebuttal of this argument must target either P1 or P2. The bulk of responses has typically been against P1, and we shall examine some of those approaches in chapters 5 and 6. However, it is P2 that can be refuted much more effectively. The reason for that is that Searle's formulation of the principle of energy conservation is false. Instead of *globally* and *categorically*, as Searle would have it, energy conservation applies *locally* and *conditionally*[26]. According to Searle's rendering, energy conservation applies

[25] The same argument can be run, *mutatis mutandis*, with momentum instead of energy.
[26] For most of the physical insights in the following paragraphs, as well as in section 5.2, I am deeply indebted to J. Brian Pitts.

to the whole universe at any given moment (globally) and without exception, perhaps even with necessity[27] (categorically). However, energy (and momentum) conservation is conceived of locally, i.e. with respect to single physical systems making part of the universe. The system under consideration can be made as small or large as one wants; in any event, it will have finite boundaries across which energy can 'flow' in or out. "Energy is conserved in system *s*" then means that the time differential of energy in- and outflux is zero. More precisely, energy conservation takes the form

$$(1) \qquad (t)(x)\left[\frac{\delta\rho(t,x)}{\delta t} + \nabla \cdot J(t,x) = 0\right]$$

where $\rho(t,x)$ is the density of energy and

$$\nabla \cdot J = \frac{\delta J_x}{\delta x} + \frac{\delta J_y}{\delta y} + \frac{\delta J_z}{\delta z}$$

is the 'divergence' of the current density vector field J (with magnitude and direction at each point), a measure of how fast energy 'flows' out of the place. (1) is called the 'continuity equation'. Its content can be compared to a room in which there are several people. Barring conceptions and deaths, the number of people can always be accounted for by tracking how many people leave or enter through the doors.

Seen this way, energy non-conservation is ubiquitous. Define the water in your kettle as the physical system under consideration: above equation will not come out zero, because more energy will get into the water from the environment than from the water out into the environment. Of course, one can enlarge the system by including the kettle's energy source, but even that will likely not suffice, for all depends on the balance of energy in- and outflux, which can vary at any moment in systems that allow the exchange of energy with their environment. To truly get an energy density flux equation that yields zero, one would have to draw the boundaries around the whole universe. However, there is considerable doubt among cosmologists whether the total amount of the universe's energy can even be defined (Peebles 1993, 139). Even if calculation is meaningful, the results sometimes disagree or give surprising results (e.g. Nester, So, and Vargas 2008). Total energy could fail to be defined because there is just too much of it, giving an integral that diverges to infinity, or perhaps because energy is bound up with coordinate systems and hence cannot be globally integrated if the universe requires multiple coordinate systems.

So much for locality. Even so, it could still be possible that energy cannot be non-conserved *all things considered*. This is where the other aspect, conditionality, comes into play. Conservation principles (for energy and momentum) are a consequence of symmetries specified by Noether's first theorem (Noether 1918; Goldstein 1980, ch. 12-7). 'Symmetry' must here be conceived more widely than just, say, the mirror symmetry between one's two hands. A *continuous* symmetry, as figuring in Noether's theorem, can essentially be a *translation* (in space, but also in time!) by an arbitrarily small quantity. A perfect sphere, for instance, is continuously symmetrical, because any amount of rotation will yield the exact same form. By contrast, a perfect cube is symmetrical, but not continuously symmetrical. Its symmetry is only preserved

[27] To be sure, Searle makes no mention of necessity, but it stands to reason that some form of necessity is required to rule out interaction *a priori* (see also the pertinent discussion of CCP in chapter 2).

by rotations of n • 90° (where *n* is an integer). In general, there is a biconditionality between continuous Noether symmetries and conserved quantities of the form

Continuous symmetry ≡ Conserved quantity[28]

More precisely, continuous *temporal* symmetry (in physical jargon 'time translation invariance') implies *energy* conservation (and vice versa); continuous *spatial* symmetry implies *momentum* conservation (and vice versa). One could also express this in terms of regularities[29] of nature: if the physical regularities are the same over time, then the amount of energy does not change over time; if they are the same across space, then the amount of momentum does not change over time. Since the above logical relation is a biconditional (expressed by ≡), the contrapositives are also true: non-symmetry implies non-conservation (and vice versa)[30].

Mathematically (see also section 5.2), Noether's theorem has the following entailments: when there is a single independent variable (time), it leads to the constancy (zero time derivative, that is, zero rate of change over time) of the corresponding quantity. When there are multiple independent variables (time plus 1-3 spatial dimensions), Noether's theorem leads to the continuity equation (1) specified above, which relates the rate of change of the density of a quantity with the amount of the quantity 'spewing out' (the divergence of the current density). In other words, if a physical system's spatial relations can be disregarded (spatial variables omitted), then energy and momentum are conserved insofar as time translation invariance is given; if spatial relations must be taken into account, then the Noether theorem implies an equation by which the change of the corresponding quantity in the system can be calculated.

The upshot is that P2 stands on shaky ground to an extent to which it becomes useless as a premise. It is far from clear that the total amount of energy in the universe is constant, and even if it were, the conditionality of energy conservation allows for an increase or decrease without violating any principle or law. Now one might ask, why bother investigating replies aimed at refuting P1, given that we could show that the argument does not work due to the failure of P2? The answer is that one can thus come to see that there is really no alternative to what I will call the 'conditionality response' (chapter 7); in other words, the approaches canvassed in chapter 5 are not alternative ways for the interactive dualist to defend his position, they are no ways at all. Chapter 5 will be concerned with proposals in the domain of classical physics, while chapter 6 deals with the idea that a certain interpretation of quantum mechanics makes interactive dualism conservation-friendly. In chapter 7, finally, I will explain what I take to be the only solid answer to the OCP: the 'conditionality response'.

---

[28] The biconditionality corresponds to the conjunction "(Continuous symmetry ⊃ conserved quantity)•(Continuous quantity ⊃ continuous symmetry)".

[29] Most people would rather use the term 'laws' here. I prefer the more neutral term 'regularities', at any rate before the (in my view) metaphysically laden notion of 'laws' of nature has been sufficiently discussed (which will happen in part V).

[30] See the truth tables of biconditionals.

5. Unsuccessful Responses to the Objection from Conservation Principles I: Classical Physics

All the following approaches have in common that they try to circumvent premise 1 of the formalized OCP argument in chapter 4: the claim that interactive dualism entails change of the energy (or momentum) of physical systems.

5.1 Redistribution of quantities

The first proposal under discussion here is that non-physical minds might be able to act on physical matter by *redistributing* energy or momentum (henceforth summarily referred to as 'quantities') *without adding or subtracting* of those quantities. Thus, the hope goes, immaterial minds could bring about physical effects without violating the conservation principles. Such an account is defended among others by (Broad 1937, 109; Dilley 2004, 142; Meixner 2008, 18; Gibb 2010). For example, Broad suggests that the mind could

> …determine that at a given moment so much energy shall change from the chemical form to the form of bodily movement; and they determine this, so far as we can see, without altering the total amount of energy in the physical world. (1937, 109)

Broad, too, invokes a global view of energy conservation. Still, his proposal might work for energy if instead of the whole physical world one takes a human body as the system under consideration: if indeed all the energy for "bodily movement", or, more to our point, certain cerebral activities, stems from chemical compounds in the cells, then energy is indeed conserved. What Broad overlooks, though, is that in order to transform chemical energy into another form of energy a so-called 'threshold energy' is required, which lifts the compound energetically above its stable state into an unstable state from where it releases some of its internal energy by transforming into a different substance; in other words, the chemical substance is 'heaved over the threshold'. Where does the threshold energy come from? If it comes from the mind, then we have an injection of energy unaccounted for, since we cannot define a physical system such that its energy balance remains zero, because the mind is non-physical. Of course, in cells, enzymes are those molecules which lower the threshold energy so much so that the cell temperature suffices for the process to take place. In the present case, however, we are dealing with processes that are under the volitional control of the mind, not processes of the kind that happens all the time without our conscious contribution. Even if it is an enzyme that explains the triggering of the energy transformation, it must have been the mind that 'sent' it there, which raises the question whether the enzyme's energy and momentum are conserved. In defense of Broad, we might suppose that the enzyme had a certain (kinetic) energy amount all along, and that the mind just altered its direction; in this case, energy would be conserved. What seems so innocuous is, however, in fact a breach of linear momentum conservation. Linear momentum is momentum ($p = mv$) with a vectorial direction, such that even a change of the vector (the direction) alone would constitute an alteration of linear momentum; however, if that alteration is brought about by the mind, then again we cannot define a physical system whose momentum balance remains zero. One can capture this even more compellingly in mathematical terms: as soon as the Euler-Lagrange equation of a system contains a term that explicitly depends on time (i.e. is not time-invariant), then the corresponding quantity cannot be conserved, and this is so when something non-physical interferes. We will go through an exemplary calculation in the next section.

The bottom line is that the proposal simply fails to do what it set out to do, namely provide an account of how the mind can interact with the body without altering either conserved quantity: conserving only energy just isn't enough.

Even apart from that, there is a difficulty with the redistribution *mechanism* which is encapsulated in Ducasse's version of the redistribution approach:

> [I]t might be the case that whenever a given amount of energy vanishes from, or emerges in, the physical world at one place, then an equal amount of energy respectively emerges in, or vanishes from, that world at another place. (Ducasse 1960, 89)

The "vanishing" and "emerging" of energy is tantamount to its 'teleportation'. If this 'teleportation' is to occur simultaneously, it runs into two problems at once: first, the problem that no transport can happen faster than the speed of light, and second, that simultaneity is relative according to relativistic physics[31]. It might of course, and more plausibly, occur at a finite subluminal speed. The 'teleportation' location must in this case not be too far away, for reasons of minimal time delay. Perhaps this is indeed an option, but, as shown, the account suffers from much more severe ailments.

On the whole, the redistribution approach, far from achieving its goal in a contentious way, not even reaches that goal to begin with. In addition, it raises questions about its metaphysical and physical mechanism, which are only in part answerable in a satisfactory way.

5.2 Alteration of constants

Another attempt to make interactive dualism conservation-friendly was made by E.J. Lowe. He hypothesizes the mind could act on the body by changing physical constants:

> According to this second line of thought, the mind exerts causal influence on the body not through the exercise of psychic (non-physical) *forces* of any sort, but through influencing the values of the so-called "constants" which feature in various *physical* force laws – for instance, by influencing (presumably only locally and to a vanishingly small degree) the value of the universal constant of gravitation G or the value of the charge on the electron. Thus it would turn out that these so-called "constants" are strictly speaking *variables*. (Lowe 1992, 270; italics in the original)

Lowe's idea is that if the mind were able to change certain physical constants, such as the gravitational constant G or the value of the electron charge e, it could influence the brain without adding to or subtracting energy from it. He admits that "postulating variability in such constants is in a sense at odds with the classical principle of the conservation of energy" (ibd.), but still prefers this to accepting the violation of energy conservation, because "it is not now being suggested that the mind has a power of creating energy ex nihilo or conversely annihilating it" (ibd.). This latter statement makes it clear that he believes that a change in physical constants does not result in energy being added or subtracted.

---

31 Relativity might fall one day in favor of absolute simultaneity (cf.(Horava 2009)), which would not have any bearing on the main argument, though, since conservation principles are violated regardless of the redistribution speed.

Unfortunately, Lowe's theory does not yield his desired result of energy conservation. To see this, consider the following simple toy calculation. It serves not only as a refutation of Lowe's proposal, but also as a generic model designed to make clear what is essential for energy conservation (both mathematically and physically), and why there is ultimately no hope to make the interaction of a non-physical mind energy-(or momentum-)conserving.

To do that, we need to consider some mathematical equations. (The mathematics is reduced to the absolute essentials and kept as understandable and succinct as possible). First, there is the Euler-Lagrange field equation (henceforth ELE), also known as an equation of motion, which is an alternative description of a physical system in classical mechanics that has the advantage of being independent of the choice of the coordinate system. The ELE follows from the principle of least action, which says roughly that the action of a physical system (a time integral of the Lagrangian) is as small as possible, given the beginning and end configurations. The Lagrangian is basically kinetic energy minus potential energy. The Euler-Lagrange equations are second-order partial differential equations whose solutions are the functions for which a given functional, the action, is stationary, i.e., the system's action is "least" or perhaps "most" or at any rate unchanged by a small change in the dynamical variables.

As a representative model to examine Lowe's proposal I will consider a massive scalar field $\phi(t,x)$ (one-dimensional for ease of calculation)[32]. A scalar field consists of numbers assigned to space-time points, one number for each point; it is called a *scalar* field because the number is the same in every coordinate system. In the example calculation, this field stands in for all the usual physical fields (electromagnetism, gravity, the weak and strong nuclear forces, the electron field, neutrino fields, etc., nearly all of which are more complicated than scalar fields), while the mass parameter k (mass to the square) stands in for whatever physical constants Lowe envisages as varying due to mental influence. For a local field theory (Goldstein 1980, ch. 12), the Lagrangian L is given by adding up (integrating) the Lagrangian density $\mathcal{L}$ over all points of space (here represented by x because there is assumed to be only one spatial dimension):

(1) $$L = \int_{-\infty}^{+\infty} dx\, \mathcal{L}$$

The Lagrangian density for this field is

(2) $$\mathcal{L} = \frac{1}{2}\dot{\phi}^2 - \frac{1}{2}\phi'^2 - \frac{1}{2}k(t,x)\phi^2$$

with $\dot{\phi}$ being the partial time derivative $\partial\phi/\partial t$ (describing how ϕ changes over time at constant location x), and ϕ'being the spatial derivative $\partial\phi/\partial x$ (describing how ϕ changes from place to place at constant

---

32 Most realistic fields are more complicated; for example, the electromagnetic field comes from a 4-component potential, leaving one with 4 fields (the values of which depend on the coordinate system), as well as four independent variables t, x, y, and z in real 3-dimensional space. The toy calculation presented includes everything relevant to understanding the more realistic calculation with more fields and more spatial dimensions, while omitting a great deal of irrelevant complexity.

time t). Note that the 'mass' of the field k is not constant, but varies over place and time. This is crucial to mathematically capture Lowe's alteration of a constant by the mind.
One uses this Lagrangian density to infer the equation of motion or field equation for this field, i.e. the ELE for that field. The ELE in some cases takes the form $-ma + F = 0$ and thus can be a mere rearrangement of Newton's second law, but the Lagrangian formalism is more general in some respects and thus the standard starting point for fundamental physics. One can instantly see from this Lagrangian that energy is not conserved in this case, because the Lagrangian explicitly depends on time due to $k(t, x)$. The details of the calculation are useful, however, in making as accessible and simple as possible the mathematics-physics that is usually missing in discussions of conservation principles in the philosophy of mind.
Now, the ELE for the above system is

(3) $$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\phi}} - \frac{d}{dx}\frac{\partial \mathcal{L}}{\partial \phi'} = 0$$

With the Lagrangian density from above substituted, the ELE ends up reading

(4) $$-k(t, x)\phi - \ddot{\phi} + \phi'' = 0$$

For energy conservation to hold in such a system, the following continuity equation must hold (Goldstein 1980, ch. 12):

(5) $$\frac{d}{dt}\theta_0^0 + \frac{d}{dx}\theta_0^1 = 0$$

θ_0^0 and θ_0^1 are components from the canonical energy-momentum tensor (the energy density and energy flux density, respectively):

(6) $$\theta_\nu^\mu = \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial \phi}{\partial x^\mu}\right)}\frac{\partial \phi}{\partial x^\nu} - \delta_\nu^\mu \mathcal{L}$$

with μ and ν both ranging from 0 to 1 and δ (the Kronecker delta) being 0 if $\mu \neq \nu$ and 1 if $\mu = \nu$. (It should be added that $x^0$is identical to t (time) and $x^1$to x (the only spatial coordinate). Thus we obtain the energy density

(7) $$\theta_0^0 = \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial \phi}{\partial t}\right)}\frac{\partial \phi}{\partial t} - \mathcal{L} = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\phi'^2 + \frac{1}{2}k(t, x)\phi^2$$

and the energy flux density

(8) $$\theta_0^1 = \frac{\partial \mathcal{L}}{\partial\left(\frac{\partial \phi}{\partial x}\right)}\frac{\partial \phi}{\partial t} - 0 = -\phi'\dot{\phi}$$

If one plugs the above expressions (7) and (8) into (5), one gets:

$$\ddot{\phi}\dot{\phi} + \boldsymbol{\dot{\phi}'\phi'} + \frac{1}{2}\left(\frac{d}{dt}k(t,x)\phi^2 + k(t,x)\dot{\phi}\phi\right) - \phi''\dot{\phi} - \boldsymbol{\dot{\phi}'\phi'} = 0$$

The terms in bold cancel out, and application of the ELE (3) yields

$$\left(-k(t,x)\phi - \ddot{\phi} + \phi''\right)\left(-\dot{\phi}\right) + \frac{1}{2}\frac{d}{dt}k(t,x)\phi^2 = 0$$
$$0 + \frac{1}{2}\frac{d}{dt}k(t,x)\phi^2 \neq 0$$

In other words, the continuity equation (5) does not hold. Hence, energy is not conserved in this system! The crux is of course precisely the one term which represents Lowe's "alterable constant", namely in this case $k(t,x)$. It does not become zero or cancel out during the calculations.

Might at least momentum be conserved, as Lowe seems to assume (1992, 268)? Let's do the calculations. Momentum conservation can be checked by the corresponding continuity equation

(9) $\frac{d}{dt}\theta_1^0 + \frac{d}{dx}\theta_1^1 = 0$

Running the above calculations on (9) yields

$$\left(-k(t,x)\phi - \ddot{\phi} + \phi''\right)(-\phi') + \frac{1}{2}\frac{d}{dx}k(t,x)\phi^2 = 0$$

which, after use of (3), results in

$$0 + \frac{1}{2}\frac{d}{dx}k(t,x)\phi^2 \neq 0$$

Thus, momentum is not conserved either. On the other hand, by the above calculation one sees (*mutatis mutandis*) that whenever and wherever the 'constant' is really constant, energy and momentum are conserved, even if conservation does not hold in some regions, such as in brains.

So, Lowe's proposal fails because the alteration of a constant results in energy *and* momentum being not conserved. This result can be obtained with other model systems as well (see e.g. (Pitts 2019a) for similar calculations with a particle in a one-dimensional gravitational field). In fact, one can read off the conservation of energy/momentum from the Lagrangian (density): If the Lagrangian explicitly depends on time, energy is not conserved, and if it explicitly depends on place, momentum is not conserved (and conversely, only if the Lagrangian does *not* depend explicitly on time or place, then energy and/or momentum are conserved). One could even, somewhat uncharitably, conclude that Lowe's proposal is after all tainted by what it tried to avoid, namely "a disagreeable air of hocus pocus" (Lowe 1992, 270), because it not only posits something as quixotic as the alteration of physical constants but also cannot avoid non-conservation. More charitably, Lowe's theory merely fails as an account of mind-brain-interaction *which upholds energy and momentum conservation in the brain* but might still succeed as an account of mind-brain-interaction *which accepts that energy and momentum are not conserved inside brains.* Such a type of account is less outrageous than it might see at first glance. It will be discussed in chapter 7 under the heading of 'conditionality responses' to the objection from conservation principles.

5.3 Nothing to ruin? Interactive dualism and general relativity

The above arguments all assumed that there is a meaningful way of expressing energy locally. There are, however, worries that this is not possible. The claim is that general relativity already excludes conservation principles for energy and momentum, so that there is nothing left for interactive dualism to 'ruin'. Such a view is taken by (Mohrhoff 1997) and (Collins 2008; 2011b). Collins writes (PEC is Collins's abbreviation for 'principle of energy conservation')[33]:

> General relativity (GR) presents a major problem for the EC objection. The problem is that no local concept of stress-energy (and hence energy-momentum can be defined for the gravitational field in GR. Consequently, PEC does not apply in GR except in the very special circumstance mentioned below. This implies that although gravitational fields and waves clearly causally influence material objects, their influence cannot be understood in terms of movement of energy through space. (Collins 2008, 36)

To underscore his point, Collins quotes physicist Robert Wald:

> In general relativity there exists no meaningful local expression for gravitational stress-energy and thus there is no meaningful local energy conservation law which leads to a statement of energy conservation. (Wald 1984, 70, n. 6)

Against this, physicist and philosopher J. Brian Pitts argues that general relativity at least formally does have conservation laws, and that it simply has been difficult to find a reasonable physical interpretation of these relations. On the whole, Pitts argues that ultimately GR, if anything, might make it harder for immaterial minds to act on bodies (Pitts 2019b). I'll leave the adjudication between those competing positions to the physicists, of whom a majority believes that energy conservation fails in GR, according to Sean Carroll[34]. In any case, nothing hinges upon the success of the GR strategy. As we will see (chapter 7), dualists can avail themselves of a response far superior to any of the approaches discussed in this chapter.

5.4 Ascribing energy to the mind

A last proposed way of constructing energy-conserving interactive dualism discussed under the heading of classical physics is to ascribe to the mind the property of carrying physical quantities. The quantity in question is mostly energy, but as should by now be clear, no treatment of energy can ignore momentum; hence, I shall be speaking of 'quantities' as comprising both energy and momentum.

---

[33] The "very special circumstance" Collins refers to is a case "in which the region of space-time is asymptotically flat (…) An example would be a star surrounded by empty space in a universe with a flat space-time." (2008, 37). It seems clear that such a setting is hardly relevant for mental causation.

[34] http://www.preposterousuniverse.com/blog/2010/02/22/energy-is-not-conserved/, accessed Aug 15, 2022

There are basically two variants of this approach: attributing to the mind *physical* energy, or *non-physical* (mental, spiritual) energy. The idea is in both cases that the conservation of energy and/or momentum is guaranteed because the mind carries those quantities, and any increase/decrease of those quantities in physical objects the mind acts upon corresponds to a decrease/increase of those quantities in the mind. Let's begin with the first option: ascribing physical quantities to the mind.

It is important to realize what the attribution of physical properties to the mind entails, namely no less than the possibility of assigning to the mind (or its physical aspect) with a Lagrangian and a Euler-Lagrange equation (see section 5.2)! To my knowledge, the only philosopher who has seriously attempted to do something along these lines is Robin Collins. To be sure, his 'dual-aspect model of the soul' (Collins 2011a) is not designed to answer the objection from energy conservation (for which he offers one answer given in section 5.3 and another under discussion in chapter 6). His idea is that the soul (mind) has two kinds of properties, subjective (or mental) and non-subjective (or physical) properties, thus making the mind, on his definitions, a physical entity with additional subjective properties. He defines a physical entity as one whose "states can be described by some mathematical function" and the evolution of whose states and their interaction with other physical systems "can be specified by a set of mathematical equations" (Collins 2011a, 234).

What his theory needs beyond that, however, is a *link* between the mental and the physical properties of the mind. After all, physical events in the brain lead to mental qualia in the mind (e.g. in perception), and mental volitions in the mind cause neuronal firings in the brain (the hypothesis of interactive dualism). Collins proposes "linking laws" between mental states/qualia and physical states within the soul. For example, a certain vibrational pattern (brought about by some specific brain activity) always produces a certain mental state/quale, and vice versa. Without such linking laws, the mind could not cause anything in the physical world, because the theory has beforehand committed itself to explaining mental causation (in a conservation-friendly way) via the mind's physical properties. Note that this situation is different from the one envisaged by the heterogeneity objection: we are dealing here with causal (or correlational) links not between two substances, but *within* one substance.

Collins's theory is an interesting and sophisticated proposal, but it does not solve the problem; it just relegates it to another level. Collins correctly realizes that one important entailment of his dual-aspect model is that the mind's physical properties strictly follow the corresponding equations of motion. Now let the physical part of the mind be 'governed' by equations of motion; but what about the mental side? Its supposed law-like link to the physical part is unproblematic in the case of physical-to-mental causation, it even is an important requirement of perceptual states to be fully in line with incoming physical stimuli. When it comes to the reverse, however (mental-to-physical causation), things look differently. Are the mental events which cause the physical events subject to mathematizable physical regularities? This seems outlandish at the outset, since mind is construed non-physically on dualism, but let's leave this aside for the moment. If indeed mental events could figure in differential equations, then at least libertarian freedom, a strong motivation for interactive dualism (see section 1.2), is gone, for then there is no way for those events to occur at the behest of a free agent. If they cannot, then the account is not conservation-friendly, for any intervention from outside physics leads to non-conservation, as we saw clearly in section 5.2. It seems that Collins is in a dilemma: either give up freedom (and end up with a picture hardly distinguishable from epiphenomenalism) or miss one goal that, if not aimed at, is nonetheless an important motivation for a theory that ascribes physical properties to the mind. If the first horn of the dilemma is unacceptable, it seems that one either must look elsewhere for a conservation-friendly account of mental causation or 'bite the bullet' of a non-conserving account.

We've been discussing the first variant, that of ascribing *physical* quantities to the mind. What about the second one of construing it as carrying *non-physical* quantities which can be converted into physical quantities? (Hart 1994, 268) explicitly offers such an account:

> Energy (or mass-energy) is conserved, and conservation is a quantitative principle. So we need intrinsically psychological quantities. (…) Once we have such psychological quantities, we may imagine that as light from objects seen reaches the region of convergence along the disembodied person's lines of sight, it passes straight through but loses some electromagnetic energy and, at a fixed rate of conversion, that person acquires or is sustained in visual experience of those objects seen. (…) So we have solved the interaction problem. To be sure, we have not imagined exactly how light energy converts into the psychic energy implicit in visual experience. But then neither do physicists tell us how mass turns into energy when an atom bomb goes off, and if their lacuna does not embarrass them, neither need ours embarrass us.

Searle uses similar talk of mental energy (see chapter 4). He identifies the problem as mental energy being "not fixed by physics" and hence not figuring in the "energetic bookkeeping" of physics. As noted in there, Searle doesn't explicitly say that physical and mental energy are interconvertible, but such convertibility is necessary to get the objection (as well as the solution) going in the first place. For if mental energy never appears in the physical world, there is no problem to begin with; but if it appears, it must appear as physical energy, and so a conversion from mental into physical energy and vice versa must take place. One might think that this is a neat solution: ascribing to the mind mental equivalents of energy and momentum does not take on board all the ballast of commitment to physical quantities, and yet the quantitative 'bookkeeping' is secured. Is Hart's convertibility suggestion the panacea we've been looking for?

A first difficulty arises when scrutinizing the conceptual content of 'mental energy'. Energy (and momentum) are inherently quantitative and quantifiable notions: we can measure and calculate the energy and momentum states of physical systems, at least relative to one another. Mental states, by contrast, have no such quantifiability (apart from perhaps their countability): it simply makes no sense to quantify the strength of a belief, the intensity of a quale or to calculate the momentum of a volition. This immediately bears on the purported convertibility of mental and physical quantities. Which rate of change of a given belief or desire offsets a speed of 2 meters per second for a mass of one gram? It seems that absent a treatment of the mind itself (not just of its interaction with the brain!) in terms of Lagrangian field theory, we are unable to understand an ascription of energy to the mind in any sense relevant to the conservation objection; but the success of such a treatment would bring back all the problems discussed above of ascribing (partial) physicality to the mind.

A closely related, though perhaps minor problem is that the exchange rates must remain exactly constant in both directions to insure the viability of the convertibility solution. Otherwise, one Joule of physical energy might today be converted into one (fictional) 'mental joule', but tomorrow one mental Joule might end up as 1.2 Joule in the physical world. Though not impossible, the requirement of fixed conversion rates raises the question what ensures their fixation.

However, Hart still has one ace up his sleeve: yes, we do not know how mental energy is converted into physical energy, he admits; but then neither do we know how mass is turned into energy when an atomic bomb goes off. Our shoulder-shrugging acceptance of ignorance in the latter case justifies our putting up with our ignorance in the former case, or so his argument goes. This, however, is a premature conclusion.

For nuclear fission, physics gives a detailed quantitative trade-off of two forms of energy, mass-energy of some nuclei made of protons and neutrons, and electromagnetic energy, all described in terms of various particles that satisfy the relativistic equation $E = \sqrt{m^2c^4 + p^2c^2}$ (with m varying with the particle type and presumably being 0 for photons). To be sure, this quantification does not answer the question of *causal mechanism*, or *how* mass gets converted into energy; it just allows us to describe the micro-systems in question in precise mathematics and calculate their respective quantity after conversion. Thus, the atomic bomb case is thus as physically and mathematically intelligible as one can ask for. By contrast, the mental-to-physical quantity conversion still lacks all hallmarks of viability.

We've seen so far that all the above suggestions in the field of classical mechanics of how to make interactive dualism quantity-conserving fail. This, however, need not be the end of the story of conservation-friendly accounts of interactive dualism. The last decades have seen an influx of quantum mechanics-based approaches to mental causation that claim to respect the conservation principles. To these we will now turn.

6. Responses to the Objection from Conservation Principles II: Quantum Mechanics[35]

In the last three decades, those who seek to have interactive dualism without non-conservation have put the bulk of their hopes in quantum mechanics. It is indeed tempting to think that the (seeming) indeterminacy of quantum processes leaves room for minds to direct the way they want without bringing about energy or momentum non-conservation. I will argue in this chapter that those hopes are unfounded. This is no setback for interactive dualism, though, because, as I've shown in the previous two chapters, conservation is basically conditional.

6.1 How can quantum mechanics help make mental causation conservation-friendly?

Given the flood of publications using quantum mechanics (henceforth 'QM') to account for mental phenomena, it may be helpful to draw some distinctions first. First, QM-based theories are considered here *solely* in so far as they are concerned with dualistic mental *causation*. In particular, I am not interested in accounts that claim to explain *consciousness* quantum-mechanically (Penrose 1995; Hameroff and Penrose 2014; Tuszynski 2006). Second, in this section only the aspect of the compliance with conservation principles is investigated; other aspects of QM-related accounts of interactive dualism, such as their conceptual and empirical adequacy, are explored in subsequent sections. Third, I am interested in issues around the (philosophical) interpretation of QM only inasmuch as they concern the *facilitation of interactive dualism*; in effect, as will become clear, the only QM interpretation of interest, the only one that figures regularly in QM-based accounts of interactive dualism, is that of consciousness collapse (CC). As the name reveals, CC assumes a collapse process of quantum objects, something other interpretations deny. In fact, an interpretation other than CC may be the best one (in fact, I personally tend to prefer the Bohmian one); and clearly, if CC is not true, it cannot serve as a facilitation of interactive dualism, and advocates of other interpretations may therefore disregard CC from the outset. Even so, it is my job here to investigate the viability of CC, and of CC only, for interactive dualism.

Now, how is QM, or CC in particular, supposed to make dualistic mental causation conservation-friendly? The starting point is the collapse view of the measurement processes. Suppose an electron in a superposition of spin +½ / -½ is measured in a certain direction (say, along the x-axis). Quantum formalism as well as empirical investigations show that the outcomes of spin +½ and -½ occur with a probability of 50 % each; any prediction what the outcome of the next measurement will be has essentially the same epistemic certainty as predicting that the next coin toss will yield heads. The collapse view of QM is that this happens due to a 'collapse of the wave function': prior to the measurement, the Schrödinger wave function of the electron is in a superposition of spin +½ / -½, and the measurement somehow brings about a 'collapse' to one sharp spike, i.e. one of the possible eigenstates with the corresponding value +½ or -½.
Quantum theory is not restricted to the quantity of spin. Position and momentum are other quantities (in QM jargon: observables) which a quantum system can be in a superposition of. In fact, every definite state

[35] Large parts of this chapter have been published in my "Does Consciousness-Collapse Quantum Mechanics Facilitate Dualistic Mental Causation?", *Journal of Cognitive Science* 21:3, 429-474 (2020). Reprint courtesy of *Journal of Cognitive Science*.

(eigenstate) of the position observable is a superposition of the momentum states, and vice versa. The famous Heisenberg uncertainty relation quantitatively links momentum and position in such a way that the more exact the measurement of one of them is, the more uncertain the value of the other becomes, and that there is a lower bound of the product of position and momentum uncertainty[36]:

$$\Delta x \Delta p \geq \frac{\hbar}{2}$$

The idea is that the mind could make a difference in the physical world without violating any conservation principle (or, more generally, any regularity), if it could influence spin, momentum or position (or some other observable) within the given constraints, simply because there is merely a *probabilistic* (as opposed to a determinate) answer to the counterfactual question "What would have happened had the mind not interacted?" (We will see later that the probabilistic occurrence of the outcomes poses problems for QM-aided interactive dualism). For example, if a certain particle in the brain (or a brain area) were in a superposition of several position states, and if the mind could bring about a state reduction ('collapse') to one of those positions, then it could influence brain processes, provided the submicroscopic change results in a macroscopic effect. (We will see in due course that none of those antecedents is without difficulties). An important corollary of this mind-induced state-reduction is – or so it is hoped – that conservation of energy and/or momentum is respected, again because momentum (and thereby energy) may have a range of possible values prior to state reduction.

At any rate, we can sum up the view just described as 'sometimes-consciousness collapse' ('sometimes-CC'): the idea that the conscious mind *sometimes* collapses (state-reduces) quantum superpositions in the brain, namely exactly where and when it wills a bodily action (or perhaps a purely mental action such as directing one's attention to something, in so far as this requires interaction with the brain).

However, the theory that's been around for much longer is what I call 'always-CC'. 'Always-CC' expands mental influence to *all* quantum objects. In fact, it is one of the first interpretations of QM. As early as 1932, John von Neumann (Von Neumann 1932) first pointed out that the strict application of the quantum formalism entails that conscious observers end up in a superposed, indeterminate state when observing a superposed quantum object; but no conscious observer ever seems to have experienced such a state of 'superposed phenomenal experience[37]'. This dilemma is the so-called measurement problem. Eugene Wigner (1967)[38] took up Von Neumann's thread and claimed that it is our *non-physical consciousness* (note: I shall henceforth use 'mind' and 'consciousness' interchangeably) that brings about the state reduction (= collapse); and non-physical it must be, lest it be entangled with the superposition of the observed physical object and become itself superposed. Let us call this view 'always-CC', since it claims that whenever a state reduction takes place, it was a non-physical mind that brought it about.

---

[36] The ontological status of the uncertainties is unclear. Some hold that there is as a matter of fact no definite pre-measurement value, while others think that it's just a lower bound on the statistical standard deviations of position and momentum measurements. One example is Jean Bricmont, who as a Bohmian believes that there is a matter of fact concerning particle position, whether influenced by the measurement or not (e.g. (D. Lewis 1987; Hall 2009), 42-44). Bricmont's position seems more conclusive to me, since he offers the 'no hidden variables theorem' (which he also proves) in its support. However, I will, for the sake of argument, assume genuine (ontological) indeterminacy for the observable-dyad of position and momentum, since only this view is conducive to quantity-conserving CC.

[37] I speak of 'phenomenal experience' rather than of 'perceptions', because perception is a success term and should be reserved for those instances of phenomenal experience where the experiences forms a perceptual belief.

[38] The idea has recently been picked up and presented in a well-structured and accessible way by Hans Halvorson (Halvorson 2011).

It is important to note that always-CC is designed to explain phenomenal experiences, not actions. Even so, if always-CC is correct, then actions are no less caused by the state-reducing influence of mind than phenomenal experiences, on the plausible assumption that the physical structures involved in action (most notably the brain) are subject to the laws of QM. If true, always-CC would, besides its explanation of definite phenomenal experiences and conservation friendly mental causation, constitute a deep connection between mind and matter. In fact, defenders of always-CC claim that QM is at best a pragmatic, unsatisfactory piecemeal theory which can only be made complete by incorporating dualistic mental influence. However, there are other ways to solve the measurement problem, which will, as announced, not be part of the present discussion.

6.2 Does CC respect the conservation principles?

Let us now turn to our primary interest: the question whether CC can give us conservation-respecting mental causation. I will show that apart from a small loophole, all the evidence suggests that it cannot.

Take the position wavefunction $\Psi(\mathrm{x})$ of a spinless particle, with just one spatial dimension for simplicity. It indirectly represents (via $|\Psi^2(x)|$) the probability distribution of the particle being in places along the x-axis. This is nothing but a superposition of positions; standard QM has it that it is a genuinely ontological matter that the particle is not in one precise place, but just has a certain probability $P(x)$ of being there (see fig. 6.1).

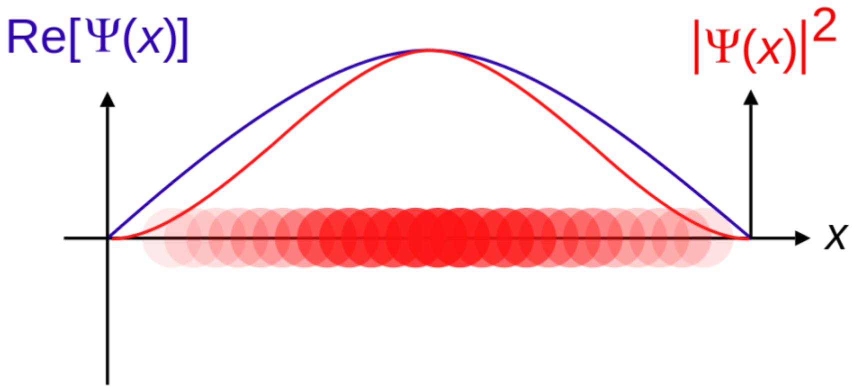

Fig. 6.1: Standing wave for a non-relativistic single particle in a box. The blue curve depicts the real part of the wave function, and the red curve the squared wave function, representing the probability density of particle position. This is also illustrated by the particle shades below – the more opaque, the higher the probability. (Image adapted from: Maschen / Wikimedia / Public Domain)

Now suppose a collapse of the wave function occurs. Immediately after the collapse, the $|\Psi^2(x)|$) (position probability distribution) curve looks like in fig. 6.2:

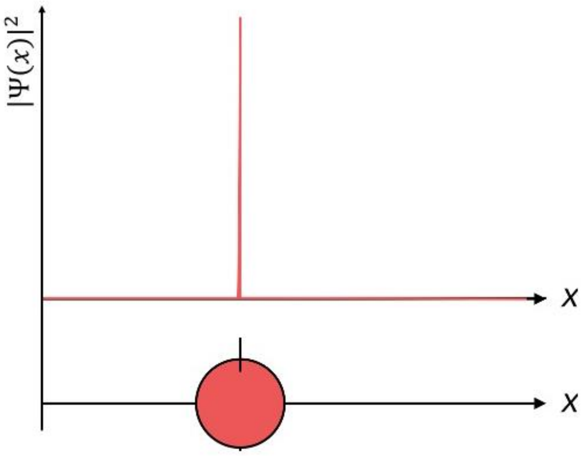

Fig. 6.2: Squared wave function after a position collapse. The particle is perfectly localized, represented by the spike. (Image adapted from: Maschen / Wikimedia / Public Domain)

This means that the position of the particle has now become definite. But via the uncertainty relation, a decrease in position uncertainty entails an increase in momentum uncertainty (and, since momentum ($p = mv$) and (kinetic) energy ($e_{kin} = \frac{1}{2}mv^2$) share the same basic quantities, an increase in kinetic energy uncertainty).

It is not immediately clear what to make of this. All we've established so far is that a collapse increases either energy-momentum or position *uncertainty*. To understand better what a collapse does to energy-momentum conservation, let us consider a particle in a superposition of energy eigenstates, say $c_1|E_1\rangle + c_2|E_2\rangle$. The Schrödinger dynamics will conserve the modulus squares $|c_i|^2$ and thus the (pre-measurement) expectation value $\langle\Psi|H|\Psi\rangle = |c_1|^2E_1 + |c_i|^2E_2$. However, what if a collapse occurs? Then, the system will collapse onto either $|E_1\rangle$or $|E_2\rangle$, and the possible values are E_1 or E_2, and energy seems not to be conserved. Carroll and Lodman (Carroll and Lodman 2021), assuming non-conservation for quantum collapses to be true, even claim that such non-conservation events should in principle be measurable, because the change in energy is not compensated by the measuring apparatus or environment. Several replies are possible. One might hold that the particle had no pre-measurement energy-momentum value at all. However, this anti-realist view looks more like 'ostrich tactics' designed to eschew any ontological commitment. If a plausible scientific realist view can be found, I urge that it should be preferred. And indeed, there are at least four such views. First, one might assume that the system had either E_1 or E_2 all along, and that the measurement just brought forward the true value. The problem is that such definite pre-measurement values come into conflict with the no-hidden-variables theorem (see fn. 36). This theorem does not rule out a fixed pre-measurement value for *all* quantities; one may be held fixed. However, that quantity is usually position (at least in Bohmian mechanics). Perhaps, inasmuch as energy may serve as the fixed quantity, the collapse process can be made energy-conserving; however, since position is then 'smeared out', linear momentum cannot be conserved, and such a trade-off gets us no further on the big picture.

Second, it might be held that conservation laws apply to closed systems only, and since a measurement entails that the measured system isn't closed any longer, energy conservation need not hold for that system. But of course, open systems exchange energy with their environment, and energy conservation can be formulated via the energy density in- and outflux across the system's borders (see section 5.1 and (Pitts

2019a)). Hence, energy conservation either holds if the continuity equation yields zero, or it does not hold, which is the case if a non-physical mind acts on the system (see section 5.2).

A third possibility is to claim that energy is conserved on the 'ensemble' level. The idea is that if one measured a large number of identically prepared systems, one would get as a mean value (at least approximately) $|c_1|^2 E_1 + |c_i|^2 E_2$, which corresponds to the expectation value $\langle\Psi|H|\Psi\rangle$.

In the following, I shall work with that third understanding of energy conservation in quantum systems. Gao (2017) offers an account along those lines. According to it, energy is conserved at the 'ensemble' level, i.e. at the level of an ensemble of identically prepared systems. Such conservation obtains if "the probability distribution of energy eigenvalues [manifests] itself through the collapse results for an ensemble of identically prepared systems. This means that the diagonal density matrix elements for the ensemble should be precisely the same as the initial probability distribution at every step of the evolution" (Gao 2017, 136); in other words, if no quantum information is lost throughout evolution and collapse. However, in Gao's theory, the collapses and their frequency are explained wholly physically, without any recourse to minds. Does *mind-induced collapse* conserve energy at the ensemble level?

To answer this question, it is helpful to examine another mind-independent collapse theory, the GRW[39] theory. Roughly, on GRW, wave function collapses occur stochastically; also, the particle localization is 'imperfect' (the position spike still has a certain width around a collapse center) and obeys a Gaussian distribution (concerning the position of the collapse center). What is clear is that collapses in GRW do not conserve energy-momentum at the ensemble level. This is primarily because energy superpositions do not collapse to eigenstates. Now, is CC more akin to GRW or to Gao's account? This is not clear. It might be that consciousness-induced collapses conserve energy-momentum at the ensemble level, although this would be an additional constraint to mental action besides the Born rule. It might also be that the collapses are more of a GRW type (i.e., failing to conserve energy at the ensemble level) and thus plausibly preserve complete freedom for mental interaction. Therefore, on the understanding of energy conservation endorsed here, we seemingly have reached a dilemma that can only be solved through a trade-off: either sacrifice *complete* freedom of action or energy-momentum conservation. Of course, this dilemma might not arise if one of the other understandings of quantum energy conservation are true. At the same time, we probably need not worry about any such dilemma, or even if there indeed is one, I urge that we can solve it by going for the latter option: accept violations of conservation laws[40]. This option will be discussed in the next chapter.

There is, however, yet another, though much less known, proposal how QM could make interactive dualism conservation-respecting, though it is not entirely clear how it is supposed to do so. To my knowledge, Robin Collins is the only philosopher having proffered this suggestion (Collins 2008, sect. V). It picks up from the observation that in QM there are correlations (or even causal connections) without energy exchange, in particular Einstein-Podolsky-Rosen (EPR) correlations. In EPR correlations, two space-like separated sequences of events (e.g. entangled photons measured by two different sensors an arbitrary distance apart; let's call them A and B) are correlated in law-like ways that cannot, as John Bell

[39] Acronym referring to its founders, the Italian physicists Ghirardi, Rimini and Weber.

[40] A serious worry, though, is that the parameters of the 'spontaneous collapses' in GRW need to be within certain confines in order to not lead to too much energy gain and thus to 'global warming' (see also (Feldmann and Tumulka 2012)). At any rate, there does not seem to be an in-principle problem for physicists if energy/momentum conservation didn't hold in spontaneous collapses.

showed, be explained in terms of a non-instantaneous common cause (Bell 1964; 1966). Nor can the A-events have caused the B-events or vice versa – unless the cause is considered to travel superluminally – because QM predicts that EPR correlations can occur between sequences of events separated by any arbitrary distance. A superluminal transfer (of energy or momentum) is ruled out, though, because according to special relativity, any transfer must occur at less than the speed of light. Lastly, QM predicts that EPR correlations occur without energy transfer. The EPR correlations can thus be taken to demonstrate that law-like correlations do not require an exchange of energy or momentum. Collins concludes that the EPR correlations show that "an interaction (or at least a correlation) between the mind and brain that does not involve an energy exchange, or any other mediating field, has precedent in current physics, thus severely weakening the EC objection" (Collins 2008, 39). Indeed, if EPR correlations can be treated as a precedent for mind-brain-interaction, this conclusion holds quite regardless of whether one adheres to causal realism or causal anti-realism. The causal realist believes that the correlation is grounded in an actual cause, which could in this case be a causal connection between A and B (ruled out by special relativity) or a common instantaneously acting cause; the causal anti-realist believes that the correlation is a brute fact and not grounded in any further facts. The former option models the classic interactive dualist contention that the mind is the cause of physical events in the body, albeit without energy and momentum transfer; the latter option simply contents itself with satisfying the counterfactual "Had the mind not acted, the events in question would not have happened" without going so far as to claim any metaphysically substantial causal connection (an example is Michael Esfeld's already mentioned Super-Humeanist theory of mind-body interaction).
As interesting as Collins's proposal is, some questions remain. First, EPR correlations take place between two physical objects[41], while the mind is wholly non-physical (unless one adopts Collins's other proposal to conceive of the mind as partly physical, which faces its own difficulties, see section 5.4). Can a non-physical mind truly be 'entangled' with any physical system? Or, less demanding, is an entangled quantum system sufficiently analogous to the mind-brain-relation to serve as a model? A problem in this context is that the ontology of entanglement is simply too poorly understood to warrant any sure statement beyond its mathematical structure. Second, even if this were possible, would the resulting mental effect be strong enough to trigger macrophysical events like neuron firings? We will look at this crucial empirical question in section 6.3, but the anticipated answer tends to be in the negative. Lastly, one must not forget that any entanglement between objects has a common cause (e.g., photons must *become* entangled before their space-like separation) which limits the degrees of freedom of the objects in question. In particular, the occurrence of measurement results will still obey the Born rule, a quite untoward consequence for one important motivation of ID, namely libertarian freedom (cf. section 8.3).

A last defense of the failure of QM-aided ID in respecting conservation principles is that although violations happen, they are too gentle to be detected. This could be the case if consciousness-collapses reduce the wave function in a 'smooth' GRW manner (in GRW the parameter of post-collapse peak width is chosen so as to not generate empirically disconfirmed 'global heating'). Thus, we may gain the impression that brains perfectly comply to the laws of nature, although, undetectable by us, non-physical minds cause irregular behavior. This may yet be the case (in fact I believe it to be so), but it completely misses the point:

[41] Mathematically, the two entangled quantum objects form one single system.

in dealing with failures of conservation principles/laws of nature, we deal with an *ontological* question, not an *epistemological* one. If conservation failures occur, no matter how small, the law-based argument against interactive dualism takes off, and the dualist needs to defend his position at the level of laws of nature in general, not at the level of specific conservation principles. It should also be noted that a GRW-type collapse leaves 'tails' in the wave function which create intra-theoretical and philosophical problems (cf. (D. Albert and Loewer 1988; McQueen 2015)).

Let's sum up. The interactive dualist's hope that mental interaction through CC conserves energy-momentum stands on shaky ground. A possible solution would be to construe quantum energy conservation to hold at the ensemble level (perhaps along the lines of Gao's account), but this almost certainly entails a significant restriction of the freedom of mental interaction. Here, we see another instance of a recurring theme that has already briefly appeared before (chapters 1-3), will accompany us throughout our discussion of the tension between interactive dualism and conservation principles, and will constitute the largest part of part IV: that of the relation of mental causation and laws of nature.

The other main line of QM-related responses to conservation-friendly is to claim that mind-brain-interaction is a correlation or causal connection that requires no energy or momentum transfer, taking EPR correlations as a precedent. Here, it is not clear that the analogy is sufficiently strong, besides empirical questions lurking in the background. Certainly, causal anti-realists like (Super-)Humeans can fall back to the less contentious claim that there are ever only correlations (and no causal connections), but then this approach could be expanded to all physical events, thus stripping EPR correlations of their distinctive appeal as physical precedence.

For now, given that the non-conserving nature of CC cannot be established conclusively, let's turn to some empirical difficulties for CC accounts of mental causation which, as a cumulative case, should make it clear that dualists ought not to rely on the 'bruised reed' of quantum mechanics, but rather on the more robust 'conditionality response' (chapter 7). Even more (conceptual) problems for CC-aided interactive dualism will appear in chapter 8.

6.3 Empirical problems

6.3.1 Energy-time uncertainty doesn't leave enough leeway

In one of the first serious attempts to show that QM-aided ID is viable, Beck & Eccles (1992) calculated that an object as heavy as six hydrogen masses could be moved by the mind within the confines of the momentum-position uncertainty relation, thereby respecting energy as well as momentum conservation. As indicated above, there are several positions as to what the uncertainty relation actually describes; but let us, for the sake of argument, assume that it captures an actual, ontological indeterminacy.

Eccles' and Beck's calculation is based on the equations of the quantum mechanical zero-point energy E_0 and the thermal energy E_{th} of the model system. The idea is that if $E_0 \gg E_{th}$, the particle is in the 'quantal regime' (with quantum effects dominating over classical effects), whereas it is in the 'thermal regime' (with classical effects preponderating) if $E_{th} \gg E_0$; finding the borderline between both regimes would then be tantamount to obtaining the critical mass a particle can maximally have while falling under the quantal regime. With $T = 300K$ and $\Delta q \cong 1$ Å the borderline equation $E_0 = E_{th}$ yields a critical mass $m_c \cong 6m_H$. And indeed, if the mind could shift around a particle that heavy without violating energy

conservation, this would support interactive dualism. Beck and Eccles consider the movement of a hydrogen bridge sufficient to increase the probability of exocytosis of a presynaptic vesicle (PVC), which is what triggers an excitatory postsynaptic potential (EPSP, also called action potential) in the next neuron. However, as we've seen above, energy-momentum conservation is in general not guaranteed for quantum collapse processes such as Beck and Eccles countenance. Furthermore, Wilson (1999) has contested Eccles' and Beck's quantitative claims. His calculations, based on the time-energy uncertainty relation, yield a time period of merely $\Delta t < 3.2 \times 10^{-15} s$ during which the breaking of one single hydrogen bond could go undetected[42] of – much too short for any relevant neurophysiological process (like ions passing through a channel) to happen. (Those processes are rather in the order of milliseconds ($10^{-3} s$)). Conversely, if one picked a reasonable (though conservative) time period like $10\ ms$, one ends up with an 'energy leeway' of $5.2 * 10^{-30} J$, which is 200,000 times smaller than a Van der Waals bond and thus much too little (see (Clarke 2014)).
Beck (1996) has presented a slightly enhanced account, where he holds on to the 1992 result but additionally calculates a signal time for quantal processes of about $30 * 10^{-12} s$, which comes relatively close to Wilson's result, and is well in the order of recent calculations of ion decoherence times in protein channels (V. Salari et al. 2011; Summhammer, Salari, and Bernroider 2012; Salari et al. 2015; Summhammer, Sulyok, and Bernroider 2018). In fact, the just-cited 2012, 2015 and 2018 papers even argue that decoherence times in the order of picoseconds are beneficial for ion transmission through protein pores, and that the transmission dynamics could not be explained without recourse to a quantum model of ions. (The import of this for dualistic mental causation will be discussed in section 6.3.3). All in all, it seems that Wilson's result stands on firmer ground. Apart from the uncertainty equation, he uses only one quite conservative assumption (one hydrogen bond[43]), while Eccles and Beck work with the additional notions of thermal and quantal regime. I'll leave the final verdict to the physicists. Perhaps, one more plus on Wilson's side is that his equation directly targets the more relevant quantity of energy (momentum isn't conserved either way). At any rate, the least that can be said about Eccles' and Beck's proposal is that it is not watertight.

Recently, Georgiev and Glazebrook (Georgiev and Glazebrook 2014) have undertaken an attempt to salvage Eccles's and Beck's basic idea. By their time, it was already established knowledge that vesicular release of neurotransmitters into the synaptic cleft is an exocytotic, protein-mediated process (Jahn and Fasshauer 2012) – not the product of the state change of a metastable paracrystalline vesicular grid, as Eccles assumed (Beck and Eccles 1992; Eccles 1994). The main protein complex involved in vesicular exocytosis is called the SNARE complex. Glazebrook and Georgiev hypothesize that a 'quasi-quantum particle' – a soliton – might propagate along the 4-alpha-helix-bundle of the SNARE complex and thus trigger exocytosis which is foundational for neuron firing. The posited soliton is assumed to have 5% hydrogen mass, which would allow it to tunnel through potential barriers of 1-2 nm width, which seems

[42] No matter what further ontological conclusions one draws from the uncertainty relation, it is clear that it denotes a limit of detectability.
[43] The smaller the energy, the larger the time interval. One could replace the hydrogen bond with an even weaker bond type, the Van der Waals force, which has a bonding energy of 0,5-5 kJ/mol (H bond: 20 kJ/mol). But this would increase the time frame by 4 to 40 times, which is still by far not enough.

to correspond well with the conformational situation in a SNARE complex. As support for their hypothesis, they offer three arguments:

- Solitons are not a quixotic idiosyncrasy of theirs. A so-called Davydov soliton (Davydov 1973; 1977) has been posited as an explanation for the actin-myosin interactions in muscles.
- Their model is temperature-dependent, which is also true of SNARE proteins. The model thus looks more realistic than electron tunneling (apparently proposed by Beck 1996) which is temperature-independent.
- Glazebrook and Georgiev claim that 'vibrationally assisted tunneling' (under which the soliton action would fall) "has been experimentally proven for the action of a number of enzymes with dehydrogenase activity" (Glazebrook & Georgiev 2014, 20).
- The hypothesis is experimentally testable: replacing hydrogen atoms in SNARE proteins with deuterium atoms[44] should have a clear effect on the probability of synaptic vesicle release.

This certainly looks like an interesting approach. Further work is needed to make the soliton hypothesis more robust. At any rate, even if it succeeds, the other problems with quantum theory outlined above remain. But perhaps, after all, what we need is just a locus in the brain where gentle, possibly undetectable interactions can take place, and quantum theory with its restrictions and strict formalism is more ballast than help.

6.3.2 Amplification of quantum effects is doubtful

> Thus, to use Jeffrey Koperski's vivid example, utilizing chaotic systems might allow God to 'alter the arrangement of bubbles in the crest of a tsunami but not redirect its course'. (Larmer 2008, 155)

As we've seen, the uncertainty relations allow (if anything) only for quantum collapse effects which by themselves seem much too weak to make any difference in the brain. It has been suggested, therefore, that those tiny leaps might be amplified by (deterministic!) chaos in the brain (Hobbs 1991; Hong 2003; Kane 1996; King 1991). The basic reasoning seems to be this: quantum systems behave probabilistically; there are chaotic macroscopic systems governed by deterministic equations but which are sensitive to initial conditions (Hobbs 1991, 143); hence, a good explanation for chaotic systems is that the initial conditions are at least in part constituted by quantum events. In fact, there is little doubt that chaotic behavior actually occurs in the brain (Battaglia and Hansel 2011; Korn and Faure 2003; Tsuda and Fujii 2007). There are, however, some major problems with the combination of quantum events and chaos.

First, the very existence of 'quantum chaos' is doubtful (Bishop 2008).This has to do with the Schrödinger equation, which mathematically predicts 'quantum suppression of chaos' (Hobbs 1991; Koperski 2000). A more general problem is that (in fact irrespective of the means of amplification) one needs to be able to specify precise initial conditions. At least on some interpretations of QM (in particular, the very collapse theories under discussion here), the notion of precise pre-measurement conditions is void of meaning. Third, a chaos-amplified quantum-ID would have to translate the directed collapses (whose plausibility is debatable, see above) at the quantum level into macroscopic effects – whether this is plausible is at least unclear. Fourth, the noise-resistance of the brain constitutes a further difficulty. Brain cells, like cells in

[44] Deuterium (^2_1H) is a heavy isotope of hydrogen (^1_1H).

general, are quite resistant to thermal fluctuations, which are about one billion times stronger than quantum fluctuations (Clarke 2014, 115). Hence, *a fortiori* brain cells should be resistant to quantum fluctuations. If, by contrast, they're not, it then seems that thermal fluctuations will be amplified as well, something we do not observe. The solution is of course to posit low-energy quantum events in rapid succession (like (Beck 1996; 2001; 2008), a proposal whose problems we have already seen.

6.3.3 Do we have empirical evidence for relevant quantum effects in the brain?

Of course, all conceptual investigations of CC-ID are fruitless if there are *empirically* no relevant quantum collapse events in the brain which the mind could make use of. In many fields of biology, the existence of a host of such quantum effects, ranging from bacteria through the animal kingdom to the human retina (see (Jedlicka 2017; Al-Khalili and McFadden 2015; Gehlert 2019)), has been proven. This raises the prior probability of their occurrence in human brains. Do we have evidence for them?

One central issue in this context is decoherence times. Decoherence is a (deterministic) process through which quantum information 'leaks' out of a system into its environment. A decoherence time is the time it takes for a quantum system to 'decohere', i.e. to lose its quantum information to the environment. If decoherence times are too short for a given biological structure in the brain, quantum effects seem be drowned in thermal noise and doomed to irrelevance. For about ten years, Max Tegmark's (2000) calculations concerning ion transmission through the cell membrane and microtubules were taken to be decisive. He arrived at decoherence times of 10^{-13} s to 10^{-20} s, which is thought to be far too short to make any difference to neuronal function. However, in the past decade, some new simulation studies (esp. Summhammer, Salari, and Bernroider 2012; Salari et al. 2015; Summhammer, Sulyok, and Bernroider 2018) suggest that (1) decoherence times for ions in protein channels are longer (in the order of 10^{-12} s), that (2) a quantum-mechanical behavior as well as quick decoherence are even beneficial for ion conduction through channels, to the extent that conduction dynamics could hardly be explained without those factors, and that (3) the quantum model even explains the selectivity (at high throughput!) of the channel.

Remains the question what to make of this. First and foremost, the model speaks of decoherence, not of collapse; these are different things. Second, the model is about a K^+ pore in bacterial cells, not in human neurons, so applicability to human Na^+/K^+ pumps would have to be shown. Third, even if the model could be transferred to human neurons, it is not clear where mental action should set in. At the level of the ion wave packet? This seems unnecessary, since according to the model, the ion 'sneaks' through the channel (Summhammer, Sulyok, and Bernroider 2018, 9) and 'needs no further help'. At the level of ion velocity, thereby increasing the portion of ions with sufficient kinetic energy? This is not only completely independent from the channeling model, but also re-imports the issue of energy conservation. At the level of the protein (constituents)? On the one hand, the ion quantum dynamics seem sufficient to explain their passage through the tunnel, and on the other hand, the plausibility of mental manipulation decreases with the size of the objects. Perhaps influencing single oxygen atoms (as in Summhammer et al.'s model) will suffice, but that is presently just speculation.

All in all, the recent advances in quantum biochemistry have raised the plausibility of quantum effects in the brain, but likewise left dualists will little to make of it: in an 'in-out' superposition setting (as in Tegmark 2000 and Beck and Eccles 1992), the mind can influence the outcome probability in one direction (problems with the Born rule notwithstanding). In the recent 'quantum tunneling' model, the ions basically help themselves, and no mental influence seems possible without betraying the very motivation of CC-ID.

6.3.4 Why no-collapse interpretations of QM don't help, and what the real issue is

I think by now it should be clear that CC does not give the interactive dualist a 'free lunch', but, to the contrary, creates more problems than it solves. On an ironic and sardonic note, Jeffrey Koperski tosses barbs at the unwarranted enthusiasm with which quantum approaches were welcomed in the literature about divine action (which is a philosophical question largely parallel to interactive dualism):

> In my view, there is a kind of mythology that appears in the divine action literature. Determinism and closure will be seen as ancient foes to be overcome if we believe in theism instead of mere deism. But good fortune! We have discovered the new magics of top-down causation, quantum mechanics, and chaos, which can slay the beast and open the gate a bit so that God might act in the world. (Koperski 2020, 124)

Still, the identified problems refer to the consciousness-collapse (CC) interpretation of QM only. One might ask whether other no-collapse interpretations are of any help. In reply, it must be noted that those interpretations do not even *pretend* to solve the problem of conserved quantities, while CC probably cannot solve it despite sustained efforts. In fact, one important no-collapse theory (Bohmian mechanics/BM) makes quantum dynamics wholly deterministic[45] and thus offers no more help for ID than classical mechanics. And that is not all: two other eminent no-collapse interpretations – the Many Minds Theory (D. Albert and Loewer 1988) and the Many Worlds Theory (Everett 1957) – require quite extravagant ontologies (the splitting of the universe into many universes/branches of the universe or the splitting of one observer into many observers/versions of that observer, respectively). I have no space here to treat those approaches adequately, but it is to be expected that they make life harder for the dualist who seeks conservation-friendly interaction. Of course, once one abandons the hope for conservative interaction (as I urge one should, see next chapter), the range of possibilities to metaphysically integrate mental causation and quantum mechanics becomes considerably larger; but an evaluation of these possibilities likewise lies beyond the scope of this thesis.

[45] There is a probabilistic element in BM, but it is located in the initial conditions of particles. The dynamics of particle movement is completely deterministic (see (Bricmont 2016b; Oldofredi et al. 2016)).

7. The Conditionality Response to the Objection from Conservation Principles

> Nature (…) makes no trouble at all. The whole system, far from being thrown out of gear (…) digests the new situation as easily as an elephant digests a drop of water. (C. S. Lewis 1960, ch. 15)

It should have become clear through the previous two chapters that there is an in-principle problem with making interactive dualism (abbreviated ID in this chapter) conservation-upholding. Why 'in-principle' problem? Recall that the conservation of quantities derives from the biconditionality of the first Noether theorem:

Continuous symmetry ≡ Conserved quantity

The Noether theorem does not require the physical world to consist of only symmetrical systems; instead, it makes the *conditional* claim that *if and only if* there is a (continuous) symmetry (with respect to time and place), *then* energy and momentum are conserved. The continuous symmetry criterion is also expressed mathematically by the temporal and spatial independence of constants in the Lagrangian of a physical system (see section 5.2). In any system on which a mind acts, continuous symmetry, as well as the time- and space-independence of constants would be lacking, precisely because the mind's influence varies over space and time. The solutions to this tension are to either make the mind's influence constant over space and time (as would be required to render valid Collins's and Mohrhoff's general relativity approach), or to declare the mind a physical system in its own right, such that the larger system comprising the system acted upon *and* the mind becomes conservative. Both solutions violate the 'prime mover'/libertarian spirit of ID.
Thus, the best an interactive dualist can do is give a *conditional response* to his detractor: energy is conserved so long as no mind acts on the physical world, i.e. in almost all places except our brains[46]. This is what I call the 'conditionality response' to the objection from conservation principles (see also (Cucu and Pitts 2019; Pitts 2019a)).
With this response, there is nothing left of the energy conservation objection. Consider again the objection in deductive form:

P1 If non-physical minds influenced the physical universe, then the energy of the physical universe would not remain constant (ex hypothesi).
P2 The energy of the physical universe is constant (with nomological necessity).
C Therefore, it is false (with nomological necessity) that non-physical minds influence the physical universe (P1, P2, modus tollens).

In effect, the conditionality response (via the Noether theorem) confirms P1 as much as it rejects P2 as insufficiently supported by physics and dialectically unavailable in this context. But without P2, there is no way the objection can even get off the ground. Ironically, it is exactly this premise which detractors of dualism have taken to be a cornerstone of physics and which many dualists have striven to uphold in order to respect science. But submitting to physics is a good idea only if one correctly describes what physics

[46] Further places where energy conservation might fail include: animal brains (if animals have immaterial souls); all places where God, angels and demons - if they exist – act on the physical world.

actually says. Otherwise one is doing the philosophy of A-level (secondary school) chemistry[47], not the philosophy of physics (to recall a warning by (Ladyman et al. 2007, 24). The upshot is that physicalists should not invoke conservation principles for an objection against ID, since they turn out to be a 'Trojan horse' for the physicalist's project.

In this chapter, I wish to dwell a little on the conditionality response, in order to prevent some potential misunderstandings, and to elucidate what to expect from a world in which conservation does not (always) hold.

Non-conservation might be very gentle

A second point closely related to the first one is that if there are non-conserving events in the brain, the non-conservation is arguably very gentle, perhaps undetectable. Plausibly, not much alteration of energy or momentum is required on the micro-level to bring about actions on the macro-level (for more on this see chapter 9). In fact, instances of strong non-conservation can virtually be ruled out, since they have not been discovered; this much can be granted to Papineau's contention that physiology has failed to find mental forces despite thorough and long-standing investigation. However, the possibility that non-conservation could be very weak should be borne in mind when making empirical claims about ID.

The conditionality response is compatible with QM, avoids interpretive issues

The third aspect relates to QM. The conditionality response is compatible with QM, but it doesn't need it. As we've seen (chapter 6), QM is barely of help in constructing a conservation-friendly theory. Thus, even if the mind acts through quantum events, the conditionality response will still be the main foothold of ID in the natural order. It might of course be that the mind-matter connection is more than accidental; defenders of the consciousness-collapse interpretation of QM think it is necessary. We will come to that in chapter 8.

By the same token, the conditionality response is independent of interpretive issues in QM, because it is valid even if a deterministic interpretation or one that rules out conserved quantities should turn out to be true.

Non-conservation is always local

Fourth, it follows from the way fundamental physics understands conservation principles that non-conservation will be local, not jeopardizing the regular working of the universe as a whole. Thus, if Susie decides to raise her arm and her mind acts on nerves in her brain to start the relevant causal chain, then the time- and space-dependence of her mind's influence (in her brain and not on the moon, during her lifetime and not before or after) implies the non-conservation of energy and momentum in her brain at that time, but it leaves energy conservation in her fridge, her car, all the planes that take off that day intact, as well as in supernovae millions of light years away. Bunge's bane is debunked: there is no need to fear the breakdown of physics and chemistry if minds bring about (gentle) instances of non-conservation in brains.

---

[47] Personal confession: For lack of better knowledge, I used the incorrect formulation of energy conservation when I taught chemistry to high school students. Whether or not Ladyman allows for A-level chemistry to use the wrong principle (while prompting philosophers to refer to the correct one), I personally would certainly try to bring across the conditionality of conserved quantities to students had I to teach the subject again, and would urge any science teacher to do likewise.

The conditionality response works independently of any account of a causal mechanism

Fifth, answering the question *how* the mind brings about the conservation-violating changes is not crucial. The conditionality response just states what would be the case if an immaterial mind interacted, and that this conditional is perfectly admissible physically. Nonetheless, of course, it does not hurt to try some answers. The imparting of *energy* is out of questions for reasons already discussed (section 5.4).

A better idea is to have the mind exert a *force* (Averill and Keating 1981). One might worry that this would make the mind unduly physical; then again it might be trivially true that a physical system yanked out of its natural trajectory must ipso facto be subject to a force. Once non-conservation is accepted, Lowe's proposal of the mind altering physical constants becomes available again, at least in principle; however, on closer inspection, and with its main motivation gone, it still looks rather unpromising. For, what exactly would the mind do when in alters physical constants? A physical constant is presumably not a thing out there that can be acted upon. Humeans will consider it as an abstraction supervening on the actual physical events; changing it means in this picture to simply change the subvenient events, which is essentially the basic claim of ID[48]. Dispositionalists, by contrast, (or all who accept intrinsic causal properties of matter) will rather want to identify physical constants as intrinsic properties of matter, so on this view, the mind would have to change some presumably deeply engrained intrinsic properties of physical objects.

One last proposal (of which I do not know whether it is defended by anyone) is that the mind creates energy ex nihilo. This is clearly different from ascribing the property of energy to the mind; something created ex nihilo by the mind is distinct from it and not a property of it. However, this approach suffers from at least two problems. First, energy alone does not suffice. Think of the simple example of a golf ball that you wish to get into a certain hole on the driving range. Simply infusing it with energy *tout court* is of no help; after all, the same amount of energy needed to get the ball into a hole 200 meters away might heat up the golf ball by 1.5 °C, but that is not what one wants. What is needed is of course *directed* energy, in other words, energy *and* linear momentum. Second, the notion of creating energy suggests that energy is some kind of 'stuff'. This seems rather implausible to me, but lacking the requisite expertise of such matters, I will not enter a discussion of this issue.

Finally, one might suppose that the mind simply has the power to make something of it (motion?) become present in the brain/body (see chapter 3). Alternatively, if one is a causal anti-realist (Humean) one does not even have bother oneself with such questions, since there is no 'how' of ID to account for. But again, the conditionality response works whether or not one can present an underlying causal account of mental interaction.

The conditionality response works better on some worldviews than on others

The last aspect I want to discuss in relation to the conditionality response is another instance of a theme that has already appeared several times in this thesis (especially in chapters 1 and 3): namely that a certain answer to some puzzles, though in itself robust and/or explanatorily powerful (here, the conditionality response), is rejected for tacit or outspoken reasons having to do with larger worldview considerations (here, as practically always, the reason of adhering to naturalism). Concretely, the manifestation of that

---

[48] Other Humeans, like Michael Esfeld, locate the mind's influence not in an alteration of the laws of nature (or of physical constants, for that matter), but in an alteration of the initial conditions at the beginning of the universe (see Esfeld 2020, 100f.). A detailed discussion of such accounts will ensue in part V.

theme in the case at hand could be formulated as follows: is it not odd, in a way that violates the razor principle[49], to have a universe which for the most part works strictly quantity-conservative, and only in vanishingly small pockets evinces non-conservation? The tacit assumption here seems to be that on naturalism, we should expect a uniform universe, not such a "dappled" one (to borrow a phrase from Nancy Cartwright (Cartwright 1999)). Another, closely related worry might be that a universe 'perforated' by mental interactions is unaesthetic. I will try to make plausible why, and on which conditions, the conditionality response makes perfect sense of our universe.

A first response to above worry is to incorporate non-conserving mental interaction into the naturalistic worldview. There are defenders of naturalistic substance dualism (e.g. (Nida-Rümelin 2010)), so why not of naturalistic ID? Support for such a view comes, perhaps somewhat surprisingly, from physics. Physics often and successfully deals with systems in which the equations have time- and/or space-dependence (and thus energy and/or momentum non-conservation), so why not have one more example? It is also worth recalling that some physicists, aiming to address the measurement problem in quantum mechanics, already accept energy non-conservation even apart from the philosophy of mind (e.g., the GRW theory in general, see for example (Feldmann and Tumulka 2012); or (Bassi et al. 2013). Such examples show that experts are not nearly as threatened by the prospect of energy non-conservation as many philosophers of mind. Conservation principles are considered useful accounting devices and sanity checks by physicists, but it is not considered absolutely obligatory that they hold exactly and at all times and places.

However, there is a significant disanalogy between cases of non-conservation considered by physicists and non-conservation brought about by non-physical minds, a disanalogy that continues to cause worries as long as one adheres to the framework of naturalism. The disanalogy is that in cases countenanced by physicists, energy or momentum non-conservation, even if local, is in complete accord with the laws of nature. Take the GRW theory as an example: the GRW theorist has two options for integrating his theory with the laws of nature. The first is to say that spontaneous collapses just happen, with a certain probability, and that this is simply a law of nature, in the sense of a brute fact which is not further explainable. The second option is to claim that the spontaneous collapses can be explained by a deeper, yet undiscovered law of nature. On interactive dualism, however, neither of those moves is possible, at least if one wishes to grant the mind independent causal powers requisite for freedom and reason. The deeper worry is therefore that genuine ID will not be amenable to integration with the laws of nature and hence incompatible with them. After all, what law of nature could possibly accommodate such, even though infrequent, capricious behavior? A thorough resolution of this tension requires an investigation of the underlying concept of 'laws of nature', which will be done in part IV. Here I just wish to note that although perfectly coherent from a philosophical, mathematical and physical standpoint, ID is a big bullet for naturalists to bite. Beyond that, in the remainder of this chapter I want to show how the conditionality response beautifully falls into place on a theistic worldview.

First, theists have a precedence of non-conserving interventions in God's miracles. Although attempts have been made to explain miracles wholly naturally, or to argue that conservation principles preclude divine

[49] The razor principle – typically attributed to William of Ockham ("Ockham's razor"), although Robert Grosseteste before him used a similar principle – roughly states that when adjudicating between two theories with equal explanatory power, one should prefer the one that invokes fewer entities.

interaction (Fales 2010), it is best to understand miracles as macroscopic instances of non-conservation (see (Larmer 1986; 2014; Plantinga 2007). In fact, divine miracles significantly parallel ID. Not only has the same objection (from conservation principles) been levelled against both, but also have similar answers been given (e.g., Plantinga's 'divine collapse causation' (Plantinga 2011, ch. 4) mirrors the 'consciousness collapse' approach for mental causation (cf. chapters 6 and 8).

Because of this precedence, theists can make perfect sense of a world that contains both and contrasts between the (strictly regular) workings of nature and the actions of free agents. God is a free, non-physical agent; so are human beings. God interacts sometimes with nature in a non-quantity-conserving way, yet in full accord with the Noether theorem; human minds do likewise, albeit on a much more modest scale. Best of all, theism can even make sense of the parallel between the two: God is the creator of human minds/souls (which on Christianity are made in God's image, which further strengthens the point). On a certain reading, theists can even explain the interrelation of nature and mental interactions. C.S. Lewis defined nature as "what springs up, or comes forth, or arrives, or goes on *of its own accord;* the given, what is there already: the spontaneous, the unintended, the unsolicited" (Lewis 1960, ch. 2, emphasis in the original). Free agents like God or human beings are of course in sharp contrast to this: they intend their actions, and they act for reasons, and not mindlessly forced by their inner nature or external forces. One may speculate that such a contrast helps us identify ourselves as free agents, and as significantly distinct from non-human nature. At the same time, nature can be understood on theism as a "hostess" who can accommodate any change impressed on it from outside (see introductory quote and below). Lewis masterfully summed up the implications of that 'hostess' conception of nature for the interrelation of laws of nature and interventions:

> The divine art of miracle is not an art of suspending the pattern to which events conform but of feeding new events into that pattern. It does not violate the law's proviso, 'If A, then B': it says, 'But this time instead of A, A2,' and Nature, speaking through all her laws, replies 'Then B2' and naturalises the immigrant, as she well knows how. She is an accomplished hostess. (ibd., ch. 8)

This is in perfect line with the conditional nature of conservation principles. At the same time, it already gestures towards a certain understanding of the laws of nature – or should we rather say, rejects another, widespread idea about them? On a conditional view of conservation principles, it is as impossible that conservation should hold with necessity as it is that the conservation principles are suspended in case conservation fails. Likewise, on the view suggested by Lewis, laws of nature do not prescribe events with inexorable necessity; but neither does a course of events different from the natural one (*sans* intervention) rescind the validity of the laws of nature. We shall explore further in part IV how these initial ideas can be honed into a thorough philosophical theory.

Conclusion

I conclude that the conditionality response is physically and mathematically robust and gives everyone what they (said they) asked for. Since physics is no longer good for an objection against ID, any argument against it must come from philosophy. However, as we have seen (chapters 2 and 3), the two big philosophical objections levelled against ID – the causal closure objection and the causal heterogeneity objection – stand on shaky ground. ID is nonetheless not widely accepted, presumably because of staunch adherence to naturalism. A possible ease of that tension could be provided by an in-depth investigation of the notion and metaphysics of the laws of nature, something I will do in chapters 10 to 14.

8. Can Consciousness-Collapse QM Facilitate Interactive Dualism Even Apart From Conservation Principles?[50]

8.1 What can CC offer apart from the conservation of quantities?

As we've seen in chapter 6, there is little hope that the consciousness-collapse (CC) interpretation of quantum mechanics (QM) can provide energy- (and momentum-) conserving mental causation. Still, as also noted there, CC was originally developed as a solution to the measurement problem, not as an account of mental causation, and might thus offer some assets (not only) to the dualist even apart from compliance with conservation principles. For example, regarding the quantum-mechanical side of things, CC gives a relatively clear criterion for when measurements take place; it offers a straightforward solution to the measurement problem and, in so doing, takes the standard QM dynamics at face value. For the dualist in particular it might contribute significantly to the question of how the (conscious) mind causes anything in the physical world, and it promises to establish a close, perhaps even (nomologically) necessary connection between mind and matter. The latter two, although more pertinent to the causal heterogeneity objection (see chapter 3), are aspects worth considering in the context of the present inquiry.

There are several objections against CC. A first one is that it is 'mysterious'. Thus Peter Lewis:

> Wigner postulates a strong form of interactive dualism in order to justify a duality in the physical laws. Few will want to follow Wigner down this path: non-physical minds, especially causally active ones, are mysterious at best. (Lewis in (Gao forthcoming))

Pace Lewis, a mere 'mysteriousness' complaint without further qualification does not establish a good reason to reject a view. If there is a *prima facie* mystery, then why not explore it? Because the postulated account is incoherent? Then one should argue to this effect. Mysteriousness objections generally smack of an all-too-easy dismissal of an unloved view, and Lewis' rejection of dualistic CC marks no exception. One the one hand, if indeed the dualism involved constitutes the main problem, then Lewis should note that CC is in principle compatible with materialism (a way barred for the current exploration). On the other hand, interactive dualism is by no means a forlorn view, as in particular chapter 1 should have made clear.

A second objection is that CC is *imprecise*, more exactly, that consciousness is an imprecise notion. David Albert writes:

> How the physical state of a certain system evolves (on this proposal) depends on whether or not that system is conscious; and so in order to know precisely how things physically behave, we need to know precisely what is conscious and what isn't. What this "theory" predicts will hinge on the precise meaning of the word conscious; and that word simply doesn't have any absolutely precise meaning in ordinary language; and Wigner didn't make any attempt to make up a meaning for it; so all this doesn't end up amounting to a genuine physical theory either. (Albert 1992, 82-83)

Albert's worry, though more deserving of the label 'objection' than Lewis' out-of-hand dismissal, is unfounded. In spite of conflicting notions of what consciousness in the literature, one can delineate the relevant idea quite nicely as *phenomenal consciousness* as opposed to *psychological consciousness* (see Chalmers 1997, ch. 1.2). And phenomenal consciousness is a 'binary' concept: either something is phenomenally

[50] Large parts of this chapter have been published in my "Does Consciousness-Collapse Quantum Mechanics Facilitate Dualistic Mental Causation?", *Journal of Cognitive Science* 21:3, 429-474 (2020). Slightly modified reprint courtesy of *Journal of Cognitive Science*.

conscious (there is something it is like to be that thing, for example, an animal), or it is not phenomenally conscious (there is nothing it is like to be that thing, e.g. a stone). Thus, there can at best be *epistemological* imprecision in our *judgment* whether something is conscious or not, but no *ontological* imprecision.

I suggest that the most pressing challenge for CC is the quest for *dynamic principles*, that is, mathematical equations that describe how consciousness collapses the wave function. Lack of such principles prevents CC's being received into the august ranks of physical theories. Maybe this is the worry underlying the 'mysteriousness' complaint: that consciousness has so far not been – or cannot be – cast into a variable that figures in a physical equation.

In this chapter, I shall first explore the possibilities to find dynamic principles for CC. I will mainly discuss David Chalmers's and Kelvin McQueen's 'superposition-resistant observables' approach, while adding a few ramifications of my own (section 8.2). I will then explore the question of what consequences for human freedom it would have if CC could be cast in terms of dynamical equations (9.3) before summing up the main lessons from this chapter (9.4).

8.2 Finding dynamic principles for CC

8.2.1 Making consciousness accessible for physical theory

To find dynamic principles for CC, more needs first to be said about how the mind 'measures' quantum objects[51]. As a first step, non-physical consciousness needs to be linked to brain structures in such a way that it becomes available for physical theory. One necessary condition for this is positing a physical correlate of consciousness[52] (PCC), which is, so to speak, the 'foothold' of consciousness in the physical world. It is worth noting why this is necessary. In principle, one could establish a materialist version of CC: if the conscious mind is identical to the brain or some physical configuration of it, then a certain (physical) property of that configuration would account for the collapse of the wave function. On dualism, however, the mind is distinct from the brain, and so we need a physical correlate of it in the brain in order to develop mathematical dynamics for CC.

For a PCC we need assume no more than a *systematic correlation* between physical states of the brain and conscious states, something even substance dualists can accept. This *systematicity* (which does not preclude multiple realizability) is at the same time crucial for finding a dynamic for consciousness collapse (otherwise, the establishment of mathematical equations would not be possible).

What exactly is a PCC? I shall use David Chalmers' (D. Chalmers 2000) definition in an adapted way[53]:

A PCC (for content) is a minimal physical representational system P such that representation of a content in P is sufficient, under conditions C, for representation of that content in consciousness.

[51] In QM, measurement is understood as the process that leads to wave function collapse (insofar as collapse occurs at all). Measurements can be literal measurements with devices such as photon detectors, or figurative measurements like the observation by a conscious observer.

[52] The bulk of the literature deals with neural correlates of consciousness (NCC). PCC include NCC and thus have the advantage of keeping the search for correlates of consciousness more open.

[53] Chalmers focuses on *neural* correlates of consciousness (NCC) in the cited article. I replaced 'NCC' by 'PCC' and 'neural' by 'physical'.

The qualification 'for content' is important because we are not interested in a correlate for the general property of *being conscious* (as opposed to being in a coma, for example), but in a correlate for certain *contents* of consciousness (in this case, volitions). What is meant by 'representation'? When I see an apple, the content is 'apple'. Neural activities appropriately correlated with my seeing-an-apple are a *physical* representation of that content, and of course my experience of an apple is a *conscious* representation of the apple. It should also be noted that the above definition suspends judgment on causal relations. It does not say that a certain representation of a content in P *causes* a representation of that content in consciousness (or the other way around), while it certainly leaves that possibility open.

What could P be? It could in principle be a certain anatomical brain area, but it might also be 40-hertz oscillations in the cerebral cortex (Crick and Koch 1990) or a field generated by the brain (Libet 1994). Of course, different proposals pertain to different aspects of consciousness, so not all of them will be equally viable for our purposes here. However, I am not concerned with those details here; all we need is a candidate for P which allows CC to be formulated more rigorously.

The next question we need to tackle is to which extent the content of P is mapped onto the content of consciousness. In particular, will a superposed P be mapped onto a superposed consciousness (*modulo* consciousness-caused collapse)? If it does, a CC theory must find an explanation why consciousness in fact never seems to be superposed. If it fails to do so, the mysteriousness objection does seem to gain some traction after all, and the whole project should probably be abandoned. I will therefore assume that the aforementioned correlation does hold. Without having too much space to argue for this, here is why: in order to make the role of consciousness accessible for physical theory, we need to claim a strict (though multiply realizable) correlation of P-states with consciousness states. The physical collapse will then be caused by a property of P (or P *tout court*, if P *is* the collapse-causing property, see below) which is strictly mapped onto some property of consciousness (both of which, as we will see, must at least statistically resist superposition). But then *some* P-properties (or P itself), if they can enter superposition at all (due to their being physical properties), will, when in superposition, correlate with consciousness properties (or consciousness *tout court*) in superposition, due to the strict mapping.

Thus, we can sum up the relation between states of P and states of consciousness as follows:

superposed P → superposed consciousness[54]

and

definite P → definite consciousness.

8.2.2 CC and the 'superposition-resistant property' framework

On CC, consciousness 'collapses' the wave function by 'measuring' it. Quantum objects in superposition have no single classical state; for example, an electron in a superposition of spin +½ and spin -½ is neither in the state of spin +½ nor of spin -½, but in what is called a superposition of the two states; in fact, one speaks of a superposition of those two states because those states are the only possible results when measuring the superposed electron. Measurement in the language of QM is, consequently, what leads to definite states (and results), whether it be carried out with actual measuring devices or otherwise. In terms

[54] The arrows signify entailment.

of a wave function diagram, a measurement turns the extended 'hills and valleys' of a wave function into one sharp peak, at least in case of a collapse onto an eigenstate, which I shall assume here (see fig. 6.1 and 6.2). Such a process (if it indeed takes place) is a disruption of the temporal evolution of the wave function dictated by the dynamical Schrödinger equation.
One can speak of the wave function collapse merely as an ontologically non-committal description of the transition from the mathematically dictated superposition to the definite result; in this sense, collapse is part of the standard dynamics of QM. However, 'collapse' as understood here does assume the ontological reality of the collapse process. It should be noted that in virtue of this, CC belongs to the greater family of collapse interpretations of QM.
It is important to note that the Schrödinger dynamics is *deterministic* (not probabilistic!) and *unitary* (roughly, it preserves all the information contained in the wave function across time). The collapse, by contrast, is *non-unitary*, i.e. it 'destroys' most of the information contained in the wave function prior to the collapse, by reducing the multitude of possible values contained in the wave function (corresponding to a superposition of states) to just one. After a collapse, the Schrödinger dynamics again determines the evolution of the wave function, until the next measurement/collapse.

Spelling out a physically viable theory of CC requires an explanation of how the mind 'measures' superposed brain structures. (Candidates for such structures could be as small as a subatomic particle or as large as whole brain areas – see section 6.3). The most elaborate approach I know of is the one by David Chalmers and Kelvin McQueen. Their approach is basically to consider consciousness and/or its PCC as 'superposition-resistant observables' (I shall henceforth use the shorthand 'super-resistant'). This locates their account within the class of *super-resistance models* of QM (Chalmers and McQueen forthcoming, 12), which are in turn classifiable as *fixed-locus* models of measurement collapse, as opposed to *variable-locus* theories (ibd., 11-12). Henry Stapp's account (see section 8.3) is an example of a variable-locus theory. The difference is that on variable-locus theories, the collapse can occur onto different observables (position, momentum,…), while on fixed-locus accounts there is a special observable that serves as the fixed locus of collapse.
The dynamics of super-resistance can be made precise in either of two basic ways. Either the super-resistant observable is entirely forbidden from entering superpositions (via 'super-selection rules', cf. (Wick, Wightman, and Wigner 1997), or superpositions of the super-resistant observable are unstable and tend to collapse. Let's begin with the former, stronger version to get the idea across more clearly (as we will see in section 8.2.3, the strong version faces a severe problem, which suggests its replacement by the weaker version). According to the strong (henceforth 'absolute') version, a 'super-resistant observable' (or 'super-resistant property' or 'super-resistant quantity') is characterized as follows:

- There can be no superposition of a super-resistant observable => a system's wave function is always in an eigenstate of the operator of this observable.
- Whenever a super-resistant observable is about to enter a superposition (via entanglement with a superposable property of another physical system), it collapses to definiteness.
- The probabilities of the collapse are given by the Born rule for the associated super-resistant operator.

This solves, and indeed is motivated by, the puzzle described in section 6.1: if consciousness resists superposition, this explains why an observer of a superposed quantum state never ends up in a state of superposed consciousness.

For now, just to get an idea how super-resistant properties work, let me briefly present an illustrative toy model. Suppose a super-resistant property has two eigenstates, $|m_1\rangle$ and $|m_2\rangle$ with eigenvalues μ_1 and μ_2, respectively, and that the structure possessing the super-resistant property is in state $|m_1\rangle$. Suppose further that a particle in a superposition (of observable p) of $\pi_1|p_1\rangle$ and $\pi_2|p_2\rangle$ gets entangled with the super-resistant observable. The (fictitious) superposition

$$|\Psi\rangle = \mu_1|m_1\rangle\,\pi_1|p_1\rangle + \mu_2|m_2\rangle\,\pi_2|p_2\rangle$$

results, which will in fact not occur. Rather, there will be an instantaneous collapse onto either $|m_1\rangle|p_1\rangle$ or $|m_2\rangle|p_2\rangle$.

Finally, to make things more concrete, which property might, for the purposes of ID, serve as super-resistant? In principle, any physical or mathematical property is a candidate. There are at least two constraints, however: (1) a super-resistant property must not be too common, so that there is room for superpositions (otherwise collapses could not occur); (2) a super-resistant property must not be too rare, so that measurements (which always yield definite results) always involve super-resistant properties. Constraint (1) rules out at least position, mass or very common atoms or molecules. Constraint (2) suggests that super-resistant properties must be present in brains.

8.2.3 CC and the Quantum Zeno effect

The main problem for the absolute, stronger version of super-resistant observables is presented by the so-called Quantum Zeno Effect (QZE). To understand the problem better, let's frame the collapse dynamics differently: suppose the collapse takes place at a time interval Δt such that if the system has evolved (according to Schrödinger dynamics) in the preceding Δt into a non-eigenstate of the super-resistant observable, it collapses probabilistically into an eigenstate of that operator, with probabilities given by the Born rule. For the absolute super-resistance model, the dynamics is the limiting case where Δt approaches zero. This is equivalent to, in standard QM dynamics, the system being continuously measured by an outside observer (which need not be taken literally, albeit one might imagine God as the outside observer). This 'continuous measurement' leads to the QZE. The QZE roughly consists in quantum systems becoming 'frozen' in one state if they're being measured in very rapid succession[55]. For a super-resistant property, this means that by itself it cannot evolve, at least if we take the Schrödinger evolution as a basis (which requires uninterrupted, though changing, superposition, since systems cannot leap from one eigenstate to another eigenstate, for eigenstates are orthogonal to each other). And if consciousness is strictly correlated with a super-resistant property, then it cannot evolve either, at least not by itself. But this seems unacceptable – surely our consciousness constantly evolves, i.e. changes. On the other hand, if we allowed

[55] Rigorous definition of QZE: "If system S is in eigenstate E of some observable O, and measurements of O are made N times a second, then, the probability that S will be in E after one second tends to one as N tends to infinity."

(contra the initial hypothesis of super-resistant properties) the consciousness-correlate to go into superposition, we would get the similarly implausible result of a non-definite consciousness.
Interestingly, some have thought the QZE conducive to interactive dualism, arguing that the mind could, through a fusillade of measurements, hold certain brain structures – which would otherwise 'smear out' according to the Schrödinger dynamics – in one definite state and thus trigger actions (Henry Stapp is perhaps the most famous proponent of this theory, see a.o. his 2017). What the proponents of such a view seem to have overlooked, though, is that the mind's ability to change is as important as its ability to hold a certain physical state in place for a sufficiently long time.

A first way out of this dilemma could be to stipulate a different physical dynamic for super-resistant properties, one that does not entail constant superposition[56]. The idea is that the super-resistant observables evolve in discrete steps – from eigenstate to eigenstate – not continuously as in the standard Schrödinger dynamics. I am not even sure whether such a dynamic is even possible mathematically, let alone physically, given that eigenstates are orthogonal to each other. If it is not possible, then so much the worse for this approach, which faces some other rather severe problems, as will be evident shortly. At any rate, *if* the discrete evolution were possible, then via the coupling to the PCC, we would get a consciousness that evolves in *discrete steps* rather than continuously. Is this plausible? It depends on two things: (1) whether such a 'jumpy' evolution is compatible with our first-person conscious experiences; (2) whether the super-resistant property can have enough eigenstates.
(2) should not be a problem. Take the position operator: it might have infinitely many eigenstates (or a continuous spectrum of eigenstates) or (if space is discrete) a finite, vast number of eigenstates. This could similarly be the case with the m-operator. As to (1), it needs to do justice to our conscious experiences by excluding gaps and indefiniteness in the stream of consciousness. The 'discrete evolution' theory clearly disallows indefiniteness (which is a problem for the 'approximate super-resistant property' account, see below). Regarding gaps one might argue that, provided the quantum leaps from eigenstate to eigenstate take (very little) time, this short intermediate period constitutes a gap.
It is hard to make an informed judgment about this since we simply lack pertinent background knowledge. Comparing the stream of consciousness to a movie is of no help here. We experience a movie as continuous (rather than as what it actually is, a rapid succession of pictures) because of limitations of our visual system, but more importantly we experience a movie *in our consciousness*, whereas it is nonsense to say that we experience *consciousness itself in consciousness*. While we have experiences of both continuous-appearing and discrete sequences of pictures, we have no experience of a continuous vs. a discrete consciousness. At least it is not *obviously* false that our consciousness is a sequence of ontologically discrete 'leaps'. Furthermore, there is still the option that the super-resistant operator has infinitely many eigenstates (or a continuum thereof), which certainly supports 'gap-less' consciousness.
McQueen and David Chalmers (McQueen, n.d.) propose another way out of the quantum Zeno dilemma. They call it 'approximate super-resistant observables'. The idea is inspired by the GRW theory. An approximate super-resistant property is a property which can enter superpositions, but *stochastically* resists them (in contrast to absolute super-resistant observables which *always* resist superposition); further, whose superpositions *stochastically* collapse (as opposed to *always collapse*, like in absolute super-resistant properties

[56] I am indebted to Dustin Lazarovici for this suggestion.

discussed above). Chalmers and McQueen suggest that the collapse frequency depends on the size of the superposition; the more superposed the super-resistant property is, the more probable a collapse becomes. Their favorite candidate for a super-resistant property is integrated information as it figures in Integrated Information Theory (IIT). Simply put, their theory has it that the informational quality (q-shape as identified in IIT, cf. (Oizumi, Albantakis, and Tononi 2014)) of consciousness and not its informational quantity ('φ') resists superposition; in other words, collapse is more probable the greater the difference in quality. An implication of this is that the more integrated information (the more consciousness) accumulates in a brain, the more probable a collapse onto a definite state of the super-resistant operator (and, by implication, of the physical substrate, and of consciousness) becomes.

Which of the two solutions to the quantum Zeno dilemma is more attractive? Both have their appeal, and both come at a cost. The approximate super-resistant property approach offers the advantage of not requiring any alteration of standard Schrödinger dynamics for the PCC. However, there will be tiny periods of time during which consciousness is superposed and hence indefinite. But presumably none of us ever experiences 'superposed consciousness'. Are the superposition periods so short that we don't experience them? Perhaps, but, again, consciousness is not something *of which* we can be aware but the one thing *through* or *in which* we can become aware of things. Therefore, I take it that we could not fail to experience even short superposition periods.

On the other hand, the absolute super-resistant observable approach with discontinuous dynamics has the obvious drawback that it commits one to a completely different dynamic just for this one observable. This is a heavy cost, and it does not really offer a recompense: after all, our consciousness seems as little discontinuous as it seems superposed. On this weighing, the 'approximate' approach seems to have the edge over the 'absolute' one.

8.3 Free actions in danger?

Let's take stock what we have seen so far. If a dualist wishes to hold onto CC to explain mental causation and make the theory rigorous, he needs to find a way out of the quantum Zeno dilemma. As argued above, there is no easy way out. The 'best of a bad lot' is to posit approximate super-resistant properties along the lines of Chalmers and McQueen, although this entails the rather implausible consequence that consciousness will be superposed for short periods of time.

There are, however, two big challenges for *any* theory that seeks to establish a physical dynamic for CC.

The first one is the question of *causal efficacy*. Does the mind really play a *causal* role, such that some events happen at least partly because of a causal contribution of the mind, or is it just 'consciousness dangling from physics', where the physics does all the causal work? To be sure, causal efficacy is guaranteed if the mind/consciousness is identified with the relevant super-resistant property. Since it is (causally) *in virtue of that property* that collapses in the brain occur, one can confidently speak of the mind being causally efficacious. However, this constitutes an abandonment of dualism and an adoption of physicalism.

On dualism things are a bit more complicated, because dualist theories ontologically distinguish between mind and matter, and thus between mind and the PCC. Recall, first, that a strict mapping of pertinent brain states onto mental states is a *sine qua non* for making CC scientifically rigorous. This yields the following *prima facie* situation (fig. 8.1):

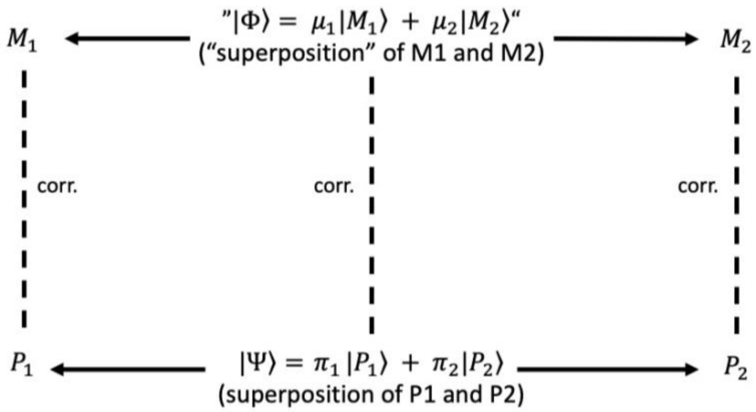

Figure 8.1

The superposition Ψ of the super-resistant observable (of eigenstates P_1 and P_2) collapses to P_1 or P_2, and the mental states M_1 and M_2, respectively, follow suit; the mind just tracks brain events, so to speak, no causal work is left for it to do. Note that this parallels the problem of supervenient property dualism gestured towards in section 1.1. The main reason is of course the tight connection between the mental and the physical in both cases (supervenience and strict correlation, respectively).

It may, however, be objected that this picture is unduly restrictive. After all, even a strict correlation between the mental and the physical is compatible with the mental having genuine causal efficacy. All it takes is a genuine influence of the mental on the physical – an influence that makes a difference to the sequence of physical events. Can this be had on a dynamically rigorous CC? What clearly stands in the way is that the dynamics of the superposition Ψ (or fictional superposition in the case of absolute super-resistant observables) fixes which eigenstate it will collapse onto (with probabilities given by the Born rule). But perhaps there is yet a possibility for the mind to get a foothold in the process, a possibility I will discuss in connection with the other crucial question regarding dualistic CC dynamics.

This other question is that of *causal independence*. By 'causal independence' I mean what other authors have termed 'being the ultimate author of one's actions' (Moreland and Rae 2000), 'being the producer of one's actions' (Hasker 1999) or 'self-movement' (Gasser 2015)[57] (see also sections 2.1 and 2.2). In effect, this is nothing but our commonsense conception of free will: I can generate actions/mental events in the style of a 'prime mover'[58], without being caused (deterministically or indeterministically) to do so by anything external (ultimate authorship condition), and I can choose between alternative actions/mental events without being externally tied down to one of them (alternate possibilities condition[59]).

Now, is the mind, on one of the outlined CC theories, causally independent in this sense? At first glance, it might seem that it cannot, for it is the super-resistant property that is doing all the work. But all the super-resistant property framework says is that superpositions (in so far as they occur) collapse to one of

[57] One should think that 'agent causality' suffices, but there are many views figuring under that label which clearly assign far less causal independence to the agent than I have in mind here (e.g. (R. Clarke 2000; Chisholm 1976; Kane 2011)).

[58] I chose this strong expression to delineate the view proposed here from other, weaker views of agent causality, which at first glance speak no less of the agent being the cause of her actions. For example, Clarke (2011) claims that the agent is a co-cause besides her reasons; Swinburne (2013) holds that it suffices for agent causality that the agent complete the otherwise insufficient physical causes.

[59] Some philosophers have attempted to show that this principle is not a necessary condition for free will (e.g. (Frankfurt 1969). I take those attacks to be unsuccessful (cf. Swinburne 2013, ch. 7.6).

the super-resistant eigenstates with a probability given by the Born rule (on the absolute super-resistant account, no collapses occur). *Prima facie*, this does not prevent the mind from making some causal contribution to the *direction* of the collapses. But we must be careful how to spell this out. First, it seems to me that we need to distinguish between the settings of *observation* and of *action*. I assume that it is wildly implausible to claim that observers somehow 'choose' whether an electron goes spin up or spin down, but that an agent choosing his action is exactly what we want to get at. So, we could propose that in observations, the mind is causally efficacious but not causally independent, while in actions it is both efficacious (regarding the collapse direction) and independent.

A possible complaint against this is that it sounds ad hoc. After all, the underlying mechanism (consciousness-collapse) is the same in both cases, which suggests an equal treatment: either the mind is merely causally efficacious in both cases, or it is causally independent in both cases. But none of these options is satisfactory, since the first one precludes genuinely free actions, and the second one posits an implausible telekinetic power of choice over what happens in the physical world.

Let's therefore grant for the sake of argument that the mind can have different causal roles in observation and action as outlined above. It then follows that the mind directs the collapses. But doesn't this run counter to quantum theory and physical theory in general? At least not obviously, and not in all respects. Remember that we allowed for some theoretical role of consciousness in the first place. Then suppose that consciousness is a super-resistant property. Do we know *in what manner* this super-resistant property brings about collapses? We don't. The collapse process is a 'black box' for us. It might as well be that the box contains the effects of conscious volitions. One could still object that the introduction of super-resistant properties was meant to naturalize consciousness by making it accessible for physico-mathematical theory, and volitions elude the grasp of physics and mathematics. But this objection fails to acknowledge that the naturalization of consciousness does not annihilate its phenomenal nature; to the contrary, it must preserve it, on pain of ceasing to be a theory of phenomenal consciousness. In fact, naturalistic theories of consciousness try to model features of phenomenality and often center on the concept of information (see e.g. (D. Chalmers 1996; Tononi et al. 2016). Hence, I conclude that claiming causal independence of the mind does not violate the spirit of a naturalistic CC theory.

There is one restriction for the causal independence of the mind, however, at least if one wishes to stay on safe ground in physical theory: the Born rule. The Born rule is a mathematical postulate that fixes the probability of outcomes of measurements of a superposed quantum system, and it is experimentally very well confirmed. For example, it follows from the Born rule that an electron in a superposition of spin-up and spin-down (relative to a given axis) has a probability of 0.5 for moving up or down in a magnetic field, respectively; and indeed, doing sufficiently many (say 10,000) measurements with prepared electrons yields probabilities of the two possible outcomes close to 0.5 each. If a CC version of ID is to respect the Born rule, then the relative frequency of a given type of outcome must converge to the value given by the Born rule. For example, if the super-resistant operator has four eigenvalues, the relative frequency of each corresponding outcome will converge to 0.25 in case of a uniform superposition (with equal weights of the superposed states), or, more generally, to a number depending on its individual weight (while the relative frequencies/probabilities must add up to 1). In other words, it is simply not possible that, in the long run, a certain action/action type (corresponding to a certain eigenvalue of the super-resistant operator) occurs

in, say, 40% of instances of willings[60], given a uniform superposition (and mutatis mutandis for non-uniform superpositions).
How much of a restriction of freedom is this? We should pause here for a moment and realize that we've reached a crucial point. The Born rule can, for all we know, reasonably be called a law of nature. Thus, if mental causation does not respect Born's rule, a law of nature is violated. We should then in principle be able to detect those violations in brains. An independent issue is how to construct the notion of 'laws of nature', if such violations are possible. This will be the topic of part IV.
But back to the question: how mutilated will freedom come out if mental causation respects the Born rule? This will not least depend on the number of eigenvalues associated with the super-resistant operator in conjunction with the possibility of multiple realizability. The more eigenvalues, the more 'destinations' there are for a collapse process. However, there must also be multiple realizability, that is, a certain action (type) needs to be realizable by different brain states (corresponding to different eigenvalues). This is because even if there are many possible target states of a collapse process, it should be ensured that a number of them (not just one) leads to a desired action; otherwise, volitions and body movements would frequently come apart. But those two conditions (multiplicity of super-resistant eigenvalues and multiple realizability) seem to be just necessary, not sufficient conditions. After all, a multiplicity of eigenvalues distributed over redundant brain states just raises the probability of the desired action being carried out; it does not guarantee it. It might still happen, with a certain probability, that I endeavor to raise my arm, but end up with stomping my feet. I presume that such aberrations are extremely rare, and that there are good physiological (as opposed to quantum-mechanical) explanations for them. Therefore, it seems that a CC theory of ID cannot afford to allow inherent action failures.

A proposal of how to reconcile free will and CC comes from Henry Stapp, one of the most prolific writers on this subject. I cannot do justice to his ample work here, especially because his theory seems to have evolved over the years. What I am going to do is draw mainly on his most recent work *Quantum Theory and Free Will* (2017), focusing on the most relevant points for ID. Stapp has already appeared as (falsely) proposing that the QZE helps explain the causal efficacy of the mind. The following discussion centers on other aspects of his account, aspects that can be evaluated independently of the misguided QZE proposal.
Stapp divides the causation of a free action into three processes[61]. Process 2 is the Schrödinger dynamics, process 3 the probabilistic collapse process, and process 1 the mental influence. He conceives of process 1 as a Y/N question, e.g. "Am I moving my finger?". The link to the brain as a quantum object consists in every volume *V* of a brain's state space having a discrete observable with the values Y and N. David Bourget (Bourget 2004, 9) succinctly sums up this view as follows:

[60] By 'willing' (or volition) I mean here the mental state that either constitutes the actual carrying out of the (conscious) action or inevitably (physiological failures notwithstanding) leads to the action. In other words, a conscious raising of my arm couldn't happen if I didn't will it in the sense here under discussion. By contrast, an expression of will like in "I want to watch a movie tomorrow night" is *not* what I mean by 'willing'.
[61] In his 2017, Stapp speaks of two processes (1 and 2) only. Although this is his most recent advance, it is still legitimate to depict him as assuming three processes, since he divides process 2 into two phases (2017, 8).

A measurement of this observable for some *V* yields a positive outcome if the brain is in *V* and negative otherwise. There should thus be a measurement that can tell us if a brain is in one of the many possible states in which any decision has been made.

In other words, in contrast to my suggestion concerning the super-resistant properties account above, Stapp is multiplying observables rather than eigenvalues (something Chalmers and McQueen (forthcoming, 11) call a 'variable-locus model'). It is important to note that for Stapp, the two pivotal points for free actions are processes 1 and 3. Process 1 determines the 'question'[62], and process 3 (probabilistically) determines the answer, thereby not – it seems – strictly observing the Born rule. He writes:

In the quasi-orthodox theory nature's choices are not random, hence lacking a sufficient reason to be what they are, but are assumed to have sufficient reasons that entail a biasing of nature's choices in favor of choices that advance the personal values of the subjects in these experiments who are posing the questions pertaining to matters of interest to themselves. (2017, 99-100).

The (in my view) insurmountable difficulties of integrating *axiological* values into a physical theory notwithstanding, it may well be that already at this point, Stapp's theory breaks the Born rule, something he emphatically denied in earlier writings (e.g., Stapp 2004). Be that as it may, on Stapp's account, each 'question' the mind may ask in order to trigger an action corresponds to a 'projection operator' belonging to an observable[63]. Now, the values of those observables are subject to the Schrödinger evolution (process 2) and thus the probabilities of receiving a 'Yes' or 'No' answer, respectively, depend on the point of time of the question. Put simply, the probability for a desired answer (Y/N) to a certain question rises to its possible maximum (0.5) within a certain time limit (call it T_S), only to result in a collapse (without the mind's interference) and re-rise of the probability within the next T_S period (see fig. 8.2). What the mind now does, according to Stapp – not, nota bene, collapsing the wave function – is to 'hold in place' the collapse outcome (in case it was the desired one) via the QZE, for time period T_a, as represented in fig. 8.2:

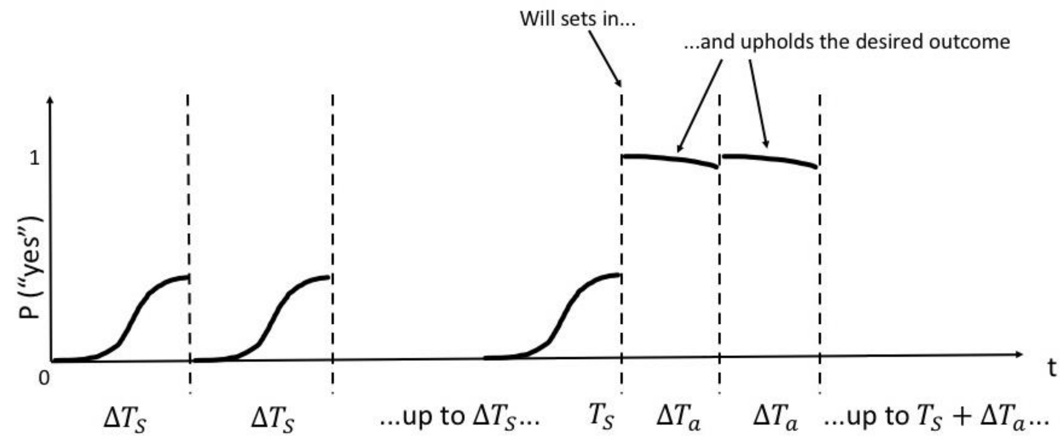

Fig. 8.2: Process of enforcing a free decision according to Stapp. The vertical axis represents the probability that a collapse to the 'yes' state occurs upon measurement (here it is assumed that the state is the desired

[62] According to the most recent version in Stapp 2017. He held this already in his (1999), then intermittently retracted it.

[63] „The detailed psychophysical rules that are presumed to fix or determine the projection operator P (which defines the empirical question) are not yet known, and I have made no basic commitment concerning the form of those rules." (Stapp 2004)

one). Dashed lines mark actual events. In the above example, the will sets in at T_S and upholds the probability of the desired outcome at ~1. (Own creation after fig. 3 in Bourget 2004)

Stapp's account can be criticized in several ways. First, his conception of 'mental measurements' as empirical questions is awkward and implausible. Asking a question simply does not seem to be what is going on when we willfully endeavor an action. (Maybe a case can be made for the construal of *phenomenal experiences* as a 'Q & A game', but I doubt that). There is no introspective evidence for this claim, and simply re-christening volitions as questions won't do, not least because we can clearly distinguish between the mental events of questions and of volitions. But admittedly, little hinges on how one construes the mental events that initiate actions; perhaps one could just claim that volitions as directed to one outcome are successful only to the extent to which the Born rule allows it. Nothing essential would be lost by dropping the 'Q & A' talk.

Another, more serious concern is that free actions may have to wait, since 'nature's answer' to an 'action request' can be negative. This stands in contradiction to our daily experiences. Stapp seeks to evade such an obvious discrepancy by making mind-matter-'communication' exceedingly rapid; for example, he holds that many 'Nos' can occur without any passage of time (2017, chapter 2) and that the mind can transmit information faster than light (2017, chapter 6 + App. A). I leave it to physicists to judge whether all this is tenable and whether his rejection of 'no superluminal transmission' does not run up against the no-signaling theorem[64].

A third objection is that Stapp's account entails ubiquitous deviations of actions from intentions. David Bourget (2004) argues to this effect. As sketched above, the mental interactions are bound to time frames T_S. Within one instance of T_S, the mind can send a finite number of 'requests' for a wave function collapse in the desired direction. Bourget assumes that T_S should be at most 100 ms, since this represents the empirically discovered lower threshold for the individuation of *complex* conscious events (Varela 1997) and coincides with Libet's 'urge' (Libet et al. 1983; Libet 1999); he further suggests that T_a should be in the order of 20 to 50 ms, which is the lower threshold for simple conscious events to be perceived as separate in time (Kristofferson 1967; Hirsh and Sherrick Jr 1961). Under those assumptions, 2 to 5 attempts of performing an action are available per timeframe. With the *ex hypothesi* stipulation that the probability of the desired response is 0.5 at each attempt, we get a probability of between 0.03 (for 5 attempts) and 0.25 (for 2 attempts) that the desired action fails. It seems clear that even a ratio of 3 failed actions per 100 is highly implausible. Stapp has replied to this criticism in the following way:

> [T]he events in a rapid sequence generated by a willful effort are not perceived as separate in time. They all run together, and their rapidity is associated with the feeling of effort. I have never found any reason why the temporal separation between the events experienced in this run-on way could not be a fraction of 1ms, if their rapidity is controlled by conscious effort. (2004, 7)

Apparently, he rejects the construal of single conscious efforts as simple conscious events falling under the 20 to 50 ms restriction. Indeed, we plausibly don't experience willful effort as a rapid sequence of events,

[64] The no-signaling theorem (or no-communication theorem) is a no-go theorem from quantum information theory which states that, during measurement of an entangled quantum state, it is not possible for one observer, by making a measurement of a subsystem of the total state, to communicate information to another observer.

but Stapp's deconstruction of conscious effort into many quantum-mechanically relevant single events is currently just a hypothesis in need of corroboration. So, the difficulty here may after all not be insurmountable.

The fourth and last objection is one of two problems of CC having to do with decoherence. Decoherence means the leaking of quantum information from a system to its environment, which could only be prevented by perfect isolation (including very low temperature). Brains, however, are of course open systems, heavily coupled to their environment, and comparatively 'hot'. In any case, decoherence times in brain structures are immensely short by the standard of time frames of our conscious lives (in the order of $10^{-12}s$ or even significantly less, depending on the exact structure in question). Whether or not decoherence times constitute a difficulty for quantum-aided mental causation was discussed in section 6.3. Here, we deal with a different problem arising from decoherence. As David Bourget (2004) argues, decohered states can and should be thought of as improper mixtures. Improper mixtures are quantum superpositions, albeit without the ability to generate interference effects. The superposition character ensures that the mind could still wield its quantum influence on the brain. At any rate, Stapp (Stapp 2000) considers decoherence as congenial to his theory.

There is a significant drawback, however. Decoherence construed this way entails that the brain is entangled with a host of external systems and that brain regions relevant for a given action are entangled with irrelevant ones. But this means that superposition collapses of brain structures must encompass the entangled systems beyond the brain as well. Applied to action, this implies that conscious effort directed at a bodily action would have *direct* consequences for physical structures beyond my body. Proponents of telekinesis should not rejoice prematurely, however. After all, a *directed* manipulation of one part of an entangled system via the other part is forbidden by the no-signaling theorem. Also, there doesn't seem to be good evidence for even random influence of mental volitions on physical objects outside the body.

More importantly even, the entanglement of relevant brain areas with irrelevant ones raises the question why only the relevant ones (e.g. the SMA[65]) collapse. This is clearly a problem for accounts like Stapp's that construe NCCs as specific brain areas, but even if one went for a neurologically more unspecific PCC, the difficulty might persist. In the super-resistant property framework, even if a super-resistant property is not a neural property, its causal role will ultimately bring about the collapse of neurons that trigger actions. It seems that we are again faced with an unpleasant dilemma: either (if this is physically tenable) one construes decoherence as the complete loss of superposition and thus runs into the problem of far too short decoherence times (see section 6.3); or one preserves the superposition character, only to face the problem of undesirable entanglements. This dilemma seems to be a problem for CC accounts in general.

8.4 Summary of conceptual problems for CC

We have seen that to elevate CC to the status of a physical theory, it needs to be supplied with its own dynamic. I presented some ways how this could be done, heavily inspired by Chalmers's and McQueen's 'super-resistant observables' approach. The QZE is a severe problem for absolute super-resistant observables that can be solved by either turning to approximate super-resistant observables, or by claiming a discontinuous dynamic; I argued that the former is preferable.

---

[65] Supplementary motor area, a brain area crucial for volitional actions (see also chapter 9).

As regards free action, any physically rigorous version of CC, due to the required strict mapping of consciousness on the PCC, runs into some problem pertaining to causality. *Prima facie*, the mind becomes causally impotent, in effect epiphenomenal, because the causal work is done entirely by the physical correlates of consciousness. A solution to this might be to grant the mind freedom to influence the *direction* (however not the *time* or *occurrence*) of collapses. The latter seems to entail violations of the Born rule in brains, though, at least if one wishes to accord unrestricted freedom to the mind. Departing from the assumption that the Born rule is a good example for a law of nature, one can already catch a glimpse of what I will identify as the most fundamental issue concerning interactive dualism: namely whether, and how, it is compatible with the laws of nature. A treatment of this issue minimally requires a theory of laws of nature that adequately accommodates the required violations.

Lastly, there is the problem of undesired entanglements, a problem specifically created by Stapp's theory. While eschewing too short decoherence times, it entails entanglements of the mind's physical correlate with all sorts of irrelevant or empirically disconfirmed structures.

At this point, one may legitimately ask why QM should be invoked at all. Note that above difficulties arise if one takes into account *all* aspects of QM, and not just the ones (seemingly) conducive to one's theory. Since those problems also pertain to the 'classical' version of CC (designed to explain definite conscious *experiences*), one may well ask if one should not look for a better QM interpretation, quite independently of interactive dualism. In conjunction with the probable failure regarding conserved quantities and the empirical problems (section 6.3), the list of motivations for CC grows thin. Perhaps a QM-based mechanism of mind-brain-interaction is, despite all the identified problems, still the best way to go as regards the physical aspect of interactive dualism. At the very least, as should have become abundantly clear by now, this, as well as the more general conditionality response, strongly suggests bidding farewell to construals of laws of nature in which local 'violations' of them are sacrosanct; in fact, as seen in chapter 7, against the backdrop of the conditionality response, 'violation' is no longer an adequate term to describe the effects of dualistic mental causation.

I conclude that the real problem an interactive dualist needs to solve is how to reconcile his view of mental causation with the laws of nature. For example, does it make sense to claim that the Born rule holds in all physical systems except those that undergo mental influence? If physical dynamics should turn out to be wholly deterministic, how can mental interaction be construed at all, without violating laws of nature? And even if dynamics were (partly) probabilistic, it seems that one needs to 'choose one's poison': either make interactive dualism law-abiding and bite the bullet of giving up on libertarian freedom or accept that probabilistic laws fail where minds intervene. Thus, the real question is how to construe laws of nature in a way that gives both physics and mental causation their due. A detailed investigation of that question will ensue in part IV.

9. Interactive Dualism and the Neuroscience of Volitional Action

If my arguments hitherto are not completely wrongheaded, then interactive dualism should by now at least look like a viable option. In chapter 2, I showed that the doctrine of the causal closure of the physical lacks warrant. In chapter 3, the causal heterogeneity objection was demonstrated to be little more than an incredulous stare at dualistic mental causation. I also tried to show that upon some reflection, physical causation is no better off than mental causation (albeit of course taken for granted), so that to the degree to which we are justified in speaking of genuine causation in the physical case, we are justified in assuming it in the mental case. Chapters 4 to 7 then dealt with the objection from conservation principles and argued against several bad and for one good way to respond to it. In chapter 8, finally, I looked at quantum-mediated mental causation from a somewhat wider angle and found that it really, if anything, exacerbates matters, especially in light of the availability of the conditionality response (chapter 7).

All this suggests is that a robust, Noether-symmetry-breaking interactive dualism is conceptually sound. What it does not show is that such mental interactions *actually take place*. This chapter seeks to fill in the gap. I must at once, however, caution the reader: the following elaboration is not going to be an empirical investigation, nor will it be a meta-study of empirical investigations *explicitly dedicated to studying nonphysical mental interactions* (mainly because there a virtually none!). Rather, although relying on neuroscientific studies of volitional action, my argument will utilize mainly generic knowledge about the functioning of brain cells. The line of thought is basically this: if a complete physical explanation of volitional actions is *in principle* unlikely, given our knowledge of the basic functioning of neurons and our knowledge of physics, then it is likely that a non-physical mind has initiated those actions.

9.1 The neurophysiological anatomy of volitional actions

It is important to make a distinction first. By 'volitional actions' in the sense relevant for our purposes here I mean those body movements that the subject reports to have 'willed' and which occur independently of external stimuli[66]. This latter addition to our everyday concept of volitional action is necessary because it exclusively considers the causal contribution of the agent and rules out external factors like sense stimuli.

Now, which brain areas are involved in the appearance of volitional actions? I am using the cautious term 'are involved in the appearance of' (instead of 'causally generate') here, because the mere sequential activity of brain regions in connection with a bodily movement doesn't by itself constitute more than a *correlation*, as opposed to a *causal* connection. However, it matters little whether we are dealing with correlation or causation here: some brain event must ultimately initiate the chain of events that lead to muscle contraction, whether one is a causal realist or a Humean (the latter will rather speak of a lawlike correlation between the brain event and muscle contraction). The following picture merely serves to show that there is a sequence of events sans external sense stimuli (to the degree to which sense stimuli can be ruled out in laboratory conditions) leading up to volitional action. In fact, it has been empirically confirmed that actions triggered by external stimuli take a different path[67]. This means that when we speak of volitional actions we refer, neurobiologically, to a specific neuronal pathway and cannot invoke the other, stimulus-induced, pathway for explaining volitional actions. The neurobiological difference between externally triggered and

---

[66] This is also the definition that the important meta-study of Haggard (2008) uses, which I am mainly referring to.

[67] Haggard 2008., 937; Brinkman and Porter 1979, 703-04

'pure' volitional actions is further buttressed by experiments showing that an activity increase in the supplementary motor area (SMA)[68] takes place during the *mental exercise* of motor actions (Roland et al. 1980; Roland 1981).

I take Haggard's (2008) overview to be sufficiently recent to reflect current knowledge about the neurophysiology involved in volitional actions. (Ironically, one conclusion that Haggard draws from his investigations is that the dualistic picture of volitional action should be abandoned for a purely physicalist one with reflexes at one end of the spectrum and volitional actions at the other end). In a nutshell, and with the caveat that the following sequence basically represents a temporal succession which may or may not be causal, the order of activated brain regions is as follows:

Basal ganglia (BG)[69] → prefrontal/frontopolar cortex (FPC)[70] → preSMA → SMA → primary motor cortex → spinal cord → muscles

Of course, all abovementioned areas (except muscles, of course) consist of neurons. Thus, when we speak of activity increases in brain regions, we really speak of activity increases of neurons[71]. It is also important to emphasize that 'activity increase' means an *increase in firing rate,* not a transition from a state of complete rest to a state of firing; in other words, neurons have a ground-state (or 'baseline'[72]) of (low-frequency) firing[73] that increases when those neurons become activated.

The question we would ideally like to get answered here is whether a *complete* physical explanation for volitional actions has been found. The answer is, at least to my knowledge, that no such thing is available to this day. Consider the above sequence: it does not offer a complete physical explanation of volitional actions (assuming the sequence to be causal) because it does not say what makes the neurons of the BG increase their firing rate. A physicalist will probably retort that there is such an explanation, but that it has just not yet been discovered. If this were all one could say about the matter, the dualist and the physicalist would be stuck in an unsatisfactory stalemate that smacks of a promissory note on the physicalist's part and of a 'soul-of-the-gaps' approach on the dualist's part.

Fortunately, we can advance the discussion even without further specific empirical investigations 'beyond the BG'. The key is to reformulate the question and ask if there can *in principle* be a complete physical explanation of volitional actions, given our best knowledge of basic neuronal functioning. This line of thought will take us to the workings of the smallest functional unit of the brain, the neuron. The question then becomes: which physical explanations are available *in principle* for increases in neuronal firing rate? To this we turn in the next section.

---

[68] The SMA is a brain region whose activity is indicative of volitional actions, at least under laboratory conditions.

[69] See e.g. Picard and Strick 1996; Akkal, Dum, and Strick 2007.

[70] Soon et al. 2008

[71] The activity measuring methods include electric potentials (e.g. Libet, Wright, and Gleason 1983; Libet et al. 1983; Deecke and Kornhuber 1978), regional cerebral blood flow (rCBF) (e.g. Roland et al. 1980; Roland 1981; Jahanshahi et al. 1995, 1995), firing rates of neurons ((Fried, Mukamel, and Kreiman 2011) (Brinkman and Porter 1979; J. Tanji and Kurata 1982)(Jun Tanji and Keisetsu 1994) and functional magnetic resonance imaging (fMRI) (e.g. Soon et al. 2008).

[72] Fried, Mukamel, and Kreiman 2011

[73] See e.g. Stevens 1993

9.2 In search of a complete microphysical explanation for the 'neuro-causal origin'

If there is a complete physical explanation of volitional action, it cannot begin with neurons in some brain region, for then the question of what made these neurons fire would still be open, and the explanation would not be complete after all. A complete explanation must begin somewhere outside the brain – not at the big bang or somewhere in outer space, to be sure, but plausibly in some region of our bodies for which Montero and Papineau's claim – that physiology has established that all physical effects have physical causes[74] – holds uncontroversially. Such a region must lie outside the brain, and thus there will be a transition (at the cellular level) from something that is not a brain cell to a brain cell (or brain cells). I shall call the first brain cell – or cells – lying in the sequence of events leading up to volitional muscle contraction the 'neuro-causal origin' (NCO). Now, how can it come about that an NCO neuron[75] increases its firing rate?

The standard way for a neuron to increase its firing rate is when it receives sufficient excitatory input from a presynaptic neuron. Consider figure 9.1:

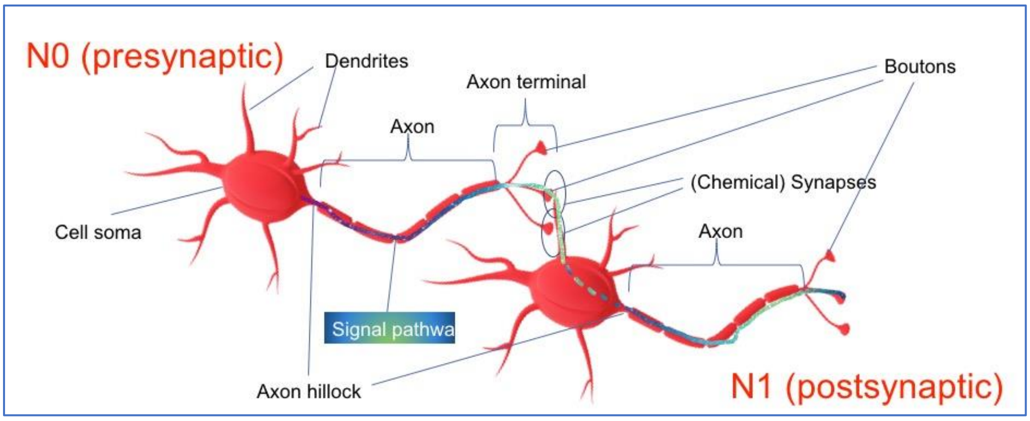

Figure 9.1: Two consecutive neurons (© Alin Christoph Cucu)

In this case, neuron 1 (N1) receives excitatory input from neuron 0 (N0). This means that N0's boutons release neurotransmitters into the synaptic cleft which then bind to N1's dendritic (postsynaptic) receptors:

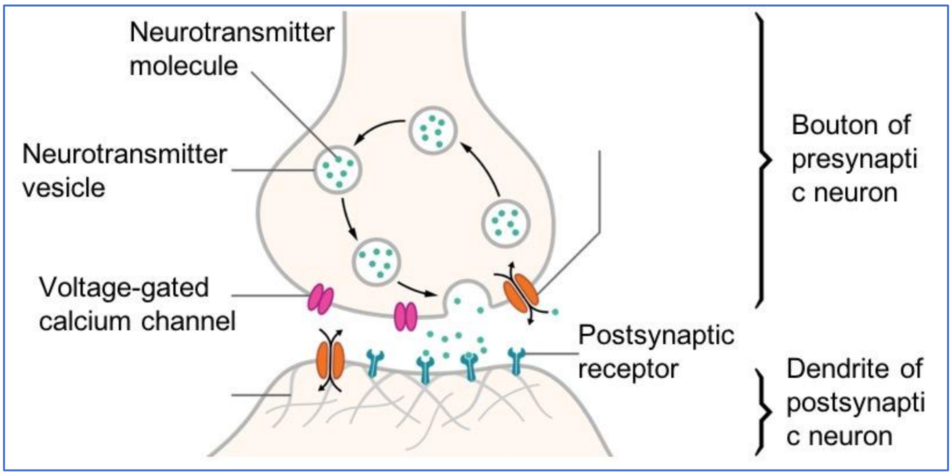

Figure 9.2: Synaptic communication (Thomas Splettstoesser – CC BY-SA 4.0; labels added)

[74] See introductory quote of chapter 2.

[75] For ease of expression, I speak of 'neuron' in the singular. The NCO might of course consist of many neurons, plausibly connected in a network-/circuit-like manner.

The binding of the neurotransmitters triggers a so-called action potential (AP) in N1, a current rapidly traveling down the axon caused by the opening of sodium and potassium (Na^+/K^+) channels. An AP in turn leads, mediated by the opening of voltage-gated calcium channels in the boutons (see fig. 9.2) to the release of neurotransmitter molecules at the boutons (into the synapse facing N2, which is not on the picture). Thus, synaptic transmission works roughly as follows: N0 fires and releases neurotransmitters which bind to N1's dendritic receptors; N1 fires, releases neurotransmitters and makes N2 fire; and so forth[76].

It is evident that this sequence of synaptic transmission cannot reach back indefinitely; it must have a beginning, that is, a genuine 'neuron zero' or perhaps several such neurons[77] (equivalent to the NCO) which do the first firings. Thus, the standard way of neuron firing brings us full circle back to the initial question, and the jury is still out on what the physical explanation of the NCO is, if there is one at all. The following hypotheses are partly mechanisms taken from the neurophysiological literature (albeit not in the context of the causation of volitional action), and partly my own ideas based on standard physics and biophysics/biochemistry. Let me make their purpose clear: they serve as 'placeholders' for the non-existing empirical research on the causal origin of volitional actions. I constructed them in order to give the physicalist a working hypothesis where dedicated empirical data are missing. To the extent to which those scenarios prove implausible, interactive dualism becomes more viable as the explanation for the causal origin.

A first natural suggestion is that the NCO is linked to a causal chain that leads outside the body. One such scenario are stimulus-driven actions: in principle, there could be an uninterrupted physical causal chain from the external stimulus (e.g. a visual or tactile stimulus) through sensory cells and afferent nerves to the cortex, and from there via efferent nerves to the muscles. For example, photons hitting our eyes could cause a conformational change in retinal proteins, which in turn cause bipolar cells to fire, which through a long chain or network of synaptic transmissions (as described above) running through the cortex to the muscles cause the arm to move. This would clearly be a perfectly acceptable, uncontroversially complete physical explanation.

However, as indicated above, such actions take a different path through the brain; by the same token, the studies underlying Haggard's overview are all 'cue-free'. Hence, this possibility is barred as long as neurophysiology does not come up with new discoveries linking volitional actions to external causes. However, this seems unlikely, since, as indicated in section 9.1, most experiments investigating volitional action lock out external causal influences.

Let's look elsewhere. I suggest another hypothesis, namely that the NCO is triggered by hormones coming from an endocrinal gland. It is well-known that hormones can act as neurotransmitters (in fact some substances have both functions, e.g. vasopressin and oxytocin). Let's call this the 'endocrinal' hypothesis. The endocrinal hypothesis requires a brain area whose neurons are directly influenced by a hormonal gland. I am not aware of any brain areas along abovementioned pathway that could fit the bill. The dopaminergic

---

[76] Of course, neurons do not just form such simple *chains*, but rather complex *networks*. But the sequentiality of synaptic transmission remains the same.

[77] There is the very real possibility that the causal origin consists of *one* neuron. Such 'command neurons' have been found in invertebrates (Stein 1978).

influence of the substantia nigra (SN) on the basal ganglia[78] unfortunately does not come into question as a candidate. The SN is not an endocrinal gland, but consists itself of neurons; also, the SN has afferences from the motor and premotor cortices, which means that while it (regulatorily[79]) influences cortical processes, it is itself influenced by the cortex. All this makes it a poor candidate for an NCO trigger.

Of course, further research might find an endocrinal candidate for an NCO trigger. However, it must be noted that endocrinal influences, being modulatory in nature, generally seem to be too slow for volitional actions[80].

A third option are pacemaker cells. Those neurons regularly self-generate APs through a cyclic mechanism of ion in- and outflux[81]. This makes them independent of external causes for AP generation. However, it is doubtful that such neurons can be found in the brain[82]; moreover, their activity is one of strict (though perhaps modifiable) regularity, which contradicts the idea of 'irregularly willed' voluntary actions. Again, the possibility of identifying such neurons at the source of volitional actions can presently not be ruled out, but is very unlikely.

The first two of the abovementioned proposals have in common that they rely on the regular receptor-mediated generation of APs as described in section 9.1. But in principle APs can be generated by other mechanisms too, as seen in pacemaker cells. Those mechanisms all involve *proteins* in some way or another. Consequently, the following hypotheses all involve conformational changes of some proteins, which basically requires energy expenditure[83] and thus at once brings back the question of energy conservation. As we know from chapter 5, energy conservation is a matter of biconditionality:

(1) Relevant Noether symmetry preserved ≡ Energy conserved

from which logically follows

(2) ~(Relevant Noether symmetry preserved) ≡ ~(Energy conserved)

which we can rewrite as

(3) [~ (Relevant Noether symmetry preserved) ⊃ ~(Energy conserved)] ∧ [~(Energy conserved) ⊃ ~(Relevant Noether symmetry preserved)]

or, by commutativity, as

[78] cf. Haggard 2008, 936.

[79] Pessiglione et al. 2006

[80] Wilson 1999, 191-92; Hille 2001, ch. 20

[81] Hille 2001, 147-49

[82] In the human body they are known to exist in the heart.

[83] For a good overview of research concerning protein conformational change in neurons see Wilson 1999.

(4) [~(Energy conserved) ⊃ ~(Relevant Noether symmetry preserved)] ∧ [~(Relevant Noether symmetry preserved) ⊃ ~(Energy conserved)]

from which follows by simplification

(5) [~(Energy conserved) ⊃ ~(Relevant Noether symmetry preserved)]

As we saw, the interaction of a non-physical mind does break the Noether symmetry, leading to (2) and logically through (3) and (4) to (5). In other words, the conclusion is that any energetical non-conservation entails that the relevant Noether symmetry has been broken. This is the inverse of the insight gained in chapter 5, where I pointed out that the breaking of Noether symmetries (which is inevitable in case a non-physical mind interacts) leads to energy and/or momentum non-conservation. In other words, whereas chapter 5 dealt with the 'top-down' hypothetical scenario of a dualistic mind interacting, we here consider the 'bottom-up' case of there being a concrete instance of non-conservation, and what that would entail.

Of course, and that is another lesson drawn from chapter 5, (non-)conservation is basically a function of the delineation of the physical system under consideration: if one draws its boundaries wide enough, then the corresponding Euler-Lagrange equation will yield zero, and everything is fine. Conversely, if no reasonable delineation of the physical system yields energy conservation, then, by the Noether theorem, something must be responsible for breaking the symmetries. The project of a complete physical explanation then becomes hard to maintain, and consideration of a non-physical cause can no longer be screened off.

I think it important to again visualize these foundations of conservation principles, because a mere local instance of non-conservation does not warrant any interactive dualist conclusion. After all, if one draws the boundaries of a physical system narrow enough, one always ends up with non-conservation. A protein which changes its conformation, considered in isolation, would be such case. But of course one needs to take into account the enzyme, electric current, ions or whatever else caused the conformational change, to be able to make a robust statement about conservation.

Now, where might a conformational change in proteins take place such that that change is conducive to increasing a neuron's firing rate?

One first option is that sodium or potassium channels open spontaneously or are caused to open in a deviating way by the binding of molecules, thereby triggering an AP[84]. Second, voltage-dependent Ca^{2+}(calcium) channels in the boutons might open without there having been a prior voltage change (i.e., without there having been an AP)[85]. (It is the calcium influx upon the opening of those channels that causes the release of the neurotransmitter vesicles from the bouton). Third, Ca^{2+} might be released from intracellular protein stores[86]. It would then have the same effect as extracellular calcium flowing in. Fourth,

[84] To be sure, there are so-called ligand-gated sodium channels whose occurrence is, however, restricted to the neuromuscular junction (cf. Hammond 2015, ch. 6)

[85] Normally those channels respond only to *large* membrane depolarizations (Hammond 2015, 151).

[86] The intracellular stores are proteins located in the endoplasmatic reticulum, the calciosome, the mitochondria and the cytosol (the cytosolic stores are lightweight proteins like parvalbumin and calbindin) (ibd., 155). A release of calcium from there occurs normally upon an appropriate signal (e.g. the formation of inositol triphosphate) through Ca-permeable channels. the proteins primarily serve as calcium-*binders* to reduce cytosolic Ca^{2+} (which is toxic in too high concentrations) (ibd., 51).

neurotransmitter vesicles might spontaneously be released from the axon terminal by exocytosis, which also requires the conformational change of some proteins[87]. The central question concerning all these options is how the respective proteins might be modified in ways that preserve energy. The following options come to mind:

1) **Ersatz ligand molecules (distinct from neurotransmitters or hormones) bind to the proteins**
2) **'Outlier' molecules with kinetic energy far above average hit the proteins**
3) **Quantum effects are responsible for the spontaneous modification of the proteins**

As to 1): There are indeed substances which activate sodium, potassium, and calcium channels[88], but all of them are pharmacologically active chemicals supplied from outside. No endogenous substance is known to modify neuronal ion channels directly. There are still two protein targets for surrogate ligand molecules: the proteins involved in vesicular release from the bouton and the proteins involved in calcium release from intracellular buffers. As regards the former option, once again no endogenous substances come into question for such an effect; exogenous substances known to trigger vesicular release are classified as strong toxins[89]. The same holds true for calcium release from intracellular buffers (except that there are perhaps no toxins causing a 'dam-breaking' of intracellular buffers).

Concerning suggestion 2), statistical thermodynamics tells us that temperature is a measure for the *mean* kinetic energy of particles and that at any temperature, and there are very few molecules far above/below that mean energy. Could not such 'outlier' particles, e.g. water molecules, be responsible for channel opening/vesicular release? They could in principle, but there are problems. First, the frequency of such events, given their low probability, seems to be insufficient to account for volitional action. Second, the approach seems much better suited for explaining the baseline firing rate, (which seems clearly too low to account for the dense volitional action patterns present in humans, and, for that matter, animals[90]); but if it explains the baseline, it cannot also explain the *increase* in firing rate.

Could the NCO be triggered quantum-mechanically (option 3)? It is important to distinguish this hypothesis from the QM-mediated accounts of interactive dualism discussed in chapters 6 and 8. Here, in exploring an energy-conserving, purely physical explanation of the NCO, we wish to know whether quantum mechanics alone, without the meddling of a non-physical mind, suffices to explain the behavior of neurons in volitional actions. Hence, we can ignore the consciousness-collapse (CC) interpretation of QM and its troubled relationship with energy conservation (see section 6.2). Still, even with CC out of the way, there are still plenty of empirical problems with quantum-mechanical approaches to brain processes (the arguments in section 6.3 against conservation-friendly interactive 'quantum' dualism hold equally against QM explanations of neuronal firings). Moreover, there is yet a further problem, namely that the

---

[87] Südhof 1995

[88] Examples include: Alkaloid-based toxins such as aconitine, batrachotoxin or brevetoxin for sodium channels (Hammond 2015, 68); diazoxide and minoxidil for potassium channels; and Bay K8644 and Ambroxol for calcium channels (Rang 2003, 60).

[89] As an example of a vesicle-release-activating neurotoxin, latrotoxins present in black widow spiders cause *all* of the neuron's vesicles to release their neurotransmitters (Ushkaryov, Rohou, and Sugita 2008). This causes extreme pain and often death.

[90] For example, in the absence of an AP, a spontaneous vesicle release in the frog neuromuscular junction is estimated to have a rate of 10^{-2}-10^{-3} times per second and release site, which means it occurs once every 100 to 1000 seconds. (Stevens 1993, 56)

frequency of quantum collapse events (independent of their vigor) does in all likelihood not match the occurrence pattern of volitional actions; this is one important lesson of chapter 8. To fix the account, one might suggest a combination of 'outlier molecules' and quantum events: quantum events explain the baseline and the 'outlier molecules' the increase. The idea is *prima facie* tenable (provided the empirical problems with quantum events in the brain can be overcome). In fact, John Eccles, one of the most prominent defenders of a QM approach of neuronal firing, is prepared to relegate quantum events to explaining the baseline rather than the increase in firing rate[91]. However, it faces the intrinsic problems of the outlier hypothesis pointed out above. All in all, even a refined quantum account does not look as if it could satisfactorily explain the neuronal firing patterns in volitional actions.

In summary, the prospects of finding an energy-conserving, protein-based account of the NCO look bleak. One would need either special ligand molecules or a non-biochemical physical explanation to arrive at such an explanation. But neither seems to be available. Thus, it is hard to avoid the conclusion that at some point, energy (perhaps along with linear momentum) will be non-conserved in the brain; very gently, to be sure, but non-conserved all the same. This invites interactive dualism, for it is hard to see what else than a non-physical mind could be responsible for the breaking of Noether symmetries in a brain.
It is of no use to rejoin at this point that if one considers the human body as a whole, energy will be conserved after all. Whether this is really is so is an empirical matter, and has to my knowledge not been investigated in a specific manner (Rubner's 1889 respirometer experiments with dogs hardly count). The non-conservations are expected to be gentle, and so a very sensitive detection would have to be employed. Also, as long as a physical explanation of the triggering of volitional actions at the micro-neuronal level is lacking, so is energy conservation, because only a physical explanation allows the formulation of a sufficiently comprehensive Lagrangian for which the continuity equation is zero.

Let's place the arguments of this chapter in the larger context of this thesis' objective. Even without the foregoing meta-empirical considerations, there are many good philosophical reasons (chapter 1) to believe in interactive dualism (and one good empirical reason to believe in substance dualism, if not necessarily its interactive form: namely near-death experiences, which are not treated in this thesis). On the other hand, the typical objections against interactive dualism are unconvincing (chapters 2-7). As we saw specifically from the response to the objection from conservation principles, the physical way for interactive dualism is conceptually paved by the conditionality of conservation principles via the Noether theorem. This overall picture, far from depicting the physical world as a monolithic block with at best vanishingly small chinks for minds to fit into (quantum indeterminacy, which, as we saw, is much more problematic than commonly thought), actually gives the impression of a puzzle for whose only gap we hold the fitting piece in our hands: interactive dualism. Of course, it must at once be admitted that if a complete physical explanation of volitional actions should be found, things would dramatically change for interactive dualism. A quantum account would then become its last resort, but even that could be partly refuted by showing that the Born rule is strictly respected. Even then, one would just have to give up a strong notion of freedom, but not interactive dualism per se, let alone dualism in general.

[91] As Eccles (1994, ch. 4, 5) seems to suggest.

Let me conclude by making a somewhat ironical observation. If the conclusion of this section is correct, then the very boomerang physicalists have thrown at interactive dualists (energy conservation) now comes circling back to hit *them*. That interactive dualism entails energy non-conservation cannot be avoided; but if energy is not conserved, then interactive dualism can hardly be avoided, either.

9.3 Neuroplasticity

I'd like to conclude this chapter by briefly discussing another phenomenon which constitutes evidence for interactive dualism, despite its being relatively neglected by philosophers of mind, at least in comparison to the 'classical' Libet-style science about volitional actions: neuroplasticity. Neuroplasticity can be described as "directed mental effort to produce systematic and predictable changes in brain function" (Schwartz, Stapp, and Beauregard 2005, 2). As such, it may plausibly count as volitional action, even if the outcomes are not limb movements. The surprising thing about neuroplasticity is that people seem to be able to alter their brain structure by the application of mental effort in cases where no such alteration is expected, be it after the removal of large parts of the brain or pathological neurological or neuro-psychological conditions.

A particularly spectacular case is that of "I.K." which Jean Askenasy and Joseph Lehmann describe at the beginning of their review paper on neuroplasticity research (Askenasy and Lehmann 2013). I.K. had a severe ischemic stroke at age 66 which destroyed almost a quarter of his brain, something believed by neurologists to be irremediable. As a consequence, the whole right side of his body was paralyzed (hemiplegia), he lost speaking and comprehension abilities (aphasia) and his ability to write (agraphia). But four years after the incident, I.K. could already play the piano with his left hand, write books and poems, and paint and exhibit paintings. Besides having an exceptionally developed right hemisphere, I.K. also was determined to regain his abilities, as he noted in a poem. In other words, he didn't give himself up, but rather undertook conscious efforts to regain his former capacities. That constitutes, as Askenasy and Lehmann put it "a top-down effect of consciousness on the neuroplasticity of his brain" (ibd., 1-2).

There is a considerable body of literature on neuroplasticity. One important research area is the change of cerebral function by self-directed regulation of emotional response via cognitive reframing and attentional re-contextualization. There are numerous reports on such effects (e.g. (Schwartz et al. 1996; Beauregard, Lévesque, and Bourgouin 2001; Ochsner et al. 2002; Lévesque et al. 2003; Paquette et al. 2003). For example, Paquette et al. could demonstrate systematic changes in the dorsolateral prefrontal cortex and parahippocampal gyrus after cognitive-behavioral therapy for phobia of spiders, with brain changes significantly related to both objective measurements and subjective reports of fear and aversion. Jeffrey Schwartz and Mario Beauregard, two of the leading experts in the field, write about patients with obsessive compulsive disorder:

> For instance, work in the 1990s on patients with obsessive compulsive disorder demonstrated significant changes in caudate nucleus metabolism and the functional relationships of the orbitofrontal cortex– striatum–thalamus circuitry in patients who responded to a psychological treatment using cognitive reframing and attentional refocusing as key aspects of the therapeutic intervention. (Schwartz, Stapp & Beauregard 2005, 3)

Another area is the systematic altering of pathological neural circuitry through conscious efforts (Schwartz et al. 1996; Schwartz 1998; Musso et al. 1999). These observations were subsumed under the term 'self-directed neuroplasticity', coined by Jeffrey Schwartz and Sharon Begley in their 2009 [2002] review

(Schwartz and Begley 2009). In general, and this further confirms the hypothesis that conscious efforts can have an influence on brain structure, the brain area activated in all studies done on the self-directed regulation of emotional response is the prefrontal cortex (Schwartz, Stapp & Beauregard 2005, 3). According to the authors just cited, a "working hypothesis for ongoing investigation in human neurophysiology, based on a significant body of preliminary data, is that the mental action of mindful awareness specifically modulates the activity of the prefrontal cortex." (ibd., 4)

Where does this leave us with respect to interactive dualism? *Prima facie*, neuroplasticity supports it, if one is prepared to conceive of the mind dualistically in the first place. Of course, neuroplasticity is not a watertight proof for interactive dualism. First, neuroplasticity is in principle compatible with a physicalist conception of the mind: if mental states are construed as physical states, one can concede 'mental'-to-physical causality without introducing a novel ontological category, and no interaction follows. Second, there may be other factors going on besides mental efforts that might account for at least part of the observed effects; for example, in the I.K. case described above the strongly developed right hemisphere of the patient plausibly explained at least part of his astounding recovery. But the number of systematic careful experiments conducted on neuroplasticity makes it clear that this evidence cannot be easily cast aside. In combination with above considerations about the underlying neurophysiological mechanisms, the phenomenon of neuroplasticity constitutes an element in a cumulative case in favor of interactive dualism and against reductive physicalism.

IV. Interactive Dualism and the Laws of Nature

10. Laws of nature as the crux of the matter

So far, we've seen that there is neither a scientific nor a philosophical reason to reject interactive dualism. On the theoretical scientific level, the objection from conservation principles fails because the Noether theorem states those principles as conditional, not categorical. Empirically, there is no evidence against interactive dualism either; to the contrary, what we know about neuronal mechanisms suggests that we cannot even explain volitional actions without recourse on an immaterial entity. On the purely philosophical side, the causal heterogeneity objection fails to give us a reason to disbelieve that causation of physical effects by an immaterial entity is possible. Finally, there is the doctrine of the causal closure of the physical world. We saw that it is not possible to convincingly ground it in physics, and so it must remain an article of faith for physicalists.

However, it is safe to say that physicalists and other non-interactionists will not be convinced even by the combined force of the arguments given so far. The aversion against interactive dualism seems to lie deeper. Where it precisely lies is not easy to say; the pertinent authors don't exactly make full disclosure. My view is that their opposition to interaction has something to do with the laws of nature. What exactly will need to be spelled out. Here is some evidence for my claim:

> Dualism violates conservation of energy. If immaterial mind could move matter, then it would create energy; and if matter were to act on immaterial mind, then energy would disappear. In either case energy—a property of all—and only concrete things would fail to be conserved. And so *physics, chemistry, biology, and economics would collapse*. Faced with a choice between these "hard" sciences and primitive superstition, we opt for the former. (Bunge 1980, 17; italics mine)

> So if substance dualism is true then it seems that one of the most fundamental laws of physics, the law of conservation, must be false. (Searle 2004, 42)

> All physical effects are fully determined[2] by *law* by prior physical occurrences. (Papineau 2001, 8; italics mine)

We've already come across the first two quotes. Although Bunge does not mention the laws nature explicitly, it stands to reason that he somehow has them in the back of his mind. For what else would physics, chemistry and biology "collapse" if not the invalidation of their most basic objects of study, namely laws of nature? Then we have John Searle who ties his rejection of dualism to what he believes to be "one of the most fundamental laws of physics". And finally, arch-physicalist David Papineau specifies the principle of causal closure (CCP) as all physical effects being "fully determined by law by prior physical occurrences". Because of this, what Papineau and physicalism ultimately need to firmly establish CCP are laws of nature which determine physical effects based on prior physical events.

Jeffrey Koperski voices this analysis as follows[92]:

[92] CCoN stands for 'causal closure of nature' and is practically equivalent to the causal closure of the physical

> As far as I can tell, we could simply drop CCoN from the conversation. Any constraint on nonviolationism *boils down to the relevant laws of nature*. Talk of closure is an unnecessary complication. (2020, 123; italics mine)

Now, how are the laws of nature supposed to be a problem for interactive dualism? One can take Searle's route: interaction is impossible, because it would violate a law of nature, and that's impossible. However, for this to work, the laws of nature would have to possess a modal force that excludes interaction, and it is unclear how this could be achieved. Imagine a stone dropped to the ground. A law of nature, Newton's law of gravitation, is involved in its accelerating in a straight line towards the earth. But although stones usually behave as described by the law, they sometimes don't, for example if there is a strong wind. Of course we would not say that the laws of nature have been invalidated (we wouldn't even say that about Newton's law), but rather that all relevant influences and their pertinent laws must be taken into account. Fair enough. Let's now take the example one step further. Suppose the stone stops in mid-air although there is neither wind nor a strong electromagnet nor any other physical influence that could explain its deviating from the expected behavior. How is this in any sense *impossible*? We can meaningfully conceive of it. Also, though it might be true that it is *physically* impossible, that doesn't mean it is impossible *tout court*, but that impossibility is what is needed to rule out *non-physical* interaction. So, the impossibility argument simply doesn't work.

The other option is to accept that mental interaction is possible, but nonetheless to be rejected because if it occurred it would *break* laws of nature. (Perhaps this is the bane Bunge is terrified of). At least in the context of *divine* interaction law-breaking is a debated issue (Koperski 2017; Larmer 2017; Koperski 2020). But whereas there is a genuine question with respect to God's intervention – namely how, if at all, the lawgiver can act in the world without breaking His own laws – the interactive dualist can retort to the worry with a shrug of his shoulders. A law is broken, so what? If mental interaction actually happened, how is that an objection? Apart from that, at least with respect to conservation laws, the Noether theorem already bars any notion of breaking: conservation laws can't be broken, because they are conditional.

Still, I am not just going to offer a shoulder-shrug in response to the 'law-breaking' worry. If there is a way to spell out interactive dualism without breaking laws of nature, then that is preferable. To do this, we must focus on the theory and notion of laws of nature; the attempts to make interaction (supposedly) law-friendly have already been shown to be failures. In the chapter 12, I will investigate what it means for a law to be broken. Before that, however, I wish to focus on the *notion* of 'laws of nature' by dint of a historical analysis, that should help us understand much of the stubborn resistance against interaction.

11. The notion of 'laws of nature': a historical survey

> For better or worse, the view that the laws of nature were simply God's decrees soon gave way to the reified version more common today (...). If a phenomenon is explained by way of natural law, many infer that there will be no need to appeal to God, thus setting the two in opposition to each other. Such a dichotomy would have been quite puzzling to those who first used the notion of laws to describe nature. (Koperski 2020, 80)

The thesis of this chapter is that the notion of 'laws of nature' has Christian origins, and that modern science and philosophy have adopted it while eviscerating it of its theistic roots, thereby bequeathing to us the almost tyrannical ideology of laws without a lawmaker and thus without exceptions. More specifically, my claim is that labeling natural regularities 'laws' is a product of the influence of voluntarist theology on 17th century science. Before that, virtually all science that had been done in Europe was inspired and influenced by a mix of Christian and Aristotelian thought in which the notion of 'laws' barely ever emerged; after the 'voluntarist revolution', a slow drift away from the theistic roots of 'laws of nature' was followed by a landslide set off mainly by Thomas Huxley in the second half of the 19th century, after which the nowadays ubiquitous dichotomy of 'science vs. religious myths' became the standard paradigm.
The realization that analyzing nature in terms of laws is a predominantly theistic idea opens up new alleys for rethinking the relationship between free agency and natural regularities: if there is a lawmaker, then there is the possibility for exceptions from the laws for special purposes, for the lawmaker is the highest instance with respect to the enactment of the laws. However, it might also provide remedies for the diverse problems the secular accounts to be discussed in chapter 13 face.

11.1 Before the 17th century

In antiquity and the Middle Ages, the term 'law' was very seldom used in connection with nature, and often in a way clearly different from our modern scientific conception[93]. Jane Ruby characterizes the ancient Greek era as featuring a "sharp distinction between *nomos* as a human conception imposed on nature and *physis* as the natural order" (Ruby 1986, 354). She specifies further:

> [T]here survive from before the Christian era but two instances of use at all comparable to that of modem science: Plato's assertion that the sick body goes against the laws of nature, and the second-century Peripatetic Kritolaos's declaration that because of "those immutable statutes, the laws of nature," the Myrmidons could never have sprung full-grown from the ground. (ibd.)

Especially the thinking in terms of the *physis-nomos* distinction makes it clear that the ancient Greeks would positively have rejected our modern understanding in which *nomos* applies to *physis* or is even part of it.
The Roman era (including early Christianity) yields more interesting occurrences of the use of 'law' (Latin *lex* or Greek *nomos*). Edgar Zilsel claimed to have found early approximations to our modern use of the term – occurrences of "prescientific natural law", as he calls them – in Ovid, Seneca and the Christian orator Arnobius (c. 300 A.D.) (Zilsel 1942, 260). Jane Ruby writes that in ancient Roman literature, the

[93] This may sound odd, since above I expressed my view that the precise content of the concept of laws of nature is often unclear. How could one then state a difference between two things if one of them is ill-defined? As will become clear, the ancient and medieval use of 'law' in connection with nature is evidently one that mainstream science and philosophy would nowadays emphatically deny. And that of course is at once a starting point to delineate what we mean by the concept today.

term *regula* (“rule” or indeed “ruler”) is more frequent than *lex* (or *nomos*), but that both are used in a sense more like ‘guideline’ than ‘law’. For example, Cicero wrote "Who does not know that the first law [lex] of history is not to venture to say anything false?"(De Oratore, 2.62.); Seneca the Elder (53 B.C.-39 A.D.) speaks of those who "reduced the course of the seasons to a fixed law" (*ad certain legem redegerunt*); Seneca the Younger (ca. 4 B.C.-65 A.D.) of comets as "not yet grasped in fixed laws (*nondum teneri legibus certis*)"; and Manilius (1st century A.D.), in his astrological poem, the *Astronomica*, of "following a fixed law (*certa... lege*) in seeking information about individual signs." (Ruby 1986, 347-348)

In general, Roman philosophy, at least during the Imperial Era, was dominated by Stoicism, whose cosmological natural law was not just “at once physical and ethical, descriptive and prescriptive, but (...) almost never associated (...) with any physical phenomenon smaller than the cosmos itself.” (ibd., 354) By contrast, laws of nature are today understood to be purely descriptive, applying exclusively to the physical realm, and certainly also to parts of the cosmos (while a ‘super-equation’ capturing the whole universe is conceivable; think of a quantum-mechanical equation describing the wave function of the whole cosmos).
Medieval natural philosophy was clearly dominated by Aristotelian thought, which explained nature by the powers and liabilities of substances, determined by the substances’ forms. This often (but not always) went hand in hand with a refusal to apply ‘law’ to non-human creation. For example, Thomas Aquinas held that what he calls ‘natural law’ is to be limited to the human realm, designating its use for non-human creation as metaphorical[94]. Bonaventure at least granted natural law to pertain to the sentient realm in general[95], while Alexander of Hales was even more generous and granted its application even to non-sentient creation[96]. The closest one can get to a modern notion of laws of nature in medieval literature is in Roger Bacon (c. 1210-c.1292) and Robert Grosseteste (1175-1253). Bacon uses *lex* frequently, while the most interesting occurrences are phrases like “the law of refraction”, not so much the expression “laws of nature”; the latter is used in a somewhat unclear way, while the former clearly targets a specific phenomenon (Ruby 1986, 346). Grosseteste uses *regula* rather than *lex*, but the context shows that he began a “shift from forms to laws” (ibd., 343). Ruby goes even so far as to argue that Grosseteste and Bacon were the real lynchpins in the turn from Aristotelian powers metaphysics to the modern understanding of nature in terms of laws, and that this paradigm shift had nothing to do (apart from astronomy) with theological considerations. Her main reason to think so is that especially Bacon evinces a great deal of mathematization; many instances of his use of *lex* are mathematical, and this mathematics is then used in his optics. I am going to disagree with Ruby’s analysis shortly; for now, let us just also note that the supposedly opposing analyses of nature in terms of forms vs. in terms of laws need not be as irreconcilable as generally assumed (for an argument, capitalizing on Bacon’s philosophy, that the two actually go well together, see (Kedar and Hon 2017).
At any rate, Grosseteste and especially Bacon are the only two philosophers in an almost millennium-long era whose idea of laws of nature comes in any way close to the modern use. William of Conches and Bernard Silvester used ‘law’ freely but not specifically (Ruby 1986, 344). Hermann of Carinthia used *habitudo*, meaning our knowledge “*qua lege*” things occur. None of them had in mind anything like the

---

94 *Summa theologiae*, 2-1. 91. 2 ad 3; 90, 1.

95 *In primum Sententiarum Petri Lombardi*, 4.33.1.1, *Opera omnia* (10 vols.; Quaracchi, 1882-1902), IV, 747-48·

96 *Summa theologiae*, 3-2.1.3.4 (Quaracchi, 1924-), IV, 338-39.

mathematical characterization of physical reality. Then, in Greek and Arabic translations much in use from the 13th century onward, the equivalents of 'law' (lex, nomos etc.) hardly ever appear (Ruby 1986, 344-5). Even 16th century Jesuit philosopher Francisco Suárez, who did use the term "natural law" (*lex naturae*) in relation to physical nature, insisted that acceptation of the term 'law' is "metaphorical since things which lack reason are not capable of obedience" and that "those very common laws [of nature] established in the universe" reflect nothing other than God's own action in accordance with "the ordinary law which he has imposed upon himself" by his "regulated power" (Oakley 2019, 23-24). Ruby sums up the situation in late medieval and Renaissance philosophy as follows:

> [I]n the thirteenth century Thomas Aquinas, followed by Francisco Suarez in the seventeenth, declared the use of "law" for the inanimate metaphoric; in the fifteenth Lorenzo Valla found it ridiculous to speak of law with reference even to animals; and in the sixteenth Pietro Pomponazzi argued that because natural events are not responses to words or persuasion, it is meaningless to speak of "law" in nature. (Ruby 1986, 341)

When did the modern notion of laws of nature then finally gain traction? It was in the 17th century with figures like Newton, Boyle and Descartes, and it was under the influence of voluntarist theology, as will be explained in section 11.2. It should, however, be noted at this point that a number of historians of science disagree with that view. First, there are the well-known, almost 'classical' but nonetheless unconvincing socio-political accounts of Edgar Zilsel (1942) and Joseph Needham (Needham 1951). Both agree with the theological voluntarist approach that the idea of nature having laws sprung from the concept of a divine sovereign lawmaker, albeit not inspired by theological doctrine but by the rise of absolute monarchy in Europe – a view which, as both Henry (Henry 2004) and Oakley (Oakley 1961; 2019) agree, entirely lacks historical evidence and is thus little more than a plausible-sounding narrative. A more persuasive theory stems from Jane Ruby. She holds that "[s]ixteenth-century astronomers all drew on the works of Regiomontanus, and after 1543 almost all if not all, whether or not they accepted the heliocentric hypothesis, on the *De Revolutionibus*. For the most part they used *lex*, as Regiomontanus and Copernicus did, without reference to God." (Ruby 1986, 355). For her, "all the familiar modern scientific uses of "law" were in place" by 1540, albeit without yet having their eventual currency (ibd., 357). According to Ruby's reconstruction, Johannes Kepler (1571-1630) used 'law' a number of times without it being clear whence he got the term. Earlier, several ablative versions of *lex* can be found in Regiomontanus (1436-1476), but they probably do not mean what came later to be understood by 'laws of nature' (ibd., 352-53). What is certain is that Copernicus (1473-1543) took up the term from Regiomontanus and apparently used it in a more principled sense (ibd., 354). However, according to Ruby, he "did not (…) follow Manilius in asserting that God brought the whole universe under law." (ibd., 355).

Taken together with the above observations about Bacon and Grosseteste, Ruby's thesis is that the notion of laws of nature had medieval roots, albeit more in mathematics than in theology. A similar case is made by John Henry (2004), who locates the origins with Descartes rather than with Bacon: "Descartes took his concept of laws of nature from the mathematical tradition, but recognized that he could not export it to the domain of physico-mathematics, to play a causal role, unless he could show that these laws were underwritten by God." (73) (In a later paper (Henry 2009), Henry recognized the influence of voluntarism on the early modern scientists). In other words, according to Henry, Descartes imported God into a conception primarily drawn from mathematics to endow it with causal potency, rather than exploring the mathematizability of a pre-existing theological conviction.

As far as I can see, both Henry's and Ruby's accounts, for all their valuable insight, fail in one crucial point: they underestimate the background influence of the Christian worldview on the thinkers in which they do not detect explicit reference to God as regards laws of nature. The Christian worldview was prevalent throughout the Middle Ages and the Early Modern Period, so it stands to reason (and is in most cases beyond reasonable doubt) that the quoted figures had internalized it. And since the Christian worldview asserts as one of its central tenets that God is the Creator of the seen as well as the unseen world, it is simply implausible to hold that believers in this worldview would not ultimately trace back laws of nature to Him. Thus Francis Oakley on Descartes: "[I]t would seem more probable *a priori* that he was drawing on a theological rather than a political tradition. This probability is heightened by the fact that he was, after all, a devout Christian whose religion was so closely connected with his scientific thinking that Robert Boyle could comment to the effect that atheism "would subvert the very foundation of those tenets of mechanical philosophy that are particularly his. "" (Oakley 1961, 438)

To be sure, the respective writers may not have made the reference to God explicit, because they focused on another aspect (e.g. mathematical rules in optical phenomena), and plausibly because it was simply taken for granted by their audience. But then there are exceptions like Tycho Brahe (1546-1601) who did make explicit what is most likely his belief in God as the ultimate source of the laws (Ruby 1986, 356). Like Henry, Ruby neglects the voluntarist theology, which came into fashion in particular after the Condemnation of 1277 and which was championed by Descartes (and also Newton). In the Condemnation of 1277, the Bishop of Paris, Stephen Tempier, along with Robert Kilwardby, Archbishop of Canterbury, prohibited the teaching of over 200 philosophical and theological theses that were being discussed and disputed in the faculty of arts under their jurisdiction. Some of those theses had been put forward by Aquinas. The main concern of the Condemnation (or at least one of the main concerns) was the general fear of the influence of Aristotelian philosophy on theology, and more specifically of Aristotelian necessitarianism in metaphysics. Expressed positively, a main drive for Tempier was to affirm the absolute power of God to do whatever He chooses. That absolute power is clearly limited if, to pick an example relevant to our purposes here, physical objects have their own intrinsic causal powers as opposed to being wholly governed by God's external imposition. Theological voluntarism is the view that God's power is unrestricted (save perhaps by logical contradictions, but Descartes went so far as to deny even that). Although voluntarism is typically discussed in an ethical context (where it means that moral truths are nothing but God's decrees), the concept can be and has been equally applied to the workings of nature.

That theological influence must also be considered when appraising Newton's shift away from the "plausible opinions" of Scholastic philosophy on pain of making him sound like one who exorcises occult qualities with mathematics. This critical influence will be the focus of the next section.

11.2 The paradigm shift of the 17th century

The paradigm shift of the 17th century regarding the foundations of science is summed up by historian of philosophy Eric Watkins as follows:

> By placing laws at the foundation of scientific inquiry (in the broader context of the Scientific Revolution), modern thinkers redefined the order of nature, which had previously been conceived of (by medieval thinkers such as Aquinas, Scotus, Ockham, and Maimonides) as based on the natures and causal powers of finite substances. (Watkins 2013, xxvi)

How are the "natures and causal powers of finite substances" supposed to be replaced by "laws"? The following quote by Isaac Newton, one of the key figures of the paradigm shift, helps us understand:

> Such occult Qualities put a stop to the Improvement of natural Philosophy, and therefore of late Years have been rejected. To tell us that every Species of Things is endow'd with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing. . . . (*Opticks*, Query 31)

For Newton, Aristotelian substantial forms (essences) are "occult" – hidden, unobservable – and lack explanatory power ("…tell us nothing"). Other detractors of Aristotelian essences who made similar points include Pierre Gassendi (Fisher 2014, sec. 3), who complained that we cannot have epistemic access to them, and Robert Boyle, to whom they were simply unintelligible (Anstey 2002, 23-24).

Another important worry was that substantial forms constitute superfluous intermediaries between God and physical entities (Henry 2009, 93). God, the sovereign ruler of the universe, does not need such 'viceroys' to govern the world, so the complaint went.

What came to supplant Aristotelian forms in mainstream thinking was the notion of laws of nature, with God as lawmaker (Koperski 2020, ch. 4). Though the exact metaphysical nature of that conception of laws is sometimes unclear (see (Koperski 2021, fn. 8) it remains safe to say that the main theological rationale for replacing forms with laws was to make God's rule over the world more immediate; in other words, it is God's decrees understood as laws that bring about the regular workings of nature. As Isaac Barrow, Newton's predecessor as Lucasian Professor of Mathematics wrote:

> God uses no other means, instruments or applications in these productions, than his bare word or command. (Barrow 1885, 303)

Samuel Clarke, one of Newton's chief defenders held that the regularities in nature are "truly and properly speaking, . . . nothing else but the will of God producing certain effects in a continued, regular, constant, and uniform manner" (Harrison 2013, 143). Some early Newtonians took gravity to be the clearest example of such manifestations of the divine will. For example, William Whiston called gravity God's "general, immechanical, immediate power" (Brooke 1991, 146). A further champion of the new view, often considered the most important and explicit one, was René Descartes. He spoke of "the laws which God . . . put into nature" (Descartes 1841). But what spawned this emphasis on divine sovereignty in 17th century natural philosophy? It cannot be a turn towards Christian theism, for that had been the prevailing worldview for well over a millennium already. No, it had to be a shift *within* the Christian tradition. As Francis Oakley (1961, 2019), Jeffrey Koperski (2020, 2021) and Eli Lichtenstein (Lichtenstein 2022) consistently argue, it was a shift from *intellectualism* to *voluntarism*. For example, Lichtenstein argues that the "voluntarist tradition in theology made a deep impact on early modern thinking about laws of nature" (4), and Oakley concludes: "[T]here can remain little room for doubt (…) that the voluntarist conception of natural law attained a wide currency in the sixteenth and seventeenth centuries." (1961, 449)

Roughly put, voluntarism says that God can freely do whatever He wants and is not restricted by His own nature, while intellectualism insists that God makes choices based on His perfect reason and essential goodness. The voluntarism-intellectualism debate first emerged as a matter of moral philosophy. Brian Leftow characterizes the contrast between the intellectualist Thomas Aquinas and the voluntarist John Duns Scotus as follows:

> For Aquinas, all ten Commandments are necessary moral truths, which God sees and proclaims. For Scotus, necessarily, loving God is right and hating him wrong. . . . All other acts, though naturally good or bad (perfecting or destroying our nature), are right or wrong contingently, because God commands or prohibits them. (Leftow 1998)

These theological doctrines were revived in 17th century as a matter of natural philosophy. The 'neo-voluntarists' believed that the laws of nature were a product of God's unbound and sovereign will and thus could have been completely different.

> [The] laws of motion . . . did not necessarily spring from the nature of matter, but depended upon the will of the divine author of things. (Boyle 1725, 2:245)

It should also be mentioned that intellectualism was typically linked to Aristotelian/Thomist philosophy. In the case of natural philosophy, this meant that God had endowed physical entities with substantial, intelligible essences from which their behavior springs. The consequences were twofold: first, that God had delegated the bulk of work in nature to (as Aquinas called them) *secondary* causes, namely the forms of substances; and second, that the study of nature could at least in part be done by reflecting *a priori* on the essences of substances. It is safe to say that the paradigm in natural philosophy, up to the 17th century was an Aristotelian/Thomist intellectualist one[97]. What, then, propelled the thorough turn towards voluntarism whose repercussions we still feel today?

Again, Oakley, Koperski and Lichtenstein agree in that a crucial event for this development was the Paris Condemnation of 1277 which prohibited (among other things) the teaching of Aristotelian doctrines about divine freedom. Oakley:

> The year 1277 may be suggested as the overt starting point of the new tradition. It was in that year that Etienne Tempier, Bishop of Paris, and Robert Kilwardby, Archbishop of Canterbury, formally condemned, as contrary to the Christian faith, a host of philosophical propositions, including some put forward by Aquinas. (1961, 438)

Oakley also explains the motivation of the condemnation:

> Behind these condemnations lay the fear, widespread also among the more orthodox Arab and Jewish thinkers, that the metaphysical necessitarianism of Aristotle and his Arabic commentators, Avicenna (980-1037) and Averroes (1128-1198) endangered the freedom and omnipotence of the Semitic and Christian God. (ibd.)

It is well worth dwelling on the consequences of this edict. What may look like the settling of another 'ivory tower' theological debate had deep impact on the interplay of theology, philosophy and ultimately science, as well as on our understanding of the world around us and of us as part of it. If indeed God structured nature by imposing laws on it, rather than by bestowing stable natures to things (laws whose content we can no longer grasp by any purely rational effort) then the ultimate intelligibility of the world suffers a heavy blow. For it remains unclear why God chose the laws He did in fact choose – neither do they flow from any necessary truths (i.e., God could have decreed them differently), nor from the natures

[97] With the possible exception of Renaissance and early modern astronomy, which seems to have thought in different terms from its very outset. As indicated earlier, the two champions of law language among distinctly medieval natural philosophers – Robert Grosseteste and Roger Bacon – did not forsake talk of substantial forms altogether (see Ruby 1986, Kedar and Hon 2017).

of things. Moreover, the relatively fruitful integration of Greek and Judeo-Christian thought, as practiced through antiquity and the Middle Ages, received a significant crack. As Oakley puts it:

> [The] compromise which had united a transcendent Semitic Creator-God with an intelligible Hellenistic world was abrogated. (1961, 439).

And these are just the direct, short-term consequences. The indirect, long-term consequences are felt to this day. The voluntarist natural philosopher could console himself over the loss of intelligibility of the natural world with the thought that the fact of the laws being decreed by God assures their rationality (for even voluntarists would not jettison God's perfect rationality), even if we have no means of grasping it. Modern scientists, after having severed laws of nature from their (voluntarist) theistic roots, have just that – the laws, without (unless they're Aristotelians or Thomists) them being grounded in anything. That is hardly a fertile soil for making ultimate sense of the world. Add to this the emphasis on 'lawhood' and you get that rigid, mechanical, unquestioned tyranny of an unexplained order in nature that is to be kept intact at any price.

But back to our historical inquiry. How did the Condemnation of 1277 influence subsequent scholarship? Plausibly enough, by favoring philosophy that eschewed the blacklisted doctrines. And this had to be philosophy that emphasized God's unrestricted freedom and power. William of Ockham was, along with John Duns Scotus, one of the foremost proponents of such an approach. Although he worked in ethics, he introduced the distinction between God's *potentia absoluta* (absolute power) and God's *potentia ordinata* (ordained power). The former comprises everything except logical contradictions, while the latter encompasses only the actions God can do *within the confines of the order He has imposed on Himself*. In ethics, this order would comprise facts like "Murder is wrong" and "Adultery is wrong". Since God sticks to His own order, He will not murder anyone and require married couples to stay faithful to one another. What He could do, though, at least in principle, is to upend the whole order. Things now wrong would become right, and vice versa. Applied to laws of nature, the voluntarist doctrine says that by His *potentia absoluta*, God could have imposed completely different laws, but once he has established a certain order, then by His *potentia ordinata* He will respect the restricted possibilities allowed by that law[98].

This voluntarist tradition came to be known as the nominalist *via moderna* movement with its centers Paris and Oxford. Notable followers were: Pierre d'Ailly (1350-1420), Jean Gerson (1363-1429), Robert Holcot (d. 1349), Gabriel Biel (d. 1495), Jacques Almain (d. 1515), John Major (d. 1540), and Alphonse de Castro (d. 1558). The influential Jesuit philosopher Francisco Suárez (1548-1617) cited all of them as supporters of the voluntarist theory; also, his own natural law thinking bore the strong impress of their point of view.

How exactly voluntarist thought percolated to Descartes, Newton and Boyle is secondary. What should have become clear is that voluntarism was an influential school of thought, at least for some time encouraged by episcopal edict and advocated by many eminent philosophers. It was thus easy for the

98 Miracles would usually have been understood to be admissible even under *potentia ordinata*. An explicit reference is found in Jacques Almain (d. 1515) who speaks of "the common law in the absence of a miracle" (Jacques Almain, *De penitentia … lectura*, in *Aurea clarissimi et acutissimi Doctoris theologi Magistri Jacobi Almain Senonensis Opuscula* (Paris, 1518), f. 13r).

‘founding fathers’ of modern science to come across voluntarist literature (although, to my knowledge, none of them makes any reference to the medieval voluntarists). One thing should be noted, though. The mere availability of voluntarist ideas cannot explain their eventual success. For there is *prima facie* no compelling reason to give up Aristotelian/Thomist metaphysics altogether, assuming that by Newton’s time no public pressure for natural philosophers to abstain from Aristotelian hylomorphism would be in place anymore. It is probably through the arguments given by Newton and Boyle and their massive influence even on people outside their field that the voluntarist conception of laws of nature eventually gained purchase.

A last issue before we move on to the further development right up until our time. There is one question that strongly suggests itself: isn’t the voluntarist conception of laws of nature tantamount or even identical to occasionalism? Occasionalism is the doctrine that every effect is directly brought about by God, without secondary causes playing any role. Remember Isaac Barrow’s statement that “God uses no other means (…) than his bare word or command” or Boyle’s contention that the “laws of motion . . . did not necessarily spring from the nature of matter, but depended upon the will of the divine author of things”. Couldn’t this be interpreted as God bringing about all the natural events directly, with the laws being simply the *a posteriori* knowable pattern of divine activity? Although probably most neo-voluntarists did not understand their view as occasionalist (or else there would have been no room for the distinct minority position of occasionalism), the discernment cannot always be made with certainty. Lichtenstein’s summary of the voluntarist position reflects this complicated relationship:

> In contrast to current philosophers of science or metaphysicians who defend so-called ‘governing’ accounts of laws of nature, then, 17th-century voluntarists like Newton and Boyle thought that abstract patterns are the wrong kind of thing to govern, constrain, or produce physical effects, in themselves. Rather, an early modern voluntarist might see abstract patterns simply as *summarizing the impact of laws taken as acts of Will*: laws govern nature only insofar as they are free exercises of power, not just intelligible rules. To the extent that they can be viewed as intelligible rules, laws are like the abstracted ‘meaning’ of a painting, not like the pre-conceived purpose of a tool or the geometric essence that a visible line imperfectly copies. The artwork is not a copy of its meaning—the artwork grounds its meaning, not vice versa. Just so, nature grounds its intelligible form. Viewed as intelligible patterns, *laws describe but do not govern nature*. But viewed instead as acts of Will that give rise to a world displaying these patterns, laws do govern nature. God governs nature in virtue of the *act of Will through which God creates nature*. (Lichtenstein 2021, 8, italics mine)

The first emphasis sounds quite occasionalist, if taken together with the denial of laws of nature governing anything. The second highlighting even adds a tinge of Humeanism while further buttressing the occasionalist thesis: if laws don’t govern but just describe, and substantial forms are excluded, then what could possibly bring about the events in nature but God’s direct action? The third highlighted passage, however, suggests a non-occasionalist reading of the previous ones. God’s action could simply be understood as one all-encompassing act of will through which he made all nature; He would then not be bringing about every single event by itself. However, if neither God nor substantial forms bring forth physical events, what does? This is indeed a tough question for any voluntarist account of the laws of nature. For the remainder of this chapter, we will turn to the developments after the era of Newton, Boyle and Descartes, which eventually led to the naturalist laws-of-nature paradigm we find nowadays.

11.3 The naturalization of the laws of nature[99]

How did we eventually arrive at the present situation, in which the search for laws of nature has become part and parcel of science (in particular, of physics), albeit, paradoxically, accompanied by an emphatic denial of any connection to God? Given the previous analysis, it stands to reason to ask:

> How can scepticism about the existence of God *not* entail scepticism toward the laws of nature? (Koperski 2021, 69)

This is because putting laws at the center of the scientific endeavor was so tightly linked to the theological doctrine of voluntarism that one can legitimately wonder how denial of the lawmaker did not entail denial of the laws. At first, though, the voluntarist (or, as Jeffrey Koperski calls it, 'decretal'[100]) view of the laws of nature continued to be popular way into the 19th century. However, several important changes of perspective slowly crept in.

First, by the mid-eighteenth century, physicists[101] began to invoke laws of nature and other related principles with less dependence on theism. Leibniz with his stress on the Principle of Sufficient Reason (PSR) and the Law of Continuity was an exception. The Principle of Least Action, by contrast, crucial for the development of differential equations (see section 5.2), was invoked without much bothering about its source. (Which of course does not imply that anyone using the principle in that way was an atheist or agnostic). Laplace was particularly up-front about "having no need for this hypothesis [God]" (Herschel 2013, lxi) when developing dynamical equations. Beyond Laplace, France came in the late 18th to be the center of a secularized view of Newtonian physics, with d'Alembert and Voltaire being leading figures. We shall meet this secular distortion of Newton again when dealing with Huxley and the X-Club.

A second strand was the specialization and mathematization of physics. It allowed scientists to start with mathematical laws and principles of physics without heeding their divine origin. Thus William Whewell:

> Such persons are not led by their pursuits to any thing beyond the general principles, which form the basis of their explanations and applications. They acquiesce in these; they make these their ultimate grounds of truth; and they are entirely employed in unfolding the particular truths which are involved in the general truth. Their thoughts dwell little upon the possibility of the laws of nature being other than we find them to be, or on the reasons why they are not so. (Whewell 1856, 206)

In other words, the stance of those scientists was, to use a contemporary slogan widespread among quantum physicists, "Shut up and calculate".

It suggests itself to think that David Hume, whose work falls mainly in the mid-18th century, was a further major influence. However, his skeptical views of causation and by implication of laws of nature did not hold much sway over science immediately. They would play an important role in the secularization of the laws of nature in the late 19th century, though (see below).

A third influence was a change of perspective regarding laws of nature, even if still happening within the theistic tradition. Instead of seeing laws of nature as part of a broader theistic worldview, as had the pioneers of the 17th century, they became the basis for a sort of design argument for the existence of God, both for

---

[99] This section is in large part a synopsis of Koperski's 2021 paper.

[100] Koperski 2020, 2021

[101] The term is anachronistic, of course, but the relevant people (then called 'natural philosophers') did work on topics we now recognize as physics.

theological conservatives and deists (Ruse 1975). The following quote by mathematician Baden Powell expresses this new viewpoint well:

> And this is, perhaps, of all others the reflection which, to a thinking and philosophic inquirer, tends most to exalt his ideas of the Divine perfections – the regulation of all the varied and complicated actions of the material world by an unvarying system; . . . the sufficiency of a few simple laws to regulate the entire complexity of the vast mechanism; the first constitution of the world upon a principle which, without further interposition, contains within itself the means of perpetual renovation and stability. (Powell 1838, 155-156)

Note Powell's claims that what most exalts the ideas of divine perfections is "the regulation…by an unvarying system" and that all happens "without further interposition". We already have here the kernel of what would eventually become the 'de-deified' version of laws of nature we are familiar with today. Further to this contributed the doctrine of uniformitarianism. Its core thesis was that the physical world had been shaped exclusively by unchanging laws of nature – a strict uniformity in terms of causation. Its most eminent defender, Charles Lyell, argued:

> Many appearances, which had for a long time been regarded as indicating mysterious and extraordinary agency, were finally recognised as the necessary result of the laws now governing the material world; and the discovery of this unlooked-for conformity has at length induced some philosophers to infer, that, during the ages contemplated in geology, there has never been any interruption to the agency of the same uniform laws of change. (Lyell 1889, 88)

It should be noted that Lyell was a theist (or, more exactly, a deist). In a way, the exclusion of interventions (the focus back then was on divine interactions, not mental ones) had already come to be in place in the minds of uniformitarians and other deists. In other words, what was rising was a *theistically argued* non-interventionism. But the laws of nature were still traced back to God.
Finally, although being more of a background development, there was the distinction between *fundamental* and *empirical* laws of nature. It was first made by John Herschel (Herschel 1851, chap. VII, paras. 212-213) and also adopted by William Whewell. Charles Darwin also accepted it, as can be seen by his quoting Whewell at the front of *Origin of Species*:

> With regard to the material world, we can at least go so far as this – we can perceive that events are brought about not by insulated interpositions of Divine power, exerted in each particular case, but by the establishment of general laws.

Later, Darwin dropped the 'divine designer' inference, while maintaining talk of laws of nature.
The biggest landslide towards the naturalization of the laws of nature took place in the last third of the 19th century. Its leading force was the so-called British 'X-Club'[102] which massively pedaled what was nothing short of an ideological and cultural subversion. In Ronald Numbers's words:

> The relatively smooth passage of naturalism turned nasty during the last third of the nineteenth century, when a noisy group of British scientists and philosophers . . . began insisting that empirical,

[102] Its members were Thomas Huxley, George Busk, Joseph D. Hooker, Herbert Spencer, William Spottiswoode, Edward Frankland, Thomas A. Hirst, and John Lubbock.

naturalistic science provided the *only* reliable knowledge of nature, humans, and society. (…) Secularization was their goal; science, their weapon. (Numbers 2008)

The X-Club pursued a twofold strategy, political and rhetorical. On the political side, they successfully put fellow 'scientific naturalists' in charge of the university teacher examinations in new state schools (Stanley 2019, 106). Rhetorically, they attacked non-naturalistic views about science and the laws of nature in a vitriolic manner, often setting up strawman versions of their opponents' claims. A particularly telling example is Thomas Huxley's review of Robert Chambers's 1844 book *Vestiges of the Natural History of Creation* as portrayed in Koperski 2021, 71-74. What probably stuck most with the minds of the readers was a false dichotomy between (modern, enlightened, progressive) science (and laws of nature) on the one hand and (archaic, obscure, retrograde) folk belief in direct divine action on the other hand.
Huxley's obsession with the uniformity of nature went even so far as to reject the idea of evolving species, thereby setting himself (at least at first) in opposition with Darwin's newly developed theory of evolution. Huxley's almost fanatic hate for everything reminiscent of divine action went so far that he claimed that even slow, lawlike change was an indicator of teleology, which he sought to have banned from nature altogether. Still, the uniformity of laws, even their *absolute* uniformity, was compatible with theism and thus a problem for Huxley. Somewhat paradoxically (given their strong connection to theism) the prominence of the laws of nature was exactly the soil in which Huxley could breed the desired separation of natural order from theism:

> [A] natural theology of the material world in which the concept of physical law had a high profile could easily become the Trojan horse for a subversive scientific naturalism. (Brooke 1992, 87)

No doubt to avoid talk of God, Huxley endorsed something like a Humean understanding of laws of nature according to which laws do not bring about anything:

> When we have made out by careful and repeated observation that some- thing is always the cause of a certain effect, or that certain events always take place in the same order, we speak of the truth thus discovered as a law of nature. Thus it is a law of nature that anything heavy falls to the ground if it is unsupported. . . . In fact, everything that we know about the powers and properties of natural objects and about the order of nature may properly be termed a law of nature. But it is desirable to remember that which is very often forgotten, that the laws of nature are not the causes of the order of nature, but only our way of stating as much as we have made out of that order. (Huxley 1890, 12-13)

This view created an almost ironic difficulty for him. If indeed laws do not cause, hence do not assure, the order in nature – and Huxley of course utterly rejected God as taking this role – then the assumption that nature will tomorrow behave just as she has done yesterday is ungrounded. This is precisely what Hume had identified as the problem of induction. Huxley, however, charged only the antagonistic belief in miracles with this problem, not his own view. Still, he had to come up with even more to displace the theistic notion of laws of nature.
His move was to reject the distinction, made by Herschel and Whewell, between fundamental and empirical laws. Characteristically for his 'bulldoggish'[103] rhetoric, he attacked belief in fundamental laws as "pseudo-science" and mixing "the black of dogma and the white of science into the neutral tint of what they call liberal theology" (Huxley 1902, 82). But in order to get rid of fundamental laws, he had to

[103] Huxley is famous as „Darwin's bulldog".

reinterpret Newtonian mechanics, on which Herschel and Whewell had grounded their case for fundamental laws. Again, Huxley set up a strawman (of Newtonian mechanics), only this time to adopt it as its ally. His most effective technique, though, probably was the false dichotomy mentioned above, in combination with a warfare metaphor: enlightened ("white") science has been at war with obscure ("black") theological dogma – which as its most powerful weapon wields a slavish literal interpretation of the Bible – and won the palm. Sure, Huxley would grant a choice between those two paradigms, but who would want to be on the loser's side?

Others took up the thread and promoted the false dichotomy/war metaphor strategy, for example the American scientist and later historian John W. Draper, or John Tyndall in his Belfast address to the British Association for the Advancement of Science. Draper's claim that only two "interpretations…may be given of the mode of government of the world. It may be by incessant divine interventions, or by the operation of unvarying law" (Huxley 1875, 228) is reminiscent of the false dichotomy, while Tyndall's quote is dripping with belligerence:

> [As] science demands the radical extirpation of caprice and the absolute reliance upon law in Nature, there grew with the growth of scientific notions a desire and determination to sweep from the field of theory this mob of gods and demons, and to place natural phenomena on a basis more congruent with themselves. (Huxley 1874, 8)

Huxley further popularized his view of strict naturalistic laws in the highly successful book *Physiography* (1877). He traced the water of the Thames through the water cycle to solar evaporation, without losing one single word about God or His wisdom. This was in clear and intentional contrast to other science popularizers who argued for divine harmony in the natural world (Desmond 1997, 484-485). In a way, Huxley was using a narrative strategy: telling a persuasive, consistent story which appeals to the reader on a more aesthetic level and circumvents criticism by leaving critical things out.

Why was there not more pushback from Christians and theistic scientists and philosophers? One major reason seems to be that "the X-Club and their allies used all the same language that theists had used for two centuries" (Koperski 2021, 81). They were speaking of laws of nature and uniformity, terms in use also by theistic scientists like Lyell, Herschel and Thomson (Lord Kelvin). Thus, the subversive underlying change did not become obvious. Koperski aptly sums up:

> By successfully incorporating the familiar language of laws into naturalism, it became possible to eliminate God as a useless metaphysical appendage. (ibd.)

With God as a "metaphysical appendage" removed, one gets laws without a lawmaker, which is exactly the mainstream view in science and philosophy today. But the foregoing brief historical investigation, in particular the last section, reveals yet another aspect of why talk of interaction (divine or mental) often meets with such vitriol: the naturalization of the laws of nature was an important weapon for establishing secularism and crushing the Christian faith in the culture at large, and without the rhetoric venom the severing of laws of nature from their theistic roots would hardly have been possible.

To draw the most important lesson from this chapter for our present purposes: the idea to conceive of the regularities of nature as *laws* can be traced back to Christian voluntarist theology but has 150 years ago been usurped by the anti-theist ideology of scientism. This may explain the violence of the opposition against interactionism: without a divine lawmaker and ultimate guarantor of the order of nature all we are left with is the laws – laws suffering no exception on pain of ceasing to be laws, and on pain of endangering science, or so things seem to be felt by many a non-interactionist.

The previous historical analysis also opens new strategical alleys for defenders of interactionism. For one, the presumed tension between regularities in nature and free agency could be solved by re-admitting God into the picture and thereby gaining a sovereign over the laws (as opposed to the laws being themselves the highest instance), thus mitigating the visceral unpleasantness of their possible breaking (which, as I have argued, is no entailment of interactive dualism). Theists can even go so far as to turn the tables on naturalists and ask whether non-theistic accounts of the laws of nature can do the work they're supposed to do with respect to the stability and regularity of nature – indeed we will see in chapter 13.2 that there are some issues which might be solvable by reintroducing theism into the study of the laws of nature.

12. Of laws and law-breaking

12.1 An easy way out?

In dealing with the question whether mental interaction breaks laws of nature it is first of all helpful to say in more detail what I have in mind when using the term 'laws of nature', because it gets attached to a bewildering variety of concepts from the Newtonian $F = ma$ to Mendel's laws of genetics. What we are naturally interested in here are laws pertaining to the microscopic or perhaps the submicroscopic realm, for it is at the level of molecules, atoms or subatomic structures that non-physical minds most plausible interact with bodies (see chapter 9). In other words, we are interested in the laws of physics, not the laws of chemistry or biology. Also, there seems to be a significant difference between formulae that just put into relation two or more quantities (like Einstein's matter-energy equivalence $E = mc^2$) and those which specify the spatiotemporal evolution of a system. I shall mostly call the former just 'laws' and the latter 'dynamical equations' (or 'differential equations'[104], or laws of motion).

What I also wish to do is explain some terms I am going to use frequently. It is best to use a concrete example. Consider the following four statements:

(1) Paris lies on the Seine.
(2) All countries on Earth have more than 100 inhabitants.
(3) All the pieces of furniture in my office are white.
(4) Any two massive bodies attract each other with a force of $F = G\frac{m_1 m_2}{r^2}$.

All those statements express facts. However, only (2) to (4) express *regularities*, i.e. statements of the form (x)Fx ⊃ Gx: for all x, if x is an F, then x is a G; for example, for all countries, if x is a country on Earth, then x has more than 100 inhabitants. However, among those statements, only (4) expresses a law of nature. A first observation is that one can distinguish between law *statements* and the *truthmakers* of those statements. (4) is first of all just a statement, but a true statement at that; and in this case there must be something in the world that makes the statement true. Also, what we can glean from above list is that there are regularities that count as laws and such that don't. Further, there must be something that distinguishes laws from mere regularities. What that something (the truthmakers of the laws) is differs greatly between the different theories of the laws of nature. I shall also sometimes speak of 'sub-nomic facts' or 'instances' of the laws. These are the facts underlying a law of nature, for example the fact that massive bodies attract each other according to the stated equation. That this constitutes a law of nature is a 'nomic fact'. In other words, it is a sub-nomic fact that any two massive bodies attract each other with a force of $F = G\frac{m_1 m_2}{r^2}$, but it is a nomic fact that it is a law of nature that any two massive bodies attract each other with a force of $F = G\frac{m_1 m_2}{r^2}$.

Now, we have so far not encountered a persuasive argument why interactive dualism should be ruled out. So, it is certainly possible, however perhaps at the cost of breaking laws of nature. How should the interactive dualist respond?

---

[104] Because the typical mathematical form of a dynamical equation is differential.

We should first of all not forget to import one crucial insight from the discussion in chapters 4-7, namely that conservation laws hold (bi)conditionally, as opposed to categorically: *if* (and only if) the relevant Noether symmetry is preserved *then* energy (or momentum) will be conserved. This might serve as a template for laws of nature in general: laws of nature hold if and only if the uniformity of nature (by the proxy of symmetries) is preserved, which is tantamount to there being no interaction by free, non-physical agents. We can take this to be a *ceteris paribus* reply: all other things being equal – in this case, no non-physical interaction – the laws of nature hold, where "hold" should be understood as "describe the temporal evolution of the physical universe". Such an approach is taken by philosopher of religion William Alston with respect to divine interaction. He writes that the traditional way of thinking about divine action

> does involve thinking of God as bringing things about other than they would have been had only natural factors been operative. But whether that implies a "violation" of natural laws depends on how we think of the latter. To suppose that it does is to presuppose that natural laws specify *unqualifiedly* sufficient conditions. . . . But we are never justified in accepting laws like this. The most we are ever justified in accepting is a law that specifies what will be the outcome of certain conditions *in the absence of any relevant factors other than those specified in the law*. (Alston 1994, 50)

Although he does not say so explicitly, it stands to reason that he has a *ceteris paribus* approach in mind: laws specify outcomes all else being equal, in this case in the absence of non-physical interaction. One could condense this into a conditional construal of the laws of nature:

(LC) The laws of nature describe the temporal evolution of the physical world in case no interaction occurs.

The advantage of such an approach is of course that one does not get one's hands 'dirty with metaphysics', so to speak: whatever the laws are, they hold as long as no interaction occurs. LC at once limits the purview of the laws of nature. They couldn't be *broken* in case of an interaction because they have nothing to say about such cases. This looks neat – too neat maybe. Doesn't it make non-violationism *trivially* true, since breakings of the laws are ruled out by definition? The triviality charge can be averted if we can come up with a construal that rules out neither interaction nor the breaking of the laws while staying true to the 'spirit' of lawhood. Here's one:

(LE) The laws of nature describe the temporal evolution of everything that exists.

On LE, even non-physical things are captured by the laws of nature. As a corollary, nature is extended to comprise also non-physical entities. This is one version of 'liberal naturalism' (cf. Williamson 2013b; 2013a) and thus a position in fact defended by respectable philosophers. On LE, laws would be broken in case the overall evolution deviates from the expected one, just as we could say with certainty that the laws are broken if nature behaves erratically without there being interaction (except on Humeanism). The trick about immaterial persons, however, is that they are *not* under the sway of the laws of nature. Thus, on an account of personhood and freedom defended here, the universal quantification contained in LE must be rejected and something like LC must be embraced – with the assurance of non-triviality this time. LC also has the further advantage of supporting the counterfactual "had no person interacted, the laws of nature would have made it the case that p". That gives us exactly what both freedom and regularity require.

However, the drawback of this approach is that by *modus tollens* one gets the result that if an interaction occurs, then the laws don't hold. That sounds unpalatable. So, can we come up with a response that preserves the advantages of LC but avoids the conclusion that the laws of nature are sometimes 'abrogated'? Perhaps the answer is to make the widespread (though not universally recognized[105]) distinction between laws, dynamical equations and initial conditions (let's call this the 'initial conditions approach', in short ICA). According to this distinction, the *laws* are statements about the structure of physical reality, but do not specify the temporal evolution of physical systems; that is done by the *dynamical equations* which derive from the laws. The *initial conditions* are the quantities which must be fed into the dynamical equations to yield an outcome. Take Newtonian mechanics: in order to get definite results from it, one needs to derive dynamical equations, for example via the Euler procedure. However, as Jeffrey Koperski points out, "[n]othing in the Euler procedure prevents the introduction of new influences on a system" (2020, 134). The idea is that mental interaction just alters the initial conditions, but not the laws (nor presumably the dynamical equations). I try to capture the relationship between initial conditions, laws, dynamical equations and interaction in figure 12.1.

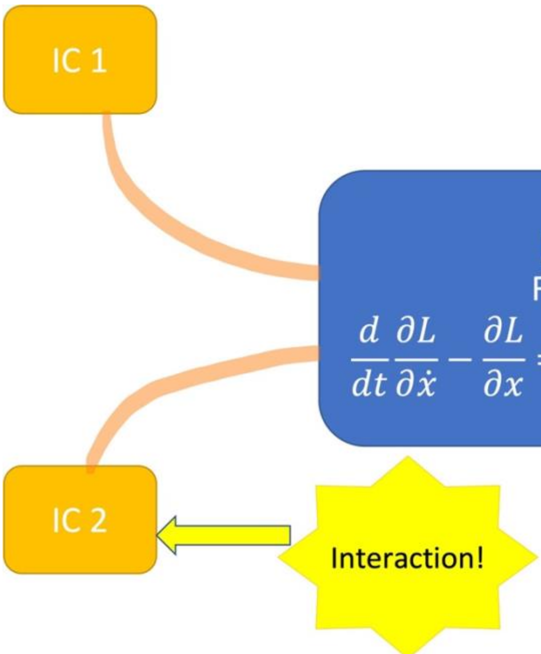

Figure 12.1. The law is the well-known Newtonian acceleration law. Below the corresponding dynamical Euler-Lagrange equation. IC 1/2 are sets of initial conditions; S 1/2 are (outcome) states.

For example, one can construct a dynamical equation (including the specific initial conditions) for an impact of billiard balls on a table, and of course the tilting of the table by one person will nullify that equation; however, what has changed is not the relevant laws (nor even the dynamical equation *sans* initial conditions), but the initial conditions. Analogously, couldn't it be that interactions never touch the laws but just change the initial conditions, and the laws 'take it from there'? I think this is what C.S. Lewis had in mind when he called nature "an accomplished hostess":

> The divine art of miracle is not an art of suspending the pattern to which events conform but of feeding new events into that pattern. It does not violate the law's proviso, 'If A, then B': it says, 'But this time

[105] For a dissenting voice see (Adlam 2022, 27-28).

> instead of A, A2,' and Nature, speaking through all her laws, replies 'Then B2' and naturalises the immigrant, as she well knows how. She is an accomplished hostess. (1960, ch. 8)

The ICA preserves the conditionality aspect without implying that they sometimes fail to hold, because the conditionality is applied to the dynamical equations only. This sounds really good – but we shouldn't jubilate just yet. The problem is this: the outcome states of one physical event (as sketched in fig. 12.1) are the initial conditions for the next transition, or conversely, the initial conditions for one physical event have been the outcome states of a previous one. Thus, if mental interaction changes the initial conditions of one physical event, it has thereby altered the outcome of a previous event; but that event started with initial conditions that implied another outcome than in fact obtained. The promise was that mental interaction should only change initial conditions; now it seems it inevitably also changes outcomes. However, what is changed is the outcome of a dynamical equation, which we beforehand distinguished from a law. The dynamical equation seems indeed to be falsified: it predicted outcome S1 from initial conditions IC1, but due to interaction S2 emerges. The underlying law, though, might still remain untouched, at least so long as the distinction between laws and dynamical equations is a robust one. On this picture, one 'pawns' dynamical equations for the safety of the laws.

However, what if the laws, whatever they are metaphysically, are directly broken by interaction? The result would be empirically the same – and the ICA distinction insufficient to detect the breaking. In order to see whether interaction necessarily does this, or if there is perhaps a way around it, we will now look at the big metaphysical options regarding laws of nature, and then ask whether laws are inevitably broken on each of them.

12.2 How can a law of nature be broken?

Let's start with a very broad categorization of the extant metaphysical theories. The first big divide (as is generally acknowledged) runs between *necessitarian* and *Humean* (non-necessitarian) views. On necessitarian views, the laws of nature possess *modal force*: in some way or other, they force their instances (the concrete physical events explainable by the laws of nature) into being – it is *nomologically necessary* that a certain event or at least one of a limited range of events happen. Conversely, it is *nomologically impossible* that any event or world-state follow any event or world-state. By contrast, Humean views deny laws such modal force[106]. According to Humeanism, physical events just occur, without anything forcing or restraining them, and therefore without any deeper explanation – they just happen. The laws of nature are, on this view, just (optimized) descriptions of the event pattern one has so far been able to observe.

Necessitarian views can be further subdivided into *generative* or *productive* and *constraining* views. Generative accounts have it that the laws of nature, or those entities on whose behavior the laws supervene, actively (productively) bring about events, whereas on a constraining view the laws just constrain the range of possible future histories a physical system can have. Among the generative views we can distinguish between those that see the productive force as *intrinsic* to matter (among which I shall discuss dispositionalism and panpsychism), and those that locate it in something *extrinsic* to matter (e.g. the DTA

---

[106] Strictly speaking, Humeans will probably not deny *all* sorts of modality in nature. Besides necessity, *possibility* is another type of modality, which Humeans should even more generously grant than non-Humeans, for if events in nature are not *necessitated* by anything, then surely many past and future world histories would have been/will be *possible*.

theory and Maudlin's nomological fundamentalism). In the camp of constraining views, two shall be considered: Marc Lange's counterfactualism and Eddy Keming Chen's and Sheldon Goldstein's recent minimal primitivism (MinP). For a better overview see the taxonomy in fig. 12.2:

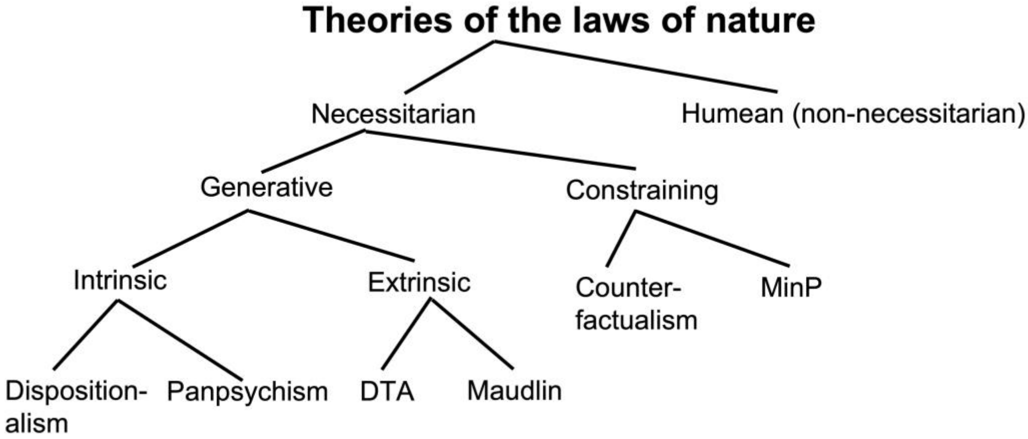

Figure 12.2: A taxonomy of the most important theories of the laws of nature

So, what does it mean for interaction to 'break' or 'violate' the laws of nature? First, a word about the metaphor of 'breaking' (or 'violation'). Despite its wide use, it is may be semantically misleading here; after all, we shouldn't conceive of laws of nature as moral or juridical laws for which we have a clear idea what it means to *break* them, namely, act against what they say ought (not) to be done. Laws of nature, however, are not expressions of what *ought* to happen, but what *does* or *will* happen. What term would be more adequate than 'breaking' is not immediately clear to me, so let's leave this question open for now and continue to use the metaphor, keeping in mind its flaws.

An important preliminary observation is that to break a law of nature it suffices to do so *locally* (in both the spatial and temporal sense); the law needn't be broken *globally*. That is in keeping with the conceptualization of conservation principles as applying to local physical systems (see chapter 5). Now, how could a law of nature be broken, depending on the type of metaphysical theory assumed? We shall look at necessitarian (productive and constraining) as well as Humean conceptions.

12.2.1 Law-breaking on necessitarian accounts

A natural starting point for coming up with a setting in which some law of nature is broken are necessitarian accounts of the laws of nature. According to above taxonomy, there are generative and constraining necessitarian theories, and among the generative accounts 'intrinsic' and 'extrinsic' ones. In investigating the generative theories, I will disregard the intrinsic/extrinsic distinction. It becomes important again for the assessment of the performance of the respective accounts in their 'core business' as theories of laws of nature; here, all that matters is that there is a modal force directed from one state to another. Whether that force is intrinsic to matter or imposed on it extrinsically is irrelevant here.

Let's start with *generative* necessitarian accounts. Consider fig. 12.3:

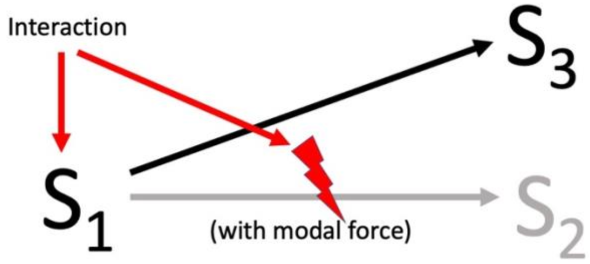

Figure 12.3: Interaction with obliteration of modal force

The states S_i should (as in the following figures) be understood as concrete physical states (sub-nomic facts), of which S_1 subsumes the initial conditions for the process under consideration. The black and grey arrows pointing away from S_1 represent, to be precise, not the laws, but the dynamical equations derived from the laws which yield outcomes only if 'fed' with quantities from which to depart, something that becomes especially clear in their mathematical formulation. At the same time, the arrows are supposed to represent the necessity involved, as distinct from the events to which it leads.

In this scenario, a mental interaction not only prevents the occurrence of S_2, but also the occurrence of the nomological necessity leading to state S_2. Instead, it exerts a different necessity that brings about S_3. Thus, neither the nomological necessity nor the expected state are actualized. If now the laws are *constituted by* the pertinent nomological necessities (as on Maudlin's nomological fundamentalism, and in some sense Lange's counterfactualism, the DTA theory and MinP) or *supervene* on them (as in dispositionalism and panpsychism), then mental interaction indeed 'breaks' the laws of nature.

This is the only clear-cut case where a law of nature is indisputably broken. To see why, let's alter the scenario just insofar as the modal force leading to S_3 is not obliterated, but remains in existence. The picture then looks something like this:

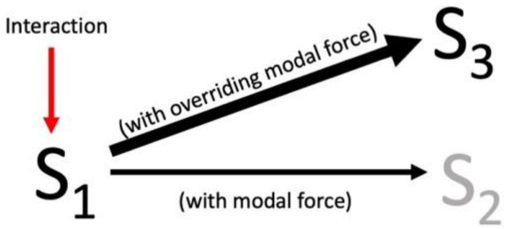

Figure 12.4: Interaction without obliteration of modal force

As can be seen from fig. 12.4, the only thing whose existence is prevented is S_2. This is equivalent to the structural (here, generative) nomological necessity staying in existence but being 'trumped'. To give an analogy, the situation is like that of an object held in mid-air by a thread. No one would say that gravity has been obliterated, rather that it is 'trumped' by the retention force. One view that affirms this metaphysical picture is Daniel von Wachter's "directedness theory" (2009, 2015). He thinks that what connects one world-state to the next is a "directedness" which can be 'trumped' by non-physical directednesses. His account thus has many parallels to dispositionalism, with the most important difference being that he rejects dispositionalism's substance ontology (Von Wachter 2001). At any rate, on this scenario, no law of nature is broken.

Above schema is directly applicable to all *generative* modal accounts (dispositionalism, DTA, nomic fundamentalism, Dolbeault's panpsychism); but what about theories where the modal force is *constraining*? In this case the situation looks as follows:

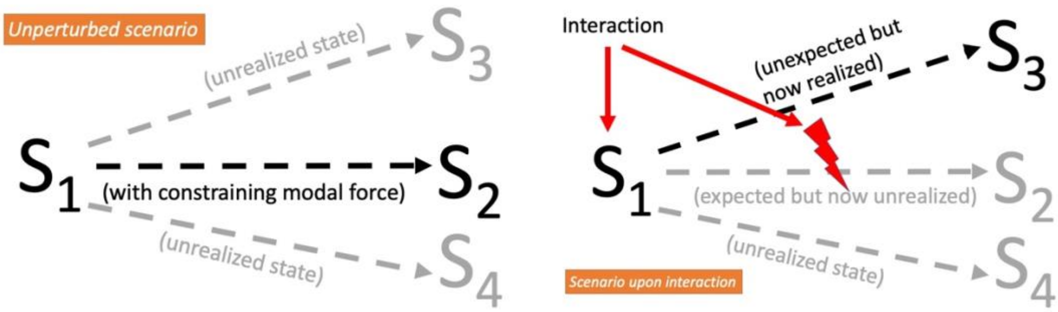

Figure 12.5: Interaction on a constraining necessitarian account of laws

The left-hand side of fig. 12.5 shows how things unfold in case no interaction occurs, and the right-hand side in case an interaction does occur. If 'nature is left to herself', the laws of nature constrain the many (if not infinite) possible futures of S_1 to one (or, in the case of probabilistic laws, to a very small number). Laws *as such* constrain the possible histories of physical systems from infinity (or a very large number) to a finite, smaller number of histories, but laws applied to initial conditions restrain the possibilities even more (since the initial conditions already embody a constraint). Is above scenario one in which a law of nature has been broken? That again depends on whether a modal force has been obliterated or just overridden. A difference to generative accounts is that the modal force does not produce anything (that must be done by some other force) but instead chooses which outcome obtains. So in order for the law of nature not to be broken that 'choosing power' must not be obliterated. We could think of it in terms of a river branching into three streams, of which two are blocked by dams (corresponding to S_3 and S_4 being blocked). The interaction would consist in blocking the hitherto free stream S_2 and removing the dam of one of the hitherto blocked streams (here S_3). Now the question is whether the continued existence of the now waterless riverbed corresponds analogously to the existence of a constraining force toward S_2. Perhaps indeed, if the constraints are something akin to 'draws' in the topology of nature, through which flows the otherwise unbridled possibility of nature.

Another analogy might help fortify this conceptualization. Suppose I stand in front of the shelf with pasta in the supermarket. I choose to get pasta of brand X, but instead my hand is forced – say, by someone stronger than me – to grab brand Y instead. Is my choice thereby obliterated? Clearly not.

Although this is far from a decisive argument – one could still hold that a choice consists solely in the realization of the chosen event – it is sufficient to establish as a viable position the idea that interaction does not break constraining necessitarian laws either.

All in all, we can say the following regarding the question whether mental interaction breaks laws of nature on a necessitarian understanding: If a law of nature consists in a certain state following a prior state, then if interaction makes it the case that another state is actualized, interaction has made the law fail. Period. However, as indicated above, in case laws are understood in terms of a certain modal 'directedness' towards a later state, and if interaction does not obliterate that directedness but 'trumps' it, then no law is broken.

12.2.2 Law-breaking on Humeanism

We have considered one *modal* scenario where the laws of nature are putatively *not* broken. It is far clearer that no law is broken in case one adopts a *non-modal* metaphysics of nature, one in which events don't occur by modal force. This corresponds, broadly speaking, to a Humean picture. Figure 12.6 seeks to capture this scenario.

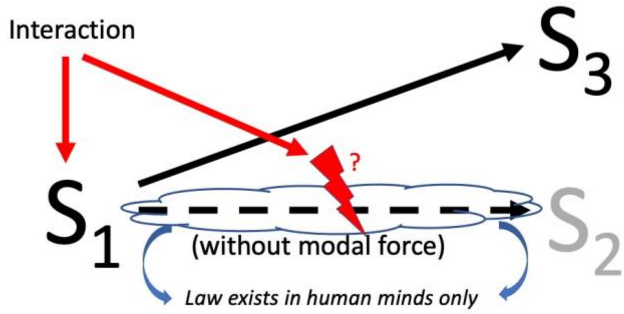

Figure 12.6: Interaction on Humeanism

S_2 simply follows S_1, without there being any necessity – whether productive or constraining – involved in its occurrence (this is expressed by the dashed arrow). The lawfulness we ascribe to this sequence of events is, according to the gold standard of Humeanism, the Best System Analysis (BSA), a matter of its belonging to the best system of analysis of the motion of matter, an analysis optimizing the opposed requirements of simplicity and informativity. Humeans, especially in the wake of David Lewis (D. Lewis 1973) claim that the integration of this optimization is sufficiently objective, because of the universal recognition of the standards of simplicity and informativity; however, not all agree, mostly because of disagreement regarding how to define simplicity and informativity. If there was a link between the way nature behaves and our intuitions regarding simplicity and informativity, then the Humean project could very well count as objective, because then the theoretical standards would directly correlate with the world out there. However, as long as such a link has not been established – and it neither has been nor does it seem something congenial to the spirit of the Humean project in general – Humeans can only, to paraphrase David Lewis, hope that nature will be kind, and Humean laws can, in virtue of the generally empiricist Humean epistemology, not really be called 'objective'. Thus, strictly speaking, Humean laws are merely subjective; they are in our minds, not 'out there' in the things, although of course predicated on truthmakers in the world. It seems therefore that interaction does not obliterate or break anything, because there is nothing to be obliterated or broken. All that happens is that the mental construction called 'laws of nature' would have to be altered, because the event patterns on which it supervenes have been (slightly) altered.

Not all Humeans, however, share such insouciance about this approach. Michael Esfeld, for example, is loath to take this route, although he recognizes that other Humeans embrace it (Esfeld 2020, 98-99). His own solution is that minds contribute to *fixing* (not to *altering*) the initial conditions at the beginning of the universe by their interaction with the physical world. One reason why he wants to leave the laws unscathed is that he considers them, in contrast to initial conditions, as something limiting our degrees of freedom:

> Only if there are such stable patterns or regularities are there laws at all. Keeping these patterns or regularities as they are expressed in the laws of nature that we can formulate fixed, it is physically or nomologically impossible for a person to travel faster than light, or to jump as high as her house. This is all that is needed for the laws to define the range within which we can act freely. (…) [R]egarding the laws of nature as being dependent on our actions comes into focus as a move in order to achieve

> the compatibility of our free will with the laws of nature only as a last means, precisely because the laws of nature delimit the range of what we can do. (2020, 100)

It seems that Esfeld proceeds from the fact that it is impossible for us to do certain things and concludes thereof that the laws must be stable, because the subvenient patterns are stable – and thus that we should rather not touch the laws. Of course, he admits that Humean laws, for lack of modal force, provide no guarantee for the observed regularities to continue into the future:

> That notwithstanding, it is metaphysically possible that, tomorrow, a person travels faster than light or jumps as high as her house. There is nothing in the present or past configuration of matter that makes it impossible for such things to happen. It is only that if we presume that the salient regularities that were discovered in the past motion of matter will prevail in the future that we get to such restrictions. But there is nothing in the universe that makes it necessary that these regularities will prevail in the future. Deliberating about her actions on the basis of the observed regularities of motion as expressed in the laws of nature that figure in our scientific theories is in any case all that a person can do, independently of whether or not there is something in nature that constrains the future evolution of the configuration of matter. (2020, 99-100)

A question arises at once: if the laws can change in the future, why not in the present, through mental interactions? There is no way around admitting that mental interactions change the matter-in-motion mosaic on which Humean laws supervene, and thus could indirectly change the laws. Esfeld recognizes that, but his own solution is to leave the laws intact and instead locate the influence of human actions in the initial conditions of the universe:

> Consequently, on Super-Humeanism, the bodily motions of human beings that are expressions of their free will contribute to determining the values of the dynamical parameters at the *initial* state of the universe, that is, the state that enters as initial condition into the laws and without which the laws would not be in the position to describe the evolution of the universe. Hence, if human beings chose to do otherwise, in the first place, slightly different initial values for the dynamical parameters at the initial state of the universe would have to be figured out in order to achieve a system that maximizes both simplicity and informational content about the change that actually occurs in the universe. (Esfeld 2022, 264; italics original)

Importantly, this proposal does not constitute retrocausation; it is about how the motion patterns of matter, to which mental causation contributes, are structured. For Esfeld, we should keep the successful BSA that construes laws of nature as a system of rules optimized for both simplicity and informativity (see section 13.2.1), and rather adjust the initial conditions that go into the laws to yield the actual history of the physical world, since the distinction between laws and initial conditions has to be made anyway. Of course, this distinction, as well as the best system analysis of the laws of nature, is something ultimately mental, wherefore the standards for breaking laws of nature are ultimately located in human minds.

All in all, it seems that on Humeanism, laws of nature cannot be broken in any interesting sense. If Esfeld's approach works, then no law is ever broken or changed by mental interaction; on other Humean accounts, like Jenann Ismael's, the laws are changed but not broken; and even if one insisted on an analysis of the laws which has them broken by mental interactions, then it would still be true that 'just' some mental rules had been broken, not something out there in the world.

The conclusion of this chapter is twofold: first, that the distinction between laws, dynamical equations and initial conditions is a necessary first step towards an interaction-friendly conception of the laws of nature;

and that, second, law-breaking is not an inevitable consequence even on necessitarian accounts so long as that which constitutes the modal force remains intact during interaction.

13. An investigation of extant metaphysical theories of the laws of nature

13.1 Desiderata

We've seen in the previous chapter that under certain assumptions, all nomological theories under consideration can, with certain assumptions, accommodate mental interaction without requiring the violation of laws of nature. Still, there are differences as to how effortlessly they do that. For example, on Humeanism, no further assumptions are necessary, whereas on necessitarian accounts on needs to embrace the existence of directednesses. And the openness to interaction is only one criterion for assessing theories of the laws of nature; their 'core business' is actually, well, accounting for the laws of nature. What is meant by that is a matter of some debate; I try to capture a reasonably wide range of desiderata by using the ten-point catalog of criteria offered by Jaag and Schrenk (2020)[107]. I will not canvass all the desiderata for each theory under examination, but since I will refer to each of them at some point or other, it is indispensable to explain them all:

I. **Objectivity:** Laws of nature are discovered and exist whether we explore them or not. Furthermore, we can be radically mistaken about whether a fact is a law or not, as a look at the history of science suggests. This supports the view that laws, as well as their law-like character, are mind-independent.
Two observations about the criterion of objectivity are in order: First, to be considered objective, a fact need not be metaphysically fundamental. The objectivity criterion is perfectly compatible with laws of nature being reducible to more fundamental objective (i.e. non-mental facts. Second, objectivity refers not only to the laws themselves, but primarily to their law *status*. Antirealist positions about laws of nature usually deny the latter, that is, the objectivity of law facts: The fact, for example, that all metals conduct electricity may exist independently of cognitive subjects. But whether this is a law of nature or not, according to anti-realism, usually depends in a relevant way on cognitive subjects. Laws of nature must exist independently of human minds.

II. **Inference:** Roughly, "If it is the case *that p is a law of nature*, then it is also the case *that p*." In other words, this criterion ensures the lawful connection between laws of nature and their sub-nomic facts or instances. For example, *if it is a law of nature that masses subjected to a non-zero net force accelerate*, it is also the case *that masses subjected to a net non-zero force accelerate*. A good theory of the laws of nature should establish the connection between the antecedent and the consequent of such conditionals.

III. **Invariance/universality:** This is usually expressed in such a way that laws of nature are generalizations of the form "All Fs are Gs" (more formally, $(x)(Fx \supset Gx)$). There are four ways to understand this generalizability: (i) at least fundamental laws of nature hold *under all circumstances* (i.e., admit of no *ceteris paribus* clauses); (ii) laws of nature hold *at all spatiotemporal locations*; (iii) laws of nature hold for *all systems*; and (iv) laws of nature hold for all determined values of corresponding quantities (e.g. Newton's laws of motion hold for *all* masses).

[107] The original German designations are Objektivität, Faktizität/Inferenz, Invarianz/Universalität, Kontingenz, Nomologische Notwendigkeit und Stützung kontrafaktischer Konditionale, Erklärungskraft, Steuerung und Unabhängigkeit, Innere Verknüpfung, Induktive Bestätigung und Stützung induktiver Schlüsse, Bedeutung für die Wissenschaften.

IV. **Contingency:** This is the idea that the laws of nature could be different than they actually are; more specifically, that their quantification is different. For example, it is at least conceivable that the gravitational force between two masses, contrary to Newton's law of gravity, is not inversely proportional to r^2, but to r^3.

V. **Counterfactual support:** Laws of nature should support counterfactuals of the form: "If it is a law of nature that metals conduct electricity, then if my spoon were made of metal instead of wood, it would conduct electricity." ("If it is a law of nature that all F are p, then if G were F, then Gp"). This distinguishes laws of nature from mere accidental regularities. For example, it is not true that "All the spoons in my drawer are made of metal" is a law of nature, because the counterfactual "If I put a fork in my drawer, then it would turn into a spoon" is not true.

VI. **Explanatory power:** Laws of nature should explain their instances. For example, the law that like charges repel each other explains why two electrons repel each other. Mere accidental regularities lack this explanatory power: The fact that all the objects in my pocket are coins and the fact that a certain object is in my pocket does not explain why it is a coin.

VII. **Governance:** Laws of nature should not just describe but 'govern' the events in nature. That governance can be conceived in at least two ways: (i) as *productive*, i.e. the laws of nature *bring about* their sub-nomic facts; (ii) as *constraining*, i.e. the laws of nature do not bring about the sub-nomic facts but rather limit the possibilities regarding which facts can occur.

VIII. **Inner connection:** Laws of nature should be such that there is an inner connection – arguably a causal one – between their concrete instances. For example, there should be an inner connection between two planets that attract each other, rather than the mere description that they move towards one another.

IX. **Inductive support:** An adequate theory of the laws of nature should do justice to the inductive confirmation of the laws of nature by their instances. When a given piece of metal conducts electricity, that confirms the law that "All metals conduct electricity". It also increases our confidence in the law-statement actually being a law of nature.

X. **Significance for the sciences:** An adequate theory of the laws of nature should do justice to their central role in the natural sciences.

Since this is no monograph on the metaphysics of the laws of nature, it is not reasonable to meticulously canvass every theory for each of those desiderata. My strategy is instead the following: (1) drop dispensable desiderata; (2) 'amalgamate' the above desiderata to a few larger 'composite' criteria; and (3) add a 'freedom' criterion. This is how it plays out concretely:

(1) Desideratum to be waived: (IV) "Contingency"

(2) Desiderata to be amalgamated: (I) through (III) and (V) through (IX) => "Metaphysical robustness"

(3) Desideratum to be added: "Receptivity for non-physical interaction"

The desiderata VII and VIII (contained in "Metaphysical robustness") admittedly constitute a certain bias against Humean theories, but they clearly are widely shared and reasonable elements for a good theory of the laws of nature. It may therefore seem a bit unfair to waive the contingency criterion, since here Humeanism could play to its strengths. However, there is one reason why, in the context of the present project, contingency does not represent a selective criterion: in a theistic framework *any* of the considered construals of the laws of nature is contingent, because God as lawmaker could always have made the laws different, even on dispositionalism, where He could have created different dispositions.

So all in all, we get three desiderata:

I. Metaphysical robustness (in short: *robustness*): contains all criteria pertinent to the explanation of nature's regularities
II. Standalone desideratum: *scientific significance*
III. Receptivity for non-physical interaction (in short: *receptivity*)

"Metaphysical robustness", I think, captures well the common theme of desiderata I-III and V-IX: laws of nature (the law facts and their status as laws) should exist in a relevant sense independently of human subjects and thus have a reality outside of human minds (I); they should ensure the inference to their instances (II), explain (VI) and even govern them (VII), assure the connections between them (VIII) and support induction (IX). They should do so counterfactually (V) and further hold universally (III). Most of these criteria are met by the ascription of *modal force* to the laws of nature (all except I and III). By modal force I understand *natural* or *physical*[108] *necessity*[109]. Metaphysical robustness stemming from modal force thus makes laws of nature some sort of 'power' out in the world that produces or at least constrains the events in nature. This compound desideratum of course puts at a crass disadvantage Humean theories which miss all these standards partly or wholly; however, there is still the *freedom* criterion (3) where Humeanism can play to its strengths.
Finally, a word about "Invariance/universality" is in order. It is crucial that it not be understood in a way that rules out interaction *a priori*. One might think *prima facie* that sense (i) is guilty of this; however, this would only be true if the *only* way to get around law-breaking interaction is via a *ceteris paribus* approach, which, as I have argued, is not the case.

13.2 The accounts under scrutiny

As we've already observed, the great divide among extant theories of the laws of nature runs between non-necessitarian (Humean) accounts and necessitarian accounts. Among generative necessitarian accounts, I will discuss the Dretske-Tooley-Armstrong (DTA) theory, dispositionalism, Tim Maudlin's nomic fundamentalism and Jean Dolbeault's panpsychist proposal. In the 'constraining camp', I shall investigate Marc Lange's counterfactual theory and the very recent 'Minimal Primitivist' (MinP) contribution by Eddy Keming Chen and Sheldon Goldstein.

13.2.1 Humean theories

Humean theories of the laws of nature get their moniker from David Hume, who famously expressed skepticism about the existence of causal connections. Hume held that all we can say for sure is that events

[108] 'Physical' strictly speaking refers to the laws of physics only. For present purposes, however, no sharp distinction needs to be made between the two terms.
[109] I will use the terms 'necessitarian' and 'modal' interchangeably.

follow each other temporally. A modal force between cause and effect – something in the cause that forces the effect into being – is imperceptible to our senses and so should not be assumed. (The origin of Humean law theories in skepticism about causation further underscores the intricate connection between laws of nature and causation).

Humean theories basically deny any modal force to the goings-on in nature. In particular, nature does not behave the way she does *because of the laws*, but the laws are what they are *because nature behaves the way she does*. Humeans often use the concept of supervenience to describe this relationship: the laws of nature supervene on the mosaic of 'perfectly natural, fundamental properties'. Thus David Lewis:

> It is the doctrine that all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. [...] For short: we have an arrangement of [fundamental] qualities. And that is all there is. There is no difference without a difference in the arrangement of [fundamental] qualities. All else supervenes on that. (Lewis 1986, ix-x)

Supervenience is not a unique feature of Humean theories, though: it also figures in dispositionalist and arguably in panpsychist accounts.

The gold standard for contemporary Humean theories is the so-called Best System Analysis (Lewis 1986; Hall 2009). It basically says that laws of nature are distinguished from mere accidental generalizations by embodying an optimum between simplicity and informativeness (see Hall 2009).

A clear advantage of Humeanism is that it poses no obstacle to non-physical interaction, as I showed in section 12.2.2. Jeffrey Koperski expresses this as follows (with respect to divine interaction):

> The regularity theorist's laws are merely observed correlations between events. No further metaphysical commitment is required. If correct, nonviolationist worries about God possibly violating the laws of nature will simply be misplaced. (Koperski 2020, 90)

> "Breaking a law of nature" under MRL [the Mill-Ramsey-Lewis theory, a.k.a. Best System Analysis; my addition] just means that God does something contrary to what we would have expected based on what we know (i.e., based on the best deductive system of scientific knowledge). But that doesn't seem to be the sort of thing that nonviolationists are worried about. Terms such as "violation" and "breaking" indicate something more significant than God acting in a way contrary to our fallible expectations. (ibd., 91)

Replacing "God" by "minds" is unproblematically possible *salva veritate* and gives us an accurate picture of the metaphysical innocuousness of mental interaction in a Humean physical world. Of course, letting minds interact with a Humean world has the counterintuitive consequence that we can change the laws of nature by our free actions. Whoever finds this unpalatable has the option of having minds fix the initial conditions of the universe (see the discussion of Michael Esfeld's proposal in 2.2.2). But as regards the metaphysics of the world outside human minds, Humean laws, per se, are fully open to interaction. Thus, *receptivity* is as fully met as one could wish.

The greatest problems for Humeanism lie on the *robustness* front. (The following are just a selection of the most important criticisms about Humeanism.) For one, even if BSA Humeanism can escape all the problems related to setting up true generalizations that satisfactorily (and non-pragmatically, non-arbitrarily) identify laws of nature, then laws of nature, including their status as laws, are still essentially mind-dependent, despite their supervenience on factual truthmakers in the world. Apart from the criteria of simplicity and informativeness, whose exact content is a matter of dispute, there is nothing in the distribution patterns of natural properties (or motion patterns of point-particles on Super-Humeanism) that commits us to construing the laws of nature in a certain way. Thus, the objectivity desideratum incorporated in *robustness* is at best only partly fulfilled. The problem is no trifle. First, the very fact of their mind-dependence renders the laws ontologically variable. Think about it: if all there is to the laws is an agreement among scientists, then when that agreement changes, the laws change. Note that it is not the degree to which the laws have been discovered that changes, but *the laws themselves*, for on Humeanism there is nothing mind-independent, neither reified laws nor other metaphysical entities on which the laws supervene, that could be discovered. According to Humeanism, we would have to say that in a strict sense the laws of nature have changed, for a lot of what passed as a law of nature in Newton's time is no longer accepted as such now.

Second, Humeanism gives no explanation where an explanation is in order, namely for the stunning fact that such simple and elegant mental concepts as our laws actually fit the nature around us. David Lewis clearly saw both aspects of the problem, but ultimately expressed hope in the 'kindness of nature':

> The worst problem about the best-system analysis is that when we ask where the standards of simplicity and strength and balance come from, the answer may seem to be that they come from us. Now, some ratbag idealist might say that if we don't like the misfortunes that the laws of nature visit upon us, we can change the laws—in fact, we can make them always have been different—just by changing the way we think! (...) The real answer lies elsewhere: if nature is kind to us, the problem needn't arise. I suppose our standards of simplicity and strength and balance are only partly a matter of psychology. It's not because of how we happen to think that a linear function is simpler than a quartic or a step function; it's not because of how we happen to think that a shorter alternation of prenex quantifiers is simpler than a longer one; and so on. Maybe some of the exchange rates between aspects of simplicity, etc., are a psychological matter, but not just anything goes. If nature is kind, the best system will robustly be the best—so far ahead of its rivals that it will come out first under any standards of simplicity and strength and balance. We have no guarantee that nature is kind in this way, but no evidence that it isn't. It's a reasonable hope. Perhaps we presuppose it in our thinking about law. [...] I can admit that if nature were unkind [...] then lawhood might be a psychological matter. [...] But I'd blame the trouble on unkind nature, not on the analysis. (Lewis 1994, 479)

Also, the desire to explain the regularities in a *generative* way (desideratum of governance) is frustrated, by definition.

Now one might object that both these wishes are simply wrong-headed. As understandable as they are commonsensically, the true explanation for an impression we gain from interacting with nature may just defy common sense. Take the intuition that since a table is solid, then all its atomic constituent parts must also be solid in a 'billiard ball' sense. That intuition is wrong; therefore, our intuitions concerning (generative) explanation could also be wrong. In reply, note first that the two intuitions are situated at different levels of abstraction. One concerns a specific explanation as opposed to another; the other the question whether there is to be expected an explanation *at all*. Second, and this is not a coincidence, the true scientific explanation in the table case relies precisely on laws of nature as a deeper explanation. How can an aggregate of particles which for the most part consist of electron shells and empty space yield a solid table? Well, because the laws of nature specify that one big electron-cloud aggregate (like my hand) will be repelled by the aggregated electron cloud that is the table, and the very fact that I cannot pass my hand through the table is what we call solidity. So I submit that our request for the laws of nature to be truly explanatory is *not* misguided.

Yet even apart from this appeal to 'common sense epistemology', there is a hard philosophical-mathematical reason why we should insist on a modal explanation of the behavior of nature. It is linked to a problem about induction (which is why Humeanism fails on the criterion of inductive support as well). Induction is admittedly an inferential extrapolation across space (from this place to another) and time (from the now to the future), but failure to support induction results from failure to explains things in the *here and now*, for if one had a strong foundation to expect things to happen similarly in other places and at other times, this expectation would be grounded in what makes it the case that things behave the way they do *here and now*. However, Humeanism *ex hypothesi* rules out any explanation for the behavior of matter, but rather holds that any behavior of nature is possible. Hence, why expect nature to behave tomorrow as it has done so far? To be sure, the induction problem also applies to modal accounts of the laws of nature, at least in an epistemic sense. For even if the laws of nature possess modal force, we may still be mistaken about what the laws of nature truly are, and so have no guarantee that our inductions for the future hold. However, one could say that on modal accounts, if we knew the true laws of nature, then our inductions could not fail, whereas on Humeanism this is not the case: even if we knew the true laws of nature now (where 'true' depends on one's axiomatic standards), we would have no guarantee that they remain that way in the future.

There is a further objection against Humeanism based on the fact of the laws' simplicity/elegance. This objection does not invoke the future, but rather holds fixed any given moment and asks why the patterns of motion are the way they are – intelligible and representable by simple and elegant laws – even though Humeanism allows for any pattern that does not include a logical contradiction, no matter how chaotic. Thus Koons and Pickavance:

> According to the Neo-Humeist, any pattern or distribution of qualities in space and time is possible. The vast majority of these distributions are very chaotic, obeying no simple generalizations at all. In fact, the ratio of chaotic to orderly patterns in a universe as large and complex as ours is a number of absolutely staggering size. Even if we restrict ourselves to patterns that conform to the observed generalizations within the space-time bubble that represents our actual history so far, the number of possible extensions of that history in which the observed order breaks down vastly outnumbers those in which things continue in much the same way.
>
> There is a principle of probabilistic reasoning, the Principle of Indifference, which dictates that the probabilities of two hypotheses should be proportional to the number of possibilities in which each of the two hypotheses would be true. Applying the Principle of Indifference to the problem of induction, given Neo-Humeist assumptions about possibility, leads to a catastrophe of skepticism. We would have to include that the probability that any generalization we have observed so far would continue to hold just in the near future and in our local neighborhood would be so small as to be negligible. (Koons and Pickavance 2015, 58)

So, according to the Principle of Indifference, we should discard the Humean hypothesis, because the probability that nature will behave in the next moment the way it used to do up until now is vanishingly small compared to the joint probabilities that it will take on another behavior.
John Foster holds a similar position:

> If there were no reason for the conformity—nothing which obliged or encouraged it—we would expect some showing of other types of outcome; not because there would have to be something which positively promoted such outcomes, but simply because it would be very strange if, without anything

to encourage things to stick to the relevant pattern, the opportunity for deviance was never in practice taken. (Foster 2001, 146)

Foster further invokes coin-tossing as an illustration. If one always got heads it would be wildly irrational to assume mere chance has been governing the process. Or conversely, if we hold fast to a chance explanation, there is no reason to extrapolate the previous pattern to subsequent tosses. By the same token, since we consistently get 'heads' (stable, intelligible regularities), we should either give up our expectation that this pattern persists into the future (if we stick to the Humean view) or rightly insist on an explanation. Michael Esfeld objects to the viability of a probability analysis of possible Humean worlds. He thinks that it makes no sense to "admit all possibilities" because such an admission would make the weighting of probabilities impossible. In his view, a proper assignment of probabilities can only take place if frequencies of events in the actual world are taken as base; something, Esfeld points out, statistical mechanics as well as Humean probability theory do (personal e-mail conversation, 8 July 2022). There is something to be learned from Esfeld's criticism, while I disagree with the overall conclusion.

First it is important to realize that there are other ways to define probabilities. Esfeld's description matches the theory of relative frequency, which is only one of three basic probability theories (see for example (Hurley and Watson 2018, ch. 11), the other two being the classical theory and the subjective theory. The classical theory is *a priori* and relies on the assumption of equal probabilities of all events in a given set (e.g. cards from a deck), but apart from that it uses the same formula to compute probabilities as the relative theory ($P = \frac{f}{n}$ with f = number of 'favorable' outcomes and n = number of possible outcomes). Admittedly, the equality of all possible Humean worlds cannot be assumed offhand, and differentiated probability assignment to subclasses of worlds founders on our ignorance, for example of how stable such worlds would be. (Of course, if there were an explanation why an orderly nature such as ours is strongly favored over more chaotic versions, then Humeanism would be in trouble, because it is precisely such an explanation it refuses to give.) At any rate, if one introduces numbers into the probability game about Humean worlds, then Esfeld's objection seems to apply.

However, there is still the subjective theory, where probabilities are not assigned any numbers, thereby dodging one of Esfeld's criticisms. He would certainly not be satisfied, by that answer, though. The subjective approach is likely to appear to him too unscientific and open to arbitrariness. I do not think that subjective probabilities are worthless merely *qua* lacking numerical value, as Hurley and Watson's example with the putative Stradivari violin nicely shows; the main reason is that our judgment is guided by background knowledge which is impossible to cast into numbers. It might be objected that in the case of a cosmic-scale question like that of the laws of nature no background knowledge can be invoked. In a way, this is true: we cannot have a background behind that against which we measure and assess all things. It then, admittedly, seems hard to formulate a rigorous probability argument against Humeanism, despite the residual unease about its refusal to give any explanation why only one of very many logically possible worlds is realized, and remains stable over vast stretches of time. I therefore leave it at that.

Another problem related to explanation is that Humean 'explanations' (strictly speaking, unifications) are circular: the laws of nature supervene on (are 'explained by', in a way) a mosaic of physical events, but those events are in turn (in scientific practice, at least) explained by the laws of nature. A solution to this circularity has been offered by Barry Loewer. Jaag & Schrenk comment on it as follows:

> A proposed solution to this difficulty comes from Barry Loewer (…). He tries to break the circle by postulating two kinds of explanations: metaphysical and (natural) scientific explanation. The laws of nature supervene on all individual instances/facts, they are metaphysically explained/constituted by them. In the practice of science, however, the laws of nature are then used to explain scientifically why single events happen the way they do. Since these are two different kinds of explanations, no circle of explanation arises, according to Loewer. However, this solution is controversial and it is considered by some philosophers as a sleight of hand. (Jaag and Schrenk 2020, 99, my translation)

I agree: differentiating between two kinds of explanation won't do. A detractor of Humeanism will simply insist, and I believe rightly so, that Loewer's scientific explanations are just mockup explanations, explanations within an agreed set of rules that could change at any time because it depends on something whose diachronic stability is not guaranteed.

Closely tied to the lack of explanatory power is the lack of nomological necessity. True, Humean (BSA) laws entail their instances logically, but they don't necessitate them nomologically. (This makes trouble for the desiderata of inference, counterfactual support, governance, and inner connection, because the natural understanding of these desiderata is that they are fulfilled by some kind of non-logical necessity). Van Fraassen and Lewis try to circumvent this problem by stipulating that nomological necessity, in the case of laws of nature, just *is* logical necessity, effectively, though Van Fraassen's words cloak this identification:

> A statement, *p*, is nomologically necessary if and only if *p* follows from the laws of nature (being the axioms of the best system). Since every statement follows from itself the laws themselves are also nomologically necessary. (Van Fraassen 1989, 44f.)

In effect, Van Fraassen claims that nomologically necessary is whatever Humeanism calls nomologically necessary and so does nothing to appease the worry that we need necessity *in the things* and not just in *statements*. It should also be noted that the insistence of Humeans on their self-styled nomological necessity makes for a deflationary theory: laws collapse into law statements, because the necessity linking laws to their instances is no other than the necessity linking general statements to their particular instances. Koperski voices this worry as follows:

> The best systems approach, like all Humean accounts of law, is not a theory about the laws of nature. It is a theory about law-statements, specifically which generalizations should be given the honorific "law." (Koperski 2020, 90)

More could be said about the problems of Humeanism (as well as Humean attempts to solve them), but I take it that the issues addressed so far raise grave enough doubts about (BSA) Humeanism being a viable metaphysics of the laws of nature, despite its openness to mental interaction.

13.2.2 Laws of nature as necessitation relations (Dretske/Tooley/Armstrong)

The problems of Humeanism described in the previous section make evident the need for accounts that remedy those problems. To understand how the following modal theories do that, it is helpful to adduce a distinction between three uses of the term 'laws of nature', made by John Foster (2001). According to him, the term can refer to either of three things:

1) Theories like Newton's theory of gravitation
2) The basic regularities themselves
3) "Forms of natural necessity" that underlie basic regularities

Sense 1) could also be understood as *law statements*, while 2) refers to the purely empirical observations of regularities in nature. Humeans can agree with these two senses, but will disagree with the third. It is that third meaning, however, that Foster identifies as the crucial one: only if there are what he calls "forms of natural necessity" can the laws of nature fulfill their explanatory role.

Most clearly formulated as a reaction to Humeanism (in particular David Lewis's account) was the theory which David Armstrong presented in his influential book *What Is a Law of Nature?* (Armstrong 2016 [1983]). Very similar accounts, in the same vein, have been proffered by Fred Dretske and Michael Tooley, which is why I shall refer to the common theoretical scaffold as the DTA theory[110]. The core of the DTA theory is that laws of nature are second-order relations between first-order properties, where the properties can be Aristotelian universals (Armstrong's version) or other types of metaphysical entities (like Platonic forms, or tropes etc.). One peculiarity of the second-order relation N(F,G) is that it is itself a second-order universal (or property); another, that it binds F and G together with necessity (modal force). This is precisely where nomological necessity, so direly absent in Humean theories, comes into play. Yet another characteristic of the theory is that *all* laws of nature share one and the same constituent, namely N.
So far, the DTA theory just claims that there is a necessary connection between properties, but not yet between the *instantiations* of those properties. It is therefore a central tenet of Armstrong's that the connection at the level of properties (universals) translates into a connection at the level of concrete instantiations. How this is done shall be one leverage point of criticism. What can be said now already is that the criticism levelled at a particular version of the DTA theory – the one that construes properties as Platonic universals – cannot be applied to the core of DTA, which is non-committal as regards the ontology of properties.

The true problem of the DTA theory lies in the N relation, which Jaag and Schrenk designate as "extravagant":

> The truly extravagant ontological postulate of the theory is the irreducible and non-supervenient higher-order necessitation relation N, no matter to which ontological category it is ultimately to be assigned. (113, my translation)

Of course, ontological extravagance is itself not a sufficient reason to discard a theory, as long as the extravagant entity does its job. However, that is precisely what can be doubted about the DTA theory. We shall come to these sorts of problems – basically, problems pertaining to the *robustness* criterion – shortly. Before that, I wish to discuss briefly how DTA fares with respect to interaction.
As we've seen in section 12.2.1, mental interaction makes it the case, as on all necessitarian accounts, that a different event occurs than the laws (or their proxies) would have brought about with modal force. We also saw that despite this, one is not committed to view mental interaction as breaking a law in this scenario. What is required is that the modal force remains untouched. One way to spell this out is Armstrong's distinction between 'iron laws' and 'oaken laws'. While the former do not admit of exceptions, the latter do: if an F is also an I (with I being an 'interference factor'), then it will not be a G. In this manner, the theory can explain why a massive object (F) made of magnetic iron can be held in mid-air by a strong

[110] There are minor differences between Dretske's, Tooley's and Armstrong's theories which are irrelevant for the present appraisal.

electromagnet (I), where it would normally fall to the ground (~G); one might also apply the 'oaken laws' conception to integrate mental interaction (with interaction being the 'I'). There will be some fundamental 'iron' laws, though, that don't admit of exceptions, which makes it hard to see how interaction could be possible without breaking *those* laws; but then again, I argued above that one should demand of theories of laws of nature that they understand invariance/universality *not* in terms of allowing no exceptions, so one could simply stipulate that DTA laws (even the 'iron' ones) do admit of exceptions (regarding outcomes) in case of non-physical interaction. One could, for example, hold that there is a psychophysical law P that connects a volition V and physical properties F and H without obliterating the N(F,G) relation such that H is instantiated instead of G. An important clause one would have to add is that the occurrence of V is not itself brought about by any further necessity relation, but by a free agent. The N relation would stand for each point in space-time; however, at some points, the modal power of this relation would be overridden by a volition. This proposal arguably neither corresponds to the original purpose nor to the spirit of the DTA theory. As far as I can see, it is still a coherent enhancement of the theory.

However, the DTA theory is not the preferable account of the laws of nature, or so I will argue. The problems lie in the field of *robustness*, albeit in a different manner than in the case of Humeanism.
The first problem is that the nature of the inference from the laws to their concrete instantiations is unclear. John Foster argues that this problem indeed arises, because a simple assertion of the N(F,G) relation won't do:

> My difficulty with it is that I cannot think of any kinds of relationship between universals which would be able to play the role envisaged. It is true that we can, in a purely trivial way, represent a law as a relationship between universals. All we need to do is to formulate the law in the ordinary way, using the nomic operator 'It is a law that', and then replace the predicates which feature in the sentence following the operator by expressions which speak of the instantiation of the corresponding properties and relations. So the sentence
>
> (1) It is a law that anything which is F is G is transformed into
>
> (2) It is a law that anything which instantiates F-ness instantiates G-ness,
>
> which asserts a relationship between F-ness and G-ness. (Foster 2001, 155)

This, Foster claims, is a mere restatement of our intuitive, pre-theoretical concept of laws of nature:

> But obviously relationships of this sort would be useless for Armstrong's purposes. For what he is claiming is that it is precisely by thinking of laws as relationships between universals that we are able to grasp their real nature; and representing laws as such relationships will not help to illuminate that nature if, as in this case, the relationships selected can only be understood in terms of the holding of laws. The difficulty is in seeing what other sorts of relationship are available to play the role that Armstrong envisages. (ibd.)

Armstrong is very explicit about the relationship being *non-logical*, which begs the question what *other kind* of relationship could do the work. Foster is skeptical about the prospect to find a kind relationship that binds together F-ness and G-ness as well as their instances:

> Where a law is of the form expressed by (1)—and, leaving aside probabilistic laws, he is inclined to think that *all* genuine laws can be represented as having this form—Armstrong speaks of the relevant relationship between F-ness and G-ness as one of *non-logical necessitation*. But I just cannot see what he

> has in mind here. He cannot literally mean that F-ness (the entity) necessitates G-ness (the entity); that would make no more sense than speaking of my pen as necessitating Australia. Nor, presumably, can he be thinking of the *existence* of F-ness as necessitating the *existence* of G-ness; for, whatever that would mean, it does not entail the relevant regularity (of all F-things being G). Perhaps what he has in mind is the nomic necessity that whatever instantiates F-ness instantiates G-ness (for I suppose we could *loosely* speak of that as a case of F-ness necessitating G-ness). But if so, we are back, in effect, with (2), and no nearer to an understanding of the nature of laws and the kind of necessity they involve. (155-56)

Jaag and Schrenk concur:

> If the implications between law(fact)s and the corresponding regularities are not simply logical, why do they exist? (Jaag and Schrenk 2020, 116; my translation)

One might postulate a necessity relation (of the non-logical type) between N(F,G) and all its instances. That, however, means one is postulating two “ontologically extravagant” things at once, so *prima facie* not a desirable way to go. And a *secunda facie* investigation makes the difficulties with this approach even look worse.

Let’s, for the sake of argument, posit a nomological-causal (rather than a logical) relation between N(F,G) and (x)(Fx -> Gx). That demands at once a ‘transcategorial’ relation between a second-order universal (N) and concrete states of affairs, another metaphysically extravagant move, albeit not a fatal one. More difficult, however, appears the fact that on such an account, laws have to be treated as causes: N(F,G) is the cause of its concrete instances as well as the corresponding regularities. That, in turn, is at odds with the basic tenet of the DTA theory that laws are just relations between universals (or properties in general). A third problem flows from the fact that the pairing of properties under N as well as the pairing of N(F,G) and (x)(Fx -> Gx) are considered contingent by the DTA theory. That makes it possible that there be a world in which N(F,G) holds, but not (x)(Fx -> Gx). In such worlds, the laws would be ‘impotent’, an entirely undesirable result. The same problem looms, *mutatis mutandis*, for induction: if the connection between laws and their instances is contingent, then it might not only vary across worlds, but also across time.

The most severe problem, however, is a looming vicious infinite regress. In order to settle the necessitation relation between N(F,G) and (x)(Fx ⊃ Gx) one needs an N* such that N*[N(F,G), (x)(Fx ⊃ Gx)]. One may, however, ask at this point how N* ensures that N(F,G) cannot occur without (x)(Fx ⊃ Gx)? It seems one needs yet another necessitation relation N** such that N**{N*[N(F,G), (x)(Fx ⊃ Gx)], [N(F,G), (x)(Fx ⊃ Gx)]} – and so forth. The regress seems indeed to be vicious, because the same question, that of what ensures the necessitation relation on a given level, reappears on each new level.

To come to terms with these vexed problems, Chris Swoyer (Swoyer 1982) has presented what might be called an ‘essentialist’ solution. It consists in saying that it is the nature of N(F,G) that (x)(Fx -> Gx) holds); in other words, that it is part of the nature of F or G, respectively, that certain necessitation relations to F or G hold. That, however, as Jaag & Schrenk point out (2020, 113-14) ceases to be a variant of the DTA theory and should better be classified as a version of dispositionalism. The reason is that this proposal stands in contradiction to the basic DTA tenet that properties have no causal essences (quidditism). By the same token, Swoyer’s account gives up the independence of laws from their instances, a criterion not only on our list but also considered indispensable by leading DTA theorists. At any rate, Swoyer’s proposal sheds light on a seemingly indissoluble tension within the DTA theory:

> So it looks like the independence criterion and the inference criterion are incompatible. (Jaag & Schrenk, 122; my translation)

The difficulties for the DTA theory continue. For instance, counterfactual support seems problematic. It is hard to see how counterfactuals can be supported if the relations between laws and their pertinent regularities are contingent. If, on the other hand, one made them (metaphysically!) necessary, then categorialism/quidditism (again, one of the basic tenets of DTA) is sacrificed, along with independence of laws from their instances.
As far as the explanation of instances by law is concerned, one again bumps into the problem of the unclear relation between laws and their instances. No proposed type of relation will do if the basic assumptions of the theory are to be maintained: a metaphysical relation (see above) stands against the basic tenets of the theory, a grounding relation would identify the laws with their instances (again loss of independence), and a logical relation is out of question anyway.
Consider, finally, the desideratum of governance and independence. Because of Armstrong's Aristotelianism (which has it that only instantiated universals exist), in a lone particle world no N(F,G) could exist, since no relations could exist – the same goes, *a fortiori,* for a completely empty spacetime. This suggests that, on an Armstrongian version of DTA, the laws depend on their instances after all, contrary to the theory's basic independence tenet. It is interesting to note that a Platonic understanding, where uninstantiated universals are allowed to exist, could help out at this point, the unpopularity of Platonic metaphysics notwithstanding.

The following quote by Jaag and Schrenk nicely sums up the two big areas of tension within the DTA theory (the "Humean assumptions" consist mainly in the categoriality as opposed to dispositionality of properties):

> So far, we have discovered two major areas of tension in Armstrong's metaphysics of laws: the first consists in the tension between his Humean assumptions, on the one hand, and his acceptance of a genuine necessity in the world. The second consists in the tension between his naturalistic Aristotelian theory of universals and the acceptance of independent elements controlling natural events. (2020, 132; my translation)

13.2.3 Dispositionalist theories

Dispositionalist theories, also often referred to as powers accounts[111], are first of all theories about the nature of causality and of properties. As such, they hold that the essences of at least some properties consist in being causal. The rival view is *categoricalism*, according to which properties are not essentially causal. According to categoricalism, properties have a qualitative or categorical nature and are individuated not by their causal effects, but by their so-called *quiddities*, where a quiddity is a primitive identity. Dispositionalists typically are discontented with the categorical picture and so one of the main motivations behind dispositionalism is to offer an alternative to categoricalism:

> The main impetus behind dispositionalism is the discontent philosophers have with the categoricalist's doctrine that the essence of a property doesn't have to do with what its instances are disposed to do under various circumstances." (Choi and Fara 2021)

[111] Further terms used to describe what I mean by dispositions are 'power' (Locke's term), 'dunamis' (Aristotle's term), 'ability', 'potency', 'capability', 'tendency', 'potentiality', 'proclivity', 'capacity', among others.

That motivation can take on a principled nature if one adheres to the Eleatic principle according to which only those things should be considered real that make a causal difference to the world.
Dispositional essentialism has it that some properties are essentially dispositional (and thus allowing others, like for example space, to be categorical), while *dispositional monism* sees *all* properties as dispositional. Dispositional essentialism has two core theses: (1) anti-quidditism, according to which properties have substantial essences and (2) the thesis that those essences are dispositional. Dispositional monism (at least Alexander Bird's version) can additionally take on board the third thesis of (3) trans-world dispositionalism of properties. (1) and (2) jointly yield the view that properties are essentially causal, that it is part of their nature to have certain causal effects, which entails that properties can be identified and individuated by their causal effects. *Trans-world dispositionalism* (3) means that the causal essences of properties remain the same in all possible worlds.

Dispositionalism can easily be elaborated into a theory of the laws of nature. On a dispositionalist reading, the laws of nature typically supervene on the dispositions of physical substances[112] (while some dispositionalists, like Mumford (Mumford 2004), eliminate laws altogether), which at least on dispositional monism entails that the laws are necessary, i.e. the same in all possible worlds. There are many versions of dispositionalism about laws of nature; the most elaborate account to date is Alexander Bird's (Bird 2007), and so the following discussion will pivot around his view, while pointing to alternatives where necessary. Bird starts with the core idea of dispositional essentialism (DE_P):

(DE_P): $\Box(Px \supset D_{(S,M)}x)$

where P is any property, x a substance and $D_{(S,\,M)}$ the dispositionality to bring about manifestation M upon stimulus S.
Bird analyses dispositions counterfactually, i.e. were x triggered by S then necessarily, M would come about:

(i) $\Box(Px \supset (Sx \Box\!\!\rightarrow Mx))$.

A universal, lawful regularity can be derived from (i) by first supposing that in some world w it is the case that

(ii) $Px \wedge Sx$

From this one can deduce that in w it is also true that

(iii) Mx

[112] Dispositionalism is generally 'thing-centered' in that it needs to localize powers in particular, individual things. Those are normally called 'substances', although the exact ontology behind this may vary.

(ii) and (iii) allow the introduction of a material conditional, wherefore in *w* the following conditional is true:

(iv) $(Px \wedge Sx) \supset Mx$

Since this deduction has been carried out for any *x*, one can quantify (iv) universally and obtain:

(v) $(x)((Px \wedge Sx) \supset Mx)$

and since *w* could be any world, we even get

(vi) $\Box(x)((Px \wedge Sx) \supset Mx)$

Let's begin, again, by examining the theory's ability to accommodate mental interaction. According to Bird's dispositionalism, a physical substance *p* has the dispositional power *D* to bring about effect *E* in circumstances *C* (which include stimulus *S*), with nomological (trans-world!) necessity such that there is no possible world in which, if *p* is in *C*, *E* will not come about. Now in case a mind intervenes and makes it the case that *p* cause *F* instead of *E* in *C*, we seem to have a problem. First, it might seem that the modal power of *D* is somehow broken. This worry, however, can be alleviated, as we saw in section 12.2.1, by considering *D* still extant but 'trumped' by the mental cause.

There is a second problem, namely that the alteration of one instance of a dispositional power invalidates the pertaining law of nature (or its universal validity), since the laws supervene on the 'dispositional mosaic'. One response strategy is to deny that fundamental, universal laws of nature exist; its most famous defender is Nancy Cartwright (Cartwright 1983). The title of her book, *How the Laws of Physics Lie*, is telling: she thinks that the (fundamental, universal) laws of physics somehow misrepresent what is going on in nature. Cartwright is a dispositionalist, but is far from obvious that dispositionalists, especially when seeking to accommodate interaction, need to take her route. It is likely that her claim of 'lying' laws rests on an equivocation between laws and dynamical equations. Daniel von Wachter comments on this that "she means by 'laws of physics' equations of motion or predictions, because the title means that often bodies do not move as the laws predict" (Von Wachter 2015, 48). Earman, Roberts and Smith further extend this analysis using the example of Universal Gravity:

> Universal Gravity *cannot* misrepresent the motion of a body, because it says nothing specific about such temporal behaviour. Only differential equations of evolution type – which might be derivable from UG together with other considerations – can be integrated to describe the temporal motion of a body or system of bodies. UG cannot be so integrated. Thus, it cannot misrepresent temporal motion. (Earman, Roberts, and Smith 2002, 286)

The above philosophers' distinction is basically one of the two I used above, namely that between laws and dynamical (differential) equations (the other one being the distinction between dynamical equations and initial conditions). So, the question how fundamental laws can emerge unscathed from instances of interaction remains unanswered.

I suggest another approach, namely to alter the dispositionalist supervenience claim slightly by stipulating that the laws supervene not on the actual *effects* (as would be given by the dynamical equations plus initial conditions), but on the *dispositions*, understood in the sense spelled out in 11.2.1. In that case no law of nature would be broken by mental interaction. This presupposes that the dispositions exist as, to borrow Daniel von Wachter's term, 'directednesses'. This is something Bird denies, though. As mentioned, he adopts a counterfactual analysis of dispositions[113], to which the concept of directednesses is alien; however, there are other, non-counterfactual analyses of dispositions, among which Molnar's (Molnar 2003) conception of dispositionality as a *sui generis* relation of physical directedness is the most relevant here (others include nomological (Swoyer 1982) and causal (Shoemaker 1980) analyses).

More flesh needs to be put to the bones of the directedness approach. The crucial point is to make plausible how a disposition can exist without the co-instantiation of a property, the relevant circumstances and the relevant stimulus leading to the manifestation. A helpful elaboration comes from Richard Swinburne, who suggests speaking of substances 'exercising causal influence' rather than 'causing' – the latter being a success verb which ties dispositions to the actual occurrences of their effects. Swinburne calls his account the SPL account (for Substance Powers Liabilities):

> [O]n the SPL account, the powers of substances must be regarded (...) as *powers not to cause effects*, but as *powers to exercise a certain kind and amount of causal influence*. This is because many of the powers which a substance may exercise can be prevented from causing any effect (or as much effect as they would otherwise cause) by the exercise of a contrary power by some other substance. A planet may exert force on another planet, attracting the latter towards it; but have no effect because a third planet is exercising a force attracting it in the opposite direction. So the SPL account must say that substances have powers to exercise causal influence of a certain kind and strength (defined in physics as a 'force') and the liability to do so always or under certain conditions (e.g. in empty space). The effect produced is a resultant of the amount of causal influence exerted by different substances. So, on the SPL account *exercising causal influence* (that is, *making naturally probable*) is a basic category. (Swinburne 2013, 131, my emphasis)

Swinburne's main rationale is thus that the dispositions of substances can be prevented from causing their pertinent effects by the dispositions of other substances. To account for such situations, one can either integrate the dispositions of other substances into *C* and thus get the result that under those circumstances, *D* simply does not bring about *E*; or one could say with Swinburne that *D* consists in exercising a certain kind and amount of causal influence, in other words, make "naturally probable" the effect *E*. This is not to be confused with a *probabilistic* account of events of type *E* where *E* occurs with an objective probability under specified circumstances *C*; rather, in the absence of counter-dispositions, *E* will occur with natural (nomological) necessity, albeit a necessity that can be thwarted by other dispositions without having to alter the dispositional mosaic.

This move, however, is only half the battle. It allows us to view the laws of nature as supervening not on the stimulus-manifestation-relation, but on the powers (dispositions) understood as directednesses: even if the manifestation of a disposition in the presence of the relevant stimulus fails to occur because some other disposition prevented it, that does not mean that any law of nature was broken. But so long as the physical universe *sans* interaction is considered, this result is unsurprising. The really interesting part is how to integrate mental interaction into the picture. I argued in section 12.2.1 that if the nomological directedness

[113] Bird holds the additional claim that the nature of a dispositional property is not only described but *exhausted by* the corresponding counterfactual, in other words, not only $\Box(Px \supset D_{(S,M)}x)$ is true, but also $\Box(Px \equiv D_{(S,M)}x)$.

is ‘trumped’ by mental causation without being obliterated, then interactive dualism does not break laws of nature. It stands to reason to assume that in such a picture, mental causation is likewise a realization of dispositions; and that is exactly what I proposed in section 3.3. Embracing an SPL account for both inanimate nature and mental causation gives us a unified picture. In combination with above picture of laws of nature consisting not in actual effects, but in causal influence (of physical substances), we get an integrative account where mental interaction is possible without either discarding freedom or breaking the laws of nature. At any rate, such a proposal seems more viable than an approach where, to make room for non-violationist interaction, mind changes the dispositions of physical substances, which seems like to strong an interference with nature.[114]

If the above proposal works, it seems that dispositionalism is receptive to mental interaction in a more comprehensive way than either Humeanism or the DTA theory. True, *receptivity* can be had significantly cheaper on Humeanism; but what exacts great efforts is almost always more valuable. In this case, an integration of mental causation and dispositional laws of nature along the lines of Swinburne’s SPL account gives us an integrated view of causation, as opposed to a Humean world in which mental causation (assuming it to be of the dispositionalist type) is of a radically different kind than the ‘causation’ occurring in the inanimate world.

In any event, dispositionalism must still prove itself worthy on the *robustness* front, which should be possible given its modal character. First, it goes without saying that dispositionalism ticks the box of *objectivity* without any difficulty: the dispositions are properties of the physical substances themselves, not concepts imposed on our representations of them. The fulfillment of the *inference* criterion is a bit more complicated, though. True, the inference from law facts (which supervene on dispositions) to regularities is, like in the case of Humeanism, virtually trivial, because laws just *are* regularities (with the difference that the laws supervene on modal facts in the case of dispositionalism). However, if one construes dispositions in a purely structural manner (as Bird does), then one runs into problems:

> Dispositional monism is the view that all there is to (the identity of) any property is a matter of its second-order relations to other properties. [...] In dispositional monism the second-order relation in question is the relation that holds, in virtue of a property’s essence, between that property and its manifestation property—which we will call the manifestation relation. (Bird 2007, 139)

The problem with this is that it raises the question in virtue of what the necessitation relation between the second-order relation (that *is* the disposition) and its instances holds. In fact, this is, *mutatis mutandis*, the same problem the DTA theory encountered: either one postulates the relation as a brute fact or a vicious infinite regress looms. The same goes, by the way, for explanatory power: if dispositions are to explain their instances, with the former depending on the latter (in Bird’s structuralism), then an explanatory circle ensues which could only be broken by positing a third-order relation that binds the second-order relation (the disposition) together with the first-order instances, but that ‘explanation’ requires a fourth-order relation to explain the third-order relation, and so forth.

A solution could be to renunciate Bird’s structuralist thesis and allow that a property *entails* all those second-order relations with other properties but is not *constituted* by them. This would, as far as I see, mean the

[114] Koperski briefly discusses such an approach with respect to divine interaction (Koperski 2020, 93-94), only to dismiss it quickly.

abandonment of dispositional monism and the at least partial re-introduction of categoricalism, something at least Bird is eager to avoid. It should be mentioned that there are attempts to reintroduce a categorical aspect to dispositional properties, see e.g. (Heil and Heil 2003) or (Jacobs 2011). An appraisal of those approaches is beyond the scope of the present investigation; at any rate, Bird's deduction of general laws does not depend on his strong 'exhaustion' thesis (see fn. 113), so that it could in principle be attained on a different conception of fundamental dispositional properties.

Bird's version of dispositionalism makes trouble on yet another account. His is a so-called single-track dispositionalism, which means that each disposition possesses but one single characteristic stimulus-manifestation-pair. The problem is that this construal cannot account for the fixed relation between a range of values of stimulus and manifestation (as it figures, for example, in Coulomb's law). It is interesting to see the link between this metaphysical objection and a historical one that Jeffrey Koperski brings forth. He argues that the modern conception of laws with its possibility for precise mathematical formulation was an important catalyst for science as we know it, and for the abandonment of the dispositionalist Aristotelian framework:

> The point is that the shift from Aristotelian essences to laws spurred the development of empiricism in science – no small matter. (2020, 92)

Indeed, mathematical formulations of laws unproblematically admit of correlating values of properties involved in the formula, often on a continuum. This is a problem for dispositionalism that requires an answer.

One reply that has been offered consists in the positing of multi-track dispositions, where each disposition possesses a multitude of stimulus-manifestation-pairings. However, multi-track dispositions don't really explain the continuum but just give us a 'list' of pairings. A better solution might be *determinable* dispositions, where the disposition is the determinable (analogically, like the color blue) and the stimulus-manifestation correlations are the determinants (like the different shades of blue) (see also Jaag and Schrenk 2020, 150-51). To the extent to which determinants quantify over a continuum as opposed to a set of distinct values can that approach salvage dispositionalism. Indeed, if determinable dispositions are viable, it seems that mathematical formulations might well be an expression of underlying dispositions. The fact that the rise of modern science was intimately connected to an abandonment of Aristotelian dispositionalism would then be at best evidence that it is easier to go about science without racking one's brains about metaphysical underpinnings. Also, Kedar and Hon (2017) make a case that early precursors of modern science in the late Middle Ages (especially Bacon and Grosseteste), themselves still largely committed to Aristotelian metaphysics, did use mathematics in their natural philosophy, thereby easing the hostility between modern law-based science and medieval Aristotelianism.

Let's get to the next desideratum. Counterfactual support seems in general better fulfilled in dispositionalism than in either Humeanism or DTA: laws are metaphysically necessary (across possible worlds), and so there is both a robust basis for counterfactuals (in contrast to Humeanism) and no need to posit a further sort of necessity (which makes trouble for the DTA theory). However, Marc Lange points to two difficulties that dispositionalism has with counterfactuals. First, it makes counterfactuals whose antecedent is contrary to the laws trivially true, because there are no possible worlds in which the antecedent is true (since the laws are metaphysically necessary). That yields implausible results: for example, it is *non-trivially* true that would the speed of light deviate just a little from what it is, then our universe would look radically different. Second, the necessity of laws does not ensure that the same natural kinds exist under all counterfactual conditions. Lange writes:

> How [...] does essentialism account for the further fact that had there been an electron at L, atomic nuclei would still have contained protons – rather than schprotons, which (I stipulate) are like actual protons except for having half their mass? (Lange 2004, 229)

To ensure this, dispositionalists would have to postulate that the same natural kinds exist under counterfactual assumptions, which seems like an *ad hoc* move. The problem of the persistence of natural kinds also affects induction. True, the metaphysical necessity of kinds having their dispositional essences ensures some sort of induction if one allows abduction to count as equivalent (the best explanation for all *F*s being *G*s is that they have an essence *K*). But nothing rules out the possibility of kinds drastically changing over time.

A last criticism comes from Jeffrey Koperski. He points to tensions between the dispositionalist metaphysics and important concepts in physics:

> One reason for this is that dispositions tend to be thing-centered: substances, in a technical sense. A substance is an entity in which properties and causal powers reside according to the dispositionalist. However, many important physical properties are not embedded in material objects (…), center of mass, for example. This is a sometimes-measurable attribute – not merely something that can be calculated, like average height – yet there need be no object that exists at the center of mass of a system. (2020, 92)

And there are more examples: the principle of least action, for example, is a hugely important principle in classical mechanics to restrict the set of possible trajectories of objects to the actual ones, but it is hard to locate within a dispositionalist universe (see (Katzav 2004)). In a similar vein, Alan Chalmers argues that the Lagrangian or Hamiltonian formulations of classical mechanics do not involve any causal statements, and that a redescription in terms of Newtonian mechanics (which can be taken to reflect causality congenial to dispositionalism) is not possible:

> In modem physics Lagrange's equations are interpreted in a more general way than the version of those equations that can be derived from Newton's laws. The energies involved are interpreted in a general way that includes all kinds of energy and not just that which results from the motion of massive bodies under the influence of forces. (…) For instance, the various conservation principles, such as the conservation of charge and parity, intimately connected with symmetries in the Lagrangian function of the energies, are not explicable by reference to some underlying process. (Chalmers 1999, 13)

Or take the by-now familiar conservation principles obtained from the Noether symmetries: they too seem to be natural necessities that cannot be accounted for by dispositionalism (see (A. Chalmers 1999) and (Fine 2002)). Finally, physicists have arrived at robust conceptual and mathematical distinctions in the properties of systems, for example the distinction between force, energy and momentum, that are 'lumped together' by a dispositionalist account. This is why Koperski complains that

> [t]he main concern among philosophers of physics is that the appeal to causal powers is not only "hopelessly vague" (…), it is a step backward in terms of understanding. There are many aspects of modern physics that are now taken for granted, like the difference between force, energy, and momentum. (ibd., 93)

Chalmers proposes a "pluralist" solution to the challenge posed especially by meta-laws: that is, to accept that not all laws of nature have the same metaphysical underpinning. I will offer a theistic formulation of the pluralist solution in chapter 14.

All in all, dispositionalism does well on *receptivity* as well as on most *robustness* criteria. One difficult challenge is to account for the inference from dispositions to their instances, something that might be mendable by the introduction of some non-dispositional aspect to properties. Another one consists in important physical principles that seemingly find no home in a dispositionalist framework. For the remainder of this sub-section, I'd like to point out two further objections to dispositionalism which pertain only to the metaphysics of dispositional properties, but which should be answered satisfactorily to bolster dispositionalism's chances as an all-out account.
The first is the worry that at least dispositional monism entails that nothing ever passes from potentiality to actuality. Thus Stephen Mumford:

> A disposition is a disposition to do something else. But if all properties are dispositional, then this something else must also be dispositional. A disposition is always, therefore a disposition to a further disposition, which in turn is a disposition to a further disposition. This has produced the charge that nothing ever passes from potency into actuality. (Mumford 2011)

Perhaps this problem can be solved by passing from dispositional monism to dispositional essentialism, where the success of such an attempt certainly depends on *which* properties one turns into categorical ones. What may be even more promising is to reintroduce a categorical/qualitative aspect into *all* properties. This is, broadly speaking, the agenda of the family of 'powerful qualities' views. A challenge for this program is to give an account of how a property can be both dispositional and qualitative. One reply is to say that "the qualitative and dispositional are identical with one another and with the unitary intrinsic property itself" (Martin 2010, 65). Choi and Fara hold against this that there is "a strong inclination to think that the dispositional and qualitative natures of a property are distinct from, if not incompatible with one another" (2021). I *prima facie* do not share this inclination, but then again before a final judgment some more investigation would be in order.
The second charge levelled against dispositionalism is that it involves Meinongianism, an ontology that includes even non-existing entities. The point is that on a counterfactual analysis, the disposition is said to persist even if the consequent of the counterfactual never comes to pass (a glass is considered fragile even if it never breaks). However, as we saw above, there are alternatives to the counterfactual analysis. Especially on the directedness interpretation, Meinongianism can be avoided, for the directedness can plausibly be considered something that exists. Alternatively, one might consider admitting possibilities into the ontology, albeit as *existing*. We shall hear more about this in chapter 14.

13.2.4 Maudlin's nomic fundamentalism

The three previous accounts all assign a relational ontological status to the laws of nature: they either supervene on more basic facts (as in Humeanism and dispositionalism), or they *are* second-order necessitation relations (as in the DTA theory). Tim Maudlin's nomic fundamentalism is a peculiar theory in comparison, because it grants laws of nature independent, non-relational ontological status, considers them as fundamental, irreducible entities. In a nutshell, the core of his theory is as follows:

> My analysis of laws is no analysis at all. Rather, I suggest we accept laws as fundamental entities in our ontology. Or, speaking at the conceptual level, the concept of a law cannot be reduced to other more primitive notions. (Maudlin 2007, 18)

Maudlin's view is thus more similar than any other account of lawhood to the view of the early modern thinkers we came across in chapter 11, albeit without the reference to God. The theory is straightforward, so the brief outline just given already captures its gist. Further details given by Maudlin will be filled in along the way.
First, the Gretchen question to each theory about laws of nature: pray tell me, how do you feel about mental interaction? Again, mental interaction passes the first checkpoint of general admission if that which constitutes the modal force (here, the laws themselves) remains intact. This seems even more plausible than on dispositionalism or the DTA theory, since it is presumably less controversial to countenance the continued existence of a particular thing (for which Maudlin's account takes a law of nature) than that of a directedness or a relation once the terminus of the directedness or one of the relata of the relation fail to be instantiated. What does Maudlin himself say to this? On the one hand, he denies the spatiotemporal universality/invariance of laws, but he also explicitly rules out exceptions for so-called FLOTEs (*fundamental laws of temporal evolution*):

> The mark that we believe a law to be a fundamental law is that exceptions are not tolerated. If the law describes the pattern that nature respects in generating the future from the past then those patterns must appear wherever the law is in force. (Maudlin 2007, 45)

The overall picture thus seems to be that some regularities we call 'laws' admit of exceptions because they are not fundamental laws, but the fundamental laws are strictly exceptionless. Now Maudlin does not discuss the possibility of interaction, it stands to reason that he would resist the idea that FLOTEs would make room for interaction; more, his conception of fundamental laws as *describing temporal evolution* (that is, as *dynamical*) bars the solution presented in section 12.1 of distinguishing between the laws, dynamical equations and initial conditions. If this is correct, then his theory would *ex definitionem* rule out interactions. However, it is far from clear that one must identify laws with descriptions of temporal evolution, which means that nomological fundamentalism could accommodate interaction after all; also, a closer inspection of the merits of Maudlin's account as a law-of-nature-theory (regardless of interaction) can reveal interesting insights. I quipped above that Maudlin's theory can be seen as the 'godless heir' of the theistic voluntarist views of the 17th century: laws are supposed to account for the behavior of nature, albeit without reference to a lawmaker. This creates several problems, or so I hope to show.

One group of difficulties cluster around the question what the relation between laws and their instances is. First, it is unclear how a primitive *sui generis* law necessitates or governs the temporal evolution of matter. Maudlin holds that laws are supposed to bring about states productively, but they are not located in spacetime, which makes it *prima facie* hard to see how they can be causally active. At any rate, the causal relationship between laws and physical states must be a different one than between physical states. This unclarity affects at once inference and explanatory power. As regards inference, Maudlin owes us an account of what establishes the connection between a law and its instances. The same goes for explanatory power: as long as the aforementioned connection remains obscure, it remains unclear why we are justified in claiming that the laws explain their instances. To anticipate: none of these problems haunts a theistic account of laws: it is God's power (active or decretal) that ensures the connection between laws (construed as the powers God has endowed physical matter with or as His decrees how nature can behave) and their instances.

Another difficulty arises with respect to independence. Maudlin argues that laws construed as fundamental entities are independent of their instances because incompatible laws could be laws in exactly alike model worlds (i.e. worlds with the exact same distribution of sub-nomic facts). For example, an empty Minkowski spacetime is both a model of General Relativity as well as of Special Relativity in combination with a theory gravitation deviating from General Relativity. On the assumption that one and the same world cannot have contradictory laws, Maudlin concludes that the overall physical state of a world cannot determine the laws (2007, 67). However, in order to get this result, Maudlin needs to make a further assumption: namely that laws are not just *true* in all their models, but that they are also *laws* in those models. After all, Humeanism has it that in order for a world w^* to be physically possible relative to a world w, the laws in w just need to be *true* in w^*. So, in order for the supplementary assumption not to be question-begging against Humeanism, Maudlin needs further support for it. Such support may exist (see Jaag & Schrenk 2020, 3.2), but what I wish to point out here is how a theistic version of Maudlin's view could sidestep this difficulty: if laws are God's decrees, then they are unproblematically independent of their instances, while also being stable (in virtue of God's nature) and endowed with modal force.

A last difficulty concerns the range of laws Maudlin's view can accommodate. In scientific practice, non-dynamical laws such as symmetry principles or the principle of least action play a significant role. But if only dynamical laws can figure as FLOTEs, it is unclear how non-dynamical laws are integrated. Again, such 'static' laws can best be conceptualized as restrictions on the behavior of nature, something a decretal theistic view or the MinP (see section 13.2.7) delivers. I conclude that although (and likely contrary to Maudlin's intentions) nomic fundamentalism can be made interaction-friendly, it is its shortcomings as a nomological theory that prompts one to look for a better alternative. A pattern seems to emerge: somewhat surprisingly, the biggest problems of all hitherto considered theories of the laws of nature lie in their 'core business'.

13.2.5 Dolbeault's panpsychist approach

Laws of nature are to govern physical processes (except on Humeanism), but as we have seen, it is notoriously difficult to spell out just exactly *how* this takes place. The DTA theory is confronted with a vicious infinite regress or, alternatively, collapses into a variant of dispositionalism; Maudlin's fundamentalism can do little more than positing a somewhat questionable causal relationship between laws and their instances. Bird's version of dispositionalism faces problems similar to those of the DTA theory. However, I indicated in that context that if could one understand dispositions differently than as constituted by the higher-order stimulus-manifestation relations – by granting it a truly intrinsic nature that is not exhausted by its causal relations – then governance might after all be satisfactorily fulfilled by dispositionalism. From there, it is only one step to Joël Dolbeault's panpsychist account of the laws of nature.

In a way, Dolbeault (2017) combines the insights from Maudlin's theory with a robust understanding of governance. Dolbeault's rationale can be summed up as follows: if laws of nature govern physical processes, and if laws of nature are something non-physical (Maudlin's view), then something non-physical has been interacting with (physical) nature all along. What we know for sure about governance, on the other hand, is that *we mentally govern (and cause) our own volitional actions.* So why not assume mind in matter in the first place? All the more since we have direct introspective evidence for processual directedness, and that in differing degrees, from the reflexive to the well-pondered, complex action. In other words, one could hypothesize that the regular processes in nature are neither brought about nor governed by ominous entities called 'laws of nature', but by minimally conscious matter acting in a manner remotely similar to us. His account might also be considered 'dispositionalism on steroids', adding to the dispositionalist framework of thinking about laws of nature claims about the inner nature of physical substances.

By way of justification for his rather daring thesis, Dolbeault argues that "the concept of 'law of nature' does not constitute a real explanation" (101). Why he thinks so is probably best summed up in a quote by C.S. Peirce:

> Uniformities are precisely the sort of facts that need to be accounted for. That a pitched coin should sometimes turn up heads and sometimes tails calls for no particular explanation; but if it shows heads every time, we wish to know how this result has been brought about. Law is par excellence the thing that wants a reason' (Peirce 1891, 6.12)

The previous appraisals of important theories should have made clear by now that this analysis cannot be dismissed out of hand. Even Maudlin's fundamentalism, which ascribes the most robust and most unambiguous ontological status to laws, is unable to explain how laws of nature bring about and therefore explain their instances. But surely it should, and most theories (apart from Humeanism) set out to do this, not least by attempting to show how laws govern nature. Dolbeault claims that his account can fill in the gap, and also that "it is more empirical in the sense that it is inspired by our conscious experience, in particular our experience of automatic actions" (2017, 87).

As far as *receptivity* goes, Dolbeault's proposal should have no problems. He explicitly rejects the causal closure of the physical (ibd., 91-92) to make room for the causal activity of non-physical (very primitive) minds. However, the result of the activity of those minds is exactly what we subsume under lawful regularities in nature, so the question how the actions of free agents affect those regularities is still open. But in principle, the door is open, and that is all we need for *receptivity*.

Of course, Dolbeault's account takes on board all the challenges that panpsychism faces, but I wish to focus on another point. Panpsychism gives us some explanation for the regularities of nature, perhaps a better one than other theories, but does it give us a satisfactory explanation? One can still meaningfully ask how it is that minimally conscious chunks of matter behave so regularly, as opposed to more capriciously. Dolbeault seems to think that the more rudimentary consciousness becomes, the more automatic – law-like – its causations become:

> [A] closer study of consciousness shows that it admits of degrees, and that, the simpler and poorer our consciousness is, the more it acts automatically (without reflection), as if it could be reduced to a set of general laws. Hence the hypothesis that we should examine: instead of being governed by laws, could physical processes be governed by a sort of minimal consciousness, only able to determine automatic actions? (ibd., 97)

That sounds reasonable, but the problem with it is that the phenomenal aspect – consciousness – is completely irrelevant for explaining the causal activity of, say, a rudimentarily conscious electron as opposed to that of a non-conscious electron? If, however, we take consciousness as a causal factor more seriously, then I suggest we link it to subjecthood, with the conscious subject being the agent. What we'd get in this case is a continuum of degrees of subjecthood (culminating in human persons) with a scale of consciousness inextricably linked to it. What remains unexplained, though, in this picture, is how it is that degrees of personhood and consciousness are linked, and how it is that more complex subjects have more degrees of freedom than primitive ones, given that those complex subjects arise from the composition of more simple subjects (a consequence of the basic tenets of virtually all variants of panpsychism except for cosmo-panpsychism).

I am not a panpsychist, but I cannot bring myself to dismissing panpsychism as crazy either. I think that, by and large, Dolbeault is on the right track. One worry about his approach is an epistemological one, expressed beautifully by C.S. Lewis:

> To think this [to conceive of the processes of nature the way we conceive of mathematics] is to imagine that the normal processes of Nature are transparent to the intellect, that we can say why she behaves as she does. For, of course, if we cannot see why a thing is so, then we cannot see any reason why it should not be otherwise. But in fact the actual course of Nature is wholly inexplicable. … the very nature of explanation makes it impossible that we should even explain why matter has the properties it has. (Lewis 2000, Kindle locations 10413-10421)

Of course, Lewis's argument is aimed at explanation in the scientific sense – that indeed is impossible, because in explaining nature we have to assume its properties, such that these properties in turn cannot be explained by the same process. Dolbeault does not attempt a scientific explanation in this sense, to be sure, but makes an argument from analogy (between our agency and the behavior of matter). Not everyone will accept this epistemological approach, however: empiricists, for example, will insist that we should rely on sense data only, and that if two theories are equivalent as regards their empirical predictions, they should be adjudicated according to 'hard' principles (like Ockham's razor). But even for non-empiricists (like Lewis) the non-transparency of nature might be an epistemological barrier upon whose crossing one should tread carefully.

Rather than criticize Dolbeault's approach, I would like to draw the attention to another possibility to make sense of his account of the behavior of matter. In a theistic worldview, it would not be surprising if matter had (proto-)mental properties, since it was created by the Supreme Mind. But even apart from this, the existence of a 'ladder' of agency (which is the crucial part of Dolbeault's proposal) would also make perfect sense. On theism, the universe was created by the Supreme Agent; on Christian theism, human beings were even created in the image of the Supreme Agent. It then stands to reason that the latter share more of the attributes of God, especially regarding freedom, while inanimate matter, whether or not conceived of as conscious or even as agential, would constitute a 'medium' in which embodied human agents can act, a medium with its own reality (independent of human minds, see the objectivity criterion) that can be acted upon but whose rules (regularities, laws) cannot be manipulated.

More on this will be said in chapter 14. For now, I hope to have at least gestured (not only in this section) towards how a theistic worldview can remedy the ailments that naturalistic theories suffer from, despite their achievements.

13.2.6 Lange's counterfactual theory

We come at last to two modal accounts that make *primitive* modal facts the truthmakers of laws, rather than modal facts which can be further analyzed (as in dispositionalism, the DTA theory or Dolbeault's panpsychist approach). The account discussed in this section is Marc Lange's (Lange 2009) theory, which makes counterfactual stability the centerpiece.

The basic idea is that laws of nature are counterfactually stable generalizations. This feature is what distinguishes laws of nature from mere accidental generalizations. For example, "all my cars are SEATs" is an accidental generalizations, because it is not counterfactually stable:

1) If I had purchased only cars of the brand SEAT, all my cars would be SEATs. (true)
2) If I had purchased a BMW today, all my cars would be SEATs. (false)

The consequents of (1) and (2) are context-sensitive. Changing the context can change the truth value of the counterfactual. By contrast, laws of nature are insensitive to context, i.e. counterfactually stable:

3) If I had purchased only cars of the brand SEAT, the speed of light would be 299792 km/s in vacuum. (true)
4) If I had purchased a BMW today, the speed of light would be 299792 km/s in vacuum. (true)

Both counterfactuals come out true, independent of the antecedent. Although Lange allows for a hierarchy of laws, such that laws not ranking highest can be falsified by suitable counterfactual conditions, there must be some most fundamental laws that hold entirely regardless of the antecedent. Counterfactual stability is also where Lange gets the nomological necessity of laws of nature:

> Accordingly, I suggest that membership in a nontrivially stable set is what it means for a truth to possess a kind of necessity. Intuitively, "necessity" is an especially strong sort of persistence under counterfactual perturbations. (Lange 2005, 425)

"Trivially stable" would, in particular, apply to statements that contained all the sub-nomic facts: of course, a counterfactual containing the sum total of all the sub-nomic facts as consequent will come out true no matter what the antecedent, not least because the antecedent will be trivially false (since it will contain only statements contradictory to what is in the consequent).
It is important to realize that Lange refuses to give any further analysis for those conditionals that make the most fundamental laws of nature true. He insists that those counterfactuals are "ontological bedrock":

> I propose that with these subjunctive facts we have reached ontological bedrock. They are [...] primitive, lying at the bottom of the world. They are the lawmakers. (Lange 2009, 136).

Let's begin the appraisal of Lange's theory by its fulfillment of *receptivity*. Is there a way to construe the modal force involved as intact upon interaction? The difficulty here is that the modal force proceeds from counterfactuals, which are basically neither things nor relations nor directednesses, but statements. And it is exactly those statements which are falsified by interaction, since interaction renders those otherwise stable counterfactuals false:

5) If I had done a double-slit experiments with photons today, their time-dependent Schrödinger equation would be $i\hbar \frac{d}{dt} \mid \Psi(\mathrm{t})\rangle = \widehat{H} \mid \Psi(\mathrm{t})\rangle$ (true)
6) If I had not done a double-slit experiments with photons today, their time-dependent Schrödinger equation would be $i\hbar \frac{d}{dt} \mid \Psi(\mathrm{t})\rangle = \widehat{H} \mid \Psi(\mathrm{t})\rangle$ (true)

But

7) If my mind had changed the trajectory of the photons, their time-dependent Schrödinger equation would be $i\hbar\frac{d}{dt} \mid \Psi(\mathrm{t})\rangle = \hat{H} \mid \Psi(\mathrm{t})\rangle$ (false)

I deliberately chose a dynamical equation, and a fundamental one at that, at least in respect of our current state of knowledge. With other candidates for fundamental laws, it is less obvious that interaction falsifies the pertinent counterfactuals. For example, is it clearly false that

8) Had God made the earth revolve around the sun twice as fast today than yesterday, $F = ma$?

I don't think so. God may be exerting a force to give the earth a push, such that $F = ma$ will be satisfied; and the same goes for other force laws. As far as principles like the conservation of energy or momentum (Lange calls them 'meta-laws') are concerned, they are conditional to begin with, and therefore no obstacle to interaction (see chapters 5 and 7). However, if dynamical equations are among the fundamental laws of nature, then it seems that Lange's demand for absolute stability and the desideratum for receptivity are mutually exclusive. But recall that an absolutely exceptionless conception of laws is not eligible to begin with, on pain of begging the question against interaction. So, we could either add to the theory LC as qualification or acknowledge the distinction between laws and dynamical equations to salvage *receptivity* for counterfactualism. I have already cast my vote for the latter strategy.

Let's get to the examination of Lange's account regarding *robustness*. By and large, it performs quite well: according to it, laws of nature are perfectly objective; inference flows naturally from it, because Lange identifies (via the counterfactual stability test) a subset of the existing facts as laws (such that if *p* is a law, then *p* is true *a fortiori*). Invariance and universality are also fulfilled, although Lange allows for some flexibility regarding spatial curiosities. Counterfactual support is of course the very core of Lange's theory, although strictly speaking, Lange's laws do not *support*, but rather *are* counterfactuals. More problematic is the desideratum of explanatory power. Here, Lange's theory runs counter to our intuitions. We normally explain a counterfactual with a law of nature; Lange, however, grounds laws of nature in counterfactuals, and refuses to give any deeper analysis or explanation of these counterfactuals. Jaag and Schrenk agree:

> This may seem counterintuitive to many readers. One might ask for an explanation of why these conditionals are true. (2020, 190, my translation)

Koperski complains that Lange's account puts the horse before the cart when it comes to the order of explanation:

> It is not counterfactual stability that lets us make inferences about the laws. It is our knowledge of the laws that grounds predictions about counterfactual stability. (2020, 96)

In his defense, Lange points to the fact that epistemic and ontological priority can come apart (Lange 2009, 136). But that is only a necessary and not a sufficient condition for his theory to be true. The demand for truthmakers of the counterfactuals can still meaningfully be made.[115]

As regards governance: because of their necessity, laws of nature do 'govern' nature. However, the necessity is limited to the reduction of possibilities; productive generation of events is beyond the theory. The criterion of inner connection is likewise not wholly unproblematic, for local connections between sub-nomic facts are not provided. Instead, the connections rely on global counterfactuals, which again goes contrary to our intuitions that something must connect events locally. A last problem for Lange is the fact that some counterfactuals are picked out as laws which no one in his right minds would consider as laws. For example, the statement "all ravens are black" seems a reasonable candidate for a counterfactually stable law; however, its logical equivalent "no non-black thing is a raven" would have to count as a law of nature, too, which is outlandish.

All in all, the counterfactual theory captures very well the stability of laws nature, especially in contrast to accidental generalizations. However, it does so at the expense of explanatory plausibility, both at the level of depth and at the level of order, because it stipulates sufficiently stable counterfactuals to be ontological bedrock, rather than ground them in something further (which could then remedy the problems the theory has with inner connection and the awkward selection of certain statements as laws). The theory raises one interesting possibility to conceive of laws, though: namely as constraints of possibilities. This does not run against the spirit of the governance desideratum (for governance can also plausibly occur via the reduction of possibilities) and avoids some vexed questions in connection with the generation of the instances of laws (see the discussions of the DTA theory and dispositionalism above). The next and last account even centers on this very notion of laws as constraints.

13.2.7 Chen and Goldstein's Minimal Primitivism (MinP) about laws

Eddy Keming Chen and Sheldon Goldstein recently published a paper titled "Governing Without A Fundamental Direction of Time" (Chen and Goldstein 2022), in which they lay out their novel account of laws of nature. They consider laws as real and irreducible (herein resembling Maudlin's nomological fundamentalism), but also as no more than constraints on the possible histories of the world. The laws thus involve no 'dynamic production'. By the same token, Chen and Goldstein do not assume causation as fundamental and hence reject the assumption of time asymmetry. They call their account "minimal primitivism" (in short MinP) about the laws of nature.

A few more words on the "possible histories" approach. On MinP, laws govern by constraining the possible histories of the physical world. The laws not only fulfill a selective role with respect to these possible histories, but also an aesthetic and practical role: instead of having to give a (possibly infinite) list of possible histories, one can, more elegantly and parsimoniously, write down simple laws (whose simplicity may be unexpected, and may create complex patterns). An explication of the governing relation in MinP is also in order before we move on. As I mentioned, governing must not be understood here as *bringing about*, but it

[115] Interestingly, Woodward defends a counterfactual theory in which dispositions are the truthmakers for the counterfactuals (Jaag and Schrenk 2020, 191). I suspend judgment on whether this makes him qualify as a dispositionalist.

should also not be conceived of in a way similar to human (political) governments. Rather, Chen and Goldstein invoke the analogy with mathematical laws to convey what they mean by 'governing'.

Now, how does MinP fare with respect to *receptivity*? Since MinP is a paragon of a necessitarian constraint account, we are dealing with the situation described in fig. 12.5. For a MinP law not to be broken it would have to be true that the constraining modal force not be obliterated. I argued in section 12.2.1 that this is a viable possibility; naturally, Chen and Goldstein do not deal with the question (since they are concerned for an account of the regularities in nature, not exceptions to them). Of course, we could always make use the *ceteris paribus* reply LC, but again I urge that it is better to rightly understand the status of the laws than to trim their purview: in this case, by invoking the distinction between laws and initial conditions. What we then get is something like this: the laws, since they constrain the possible future histories relative to initial conditions remain unscathed if an interaction occurs, because interaction ever only alters the initial conditions, and because interaction does not annihilate the constraining modal force.

MinP can really play to its strengths in the domain of *robustness*, of course. MinP laws are objective (*ex hypothesi*), unproblematically allow for inference (from a constraint statement in terms of a mathematical formula one can easily derive which sub-nomic facts result in function of the initial conditions), and they are almost by definition universal/invariant. Counterfactual support can also be easily attained, because the possible histories approach fixes which counterfactuals are true and which are false. Another advantage is that MinP can accommodate dynamical (Newton's laws, Schrödinger equation) as well as non-dynamical law statements (e.g. Noether symmetries, Principle of Least Action, Einstein GR equation, Past Hypothesis, even closed timelike curves). It is unrestrictedly open to the mathematical form of laws.
Chen and Goldstein point out some further perks of their account in contrast to other notable theories of laws of nature. With respect to the DTA theory, they argue that MinP is superior to it when it comes to accommodating differential equations:

> [D]ifferential equations conflict with Armstrong's principle about the instantiation of universals, as the values of the derivatives are calculated from values possessed by non-actual states (those in the small neighborhood around the actual one) that are not instantiated. In contrast, MinP has no difficulty accommodating laws expressed by differential equations. (ibd., 36)

Concerning Bird's dispositionalism, they claim that their account solves four problems implied by Bird's theory (of which I touched on (2) and (3) in my discussion of dispositionalism above):

> Finally, there are problems specific to accounts (such as Bird's) that analyze laws in terms of dispositions. Bird (2007) lists four problems (p.211): (1) fundamental constants, (2) conservation laws and symmetry principles, (3) principles of least action, and (4) multiple laws relating distinct properties. Problem (1) arises because slight differences in the constants do not require the properties to be different; problem (2) because conservation laws and symmetry principles do not seem to be manifestations of dispositions; problem (3) because the principles seem to commit to the physical possibilities of alternate histories, something not allowed on dispositional essentialism; problem (4) because a third law relating two properties will not be the outcome of the dispositional natures of those properties. (ibd., 37)

For all these advantages, I see some worries concerning explanatory power. Chen and Goldstein claim that their account trumps Humeanism because it gives a robust metaphysical explanation for the behavior of nature:

> There are fundamental, objective, and mind- independent facts about which laws govern the world, and we can be wrong about them. This is not a bug but a feature of MinP, symptomatic of the robust kind of realism that we endorse. For realists, this is exactly where they should end up; fallibility about the ultimate reality is a badge of honor! (ibd., 33)

However, the following quote from David Lewis equally applies to MinP:

> If nature is kind, the best system will be robustly best—so far ahead of its rivals that it will come out first under any standards of simplicity and strength and balance. We have no guarantee that nature is kind in this way, but no evidence that it isn't. It's a reasonable hope. Perhaps we presuppose it in our thinking about law. I can admit that if nature were unkind, and if disagreeing rival systems were running neck-and-neck, then lawhood might be a psychological matter, and that would be very peculiar. (Lewis 1994, 479)

The reason is that MinP makes it no less clear than Humeanism why laws of nature are as simple and elegant as they seem to be. It must be asked: why expect there to be fundamental constraints that apparently admit of being captured in simple, elegant equations, rather than in incredibly complex equations or even just in a giant list of allowed events? Theism yields an obvious answer in this case. We see once more how theism does make a difference (*pace* Hall) by firmly grounding the virtues of a theory in ultimate reality.

There is one last conundrum – 'problem' seems too strong a word – I wish to point to, albeit one MinP shares with Lange's theory: if laws do not govern in a generative sense, then what is responsible for the motion of matter? Consider an analogy: my control ('governance') of the steering wheel constrains the trajectories my car can take, but it will not go anywhere unless propelled by an engine. In this analogy, the engine governs 'productively', while I as the driver govern by constraint. What is the engine's analogue in nature? Neither Chen and Goldstein, nor Lange, nor Jeffrey Koperski (see chapter 14) address this question. Perhaps one could posit a certain (very large) amount of energy as being part of the initial conditions of the universe in a brute-fact manner. This initial energy would then provide the 'engine', potentially able to produce an infinite number of histories, albeit constrained by the laws. This may be a possible answer, but then again we may be able to come up with a better answer. That must, however, wait until chapter 14. Before that, I wish to conduct a historical analysis which shows to which degree the notion of 'laws of nature' depends on theism both historically and conceptually.

14. A theistic account of the interplay between laws of nature and interaction

A somewhat surprising result of chapter 13 is that the most severe problems for all examined nomological theories lie in their ‘core business’, namely accounting for the regularities in nature. This is notoriously true of Humeanism but goes for the other accounts as well. The DTA theory as well as Maudlin’s fundamentalism and at least Bird’s version of dispositionalism have trouble establishing the link between the laws and their instances, where dispositionalism seems to have the greatest promise to be able to mend those problems, although this requires going deep into metaphysical questions about the nature of properties. Another, *prima facie* less amenable difficulty of dispositionalism is that it seemingly cannot integrate fundamental concepts of physics.

The counterfactual theory must, counterintuitively, posit counterfactuals in a brute-fact manner. MinP and Dolbeault’s panpsychist account, finally, come closest to a satisfactory account in different respects: MinP construes laws as constraints and thus avoids many of the problems the other accounts suffer from, while Dolbeault’s account is strong on grounding the governance of laws. It is first interesting to note that the two accounts thus serve directly opposed intuitions (laws as productive vs. non-productive/constraining). Beyond this, both again face difficulties: MinP can account for the *elegance* of the laws no more than Humeanism, and Dolbeault’s theory takes on board all the challenges inherent to panpsychism.

I have repeatedly insinuated that a theistic account might solve all the aforementioned problems, and thus, in a way, integrate the ‘best of all worlds’. The hope is to ‘kill two birds with one stone’, namely to provide not only a solid account of the laws of nature, but also of the interplay between natural regularities and human free actions, and thus in a profound way to make sense of these two fundamental aspects of reality.

14.1 Theism as the best explanation

That nature evinces stable and mathematically elegant laws is undisputed. The question is how those laws are best explained. All extant theories of the laws of nature seek to do this, with the exception of Humeanism. Ned Hall (himself a Humean) argues that it makes no difference whether - once one has opted for a necessitarian account – one posits God behind the modal forces at work in nature or not:

> [Someone could have] a conception of power according to which whether something (in this case, God) has it is not a fact that reduces to the totality of nonmodal facts. Equipped with such a conception, she can formulate her preferred divine command theory. And she could just as well have formulated any of a number of distinct, and decidedly secular, anti-reductionist theories (e.g., by attributing powers to objects). (Hall 2009, 22f.)

Much, however, depends on the exact meaning of “just as well”. If one is interested only in finding a satisfactory theory for the behavior of nature, then one may be content with a secular theory; however, if one wishes to go deeper and explain the *existence* of the laws or of that on which the laws supervene, or the shape of the laws, then non-theistic theories quickly reach their limits. Also, *making sense* of the way nature behaves (and how free agents fit in) is clearly beyond secular accounts. In this section, I will only address the explanatory aspect of theism. In section 14.3 I will say some things about the question of how to make sense of the overall picture of a law-governed world in which non-physical minds can intervene.

According to John Foster, the best explanation for the existence of laws of nature is that they have been put in place by God. He writes:

> The regularities in nature, simply by being regularities, call for explanation. There are only two ways in which we could, with any plausibility, try to explain them. One way would be to suppose that they are imposed on the world by God. The other would be to suppose that they reflect the presence of laws of nature, conceived of as forms of natural necessity. But the only way of making sense of the notion of a law of nature, thus conceived, is by construing a law as the causing of the associated regularity, and the only remotely plausible account of such causing would be in terms of the agency of God. (2001, 145)

His tenet that the regularities in nature "call for explanation" makes him at once rule out Humeanism, since Humeanism doesn't even attempt to give an explanation. We discussed the pertinent arguments in chapter 13. The interesting point Foster makes is that even *necessitarian* theories face the same problem, namely failing to give a satisfactory explanation of the laws of nature; that, he thinks, is only possible if one admits God into the picture.

How does Foster arrive at this conclusion? He starts from an exclusive disjunction: either the laws are directly imposed on the world by God, or they are "forms of natural necessity" which ultimately also require God as explanation. The first disjunct roughly corresponds to the early modern voluntarist view explained in the previous chapter. The second disjunct could be realized by one of the necessitarian accounts laid out in chapter 13. Foster claims that a necessitarian theory only makes sense if the laws (construed as forms of natural necessity) cause their regularities (i.e. sub-nomic facts), and that in turn can ultimately only be accounted for in terms of God's agency.

I largely agree with Foster, though some refinement of his argument is in order. It is true that causation in the sense of generative causation is an obvious way to spell out how laws explain their instances; however, counterfactualism as well as MinP do not employ such a notion. This differentiation does not, however, change the fact that all those accounts suffer from the lack an ultimate explanation: the DTA theory cannot satisfactorily account for how the necessitation relation N binds instantiations of properties together; dispositionalism is at a loss to explain the stability of essences; nomic fundamentalism does not explain the origin of the reified laws nor how they relate to their instances; counterfactualism is silent about how counterfactuals can be fundamental, and MinP owes us an explanation of the astounding shape of the existing laws. Against this backdrop, I propose that God is a very good, if not the best explanation for the (truthmakers of the) laws. Let us see how this can go.

14.2 A theistic account - metaphysics

From the discussion of the previous chapter, dispositionalism and MinP emerged as the two most promising candidates for an adequate theory of the laws of nature. I will now combine both accounts with theism to see which one yields the more satisfactory result. But first, a preliminary observation is in order. The two metaphysics under discussion here correspond to two basic options how the laws are ultimately explained by God. Either God directly imposes them on nature (corresponding to the MinP approach), or He creates some intermediaries (in this case dispositional essences) which ensure the regular behavior of nature. The former stands in the danger of collapsing into occasionalism, which is the view that all events in nature are directly, generatively brought about by God. There is relatively widespread consensus that occasionalism is not a good solution, which is why in the following inquiry a guiding principle will be to avoid occasionalism.

14.2.1 Divine decretalism

Chen's and Goldstein's MinP, is probably the account that comes closest to being a secular version of the early modern view. The main point is that MinP neither posits intermediaries like dispositions or reified laws nor relies on counterfactuals as the basis for order in nature. In a theistic framework, the constraints with which MinP identifies laws of nature can be regarded as *divine choices*, or *impositions* on nature: nature could behave in a number of ways, but God cuts down the possibilities to one or to at least a significantly smaller number.

Jeffrey Koperski's 'divine decretalism' seems to embody a theistic version of MinP and thus constitutes an interesting competitor for an integrative theory of the laws of nature that makes room for interaction. (It should be noted from the outset that for this reason the objections levelled at divine decretalism equally apply to a theistic version of MinP). Let us therefore now turn to an appraisal of divine decretalism.

The secular MinP account left us with two main questions. The first is, if laws are just constraints on a space of possible events, then what or who accounts for the mathematical elegance of those constraints? Secondly, if laws of nature don't 'make things happen', what *does*?

Jeffrey Koperski's *divine decretalism* (2017, 2020, ch. 5.5)[116] is a theistic account that answers the first question with "God". The second question remains unanswered, which prompts an enhancement of the view (more on that below). According to divine decretalism, the laws of nature are "no more than God's choices for how physical events shall proceed" (Koperski 2020, 99). Although decretalism arguably does not entail it, Koperski combines it with what he calls the "neoclassical model" of divine interaction (Koperski 2020, ch. 7.2). The neoclassical model is basically the model I presented in section 12.1, and so rests on the two closely related pillars we've already come across: first, the non-equivalence of laws and dynamical differential equations; and second, the distinction between laws and initial conditions. For Koperski, dynamical equations are that which derives from the laws. Thus, Koperski's account is a refinement of both Plantinga's (Plantinga 2008) and Foster's accounts, both of which largely rely on a *ceteris paribus* conception of laws:

> I indicated earlier that I accept the possibility of miracles, in which something occurs that contravenes the laws of nature. To allow for this possibility, in the context of the account now envisaged, we only need to suppose that, in imposing the regularities on the world, God builds in a suitable qualification. Thus, for each such regularity, we must suppose that what God prescribes is that things are always to conform to that regularity except on those occasions, if any, when he chooses to intervene. In this way, God would both create a framework of law, embodying, as it were, the default prescriptions for the functioning of the universe, and leave himself discretion to override this framework on any particular occasion. (Foster 2001, 160)

Above distinction allows Koperski an elegant accommodation of interaction; in fact, that is one of his main interests (he is concerned with divine interaction, which is sufficiently parallel to mental interaction in the present respect). So, even if a dynamical equation should be 'falsified' by mental or divine interaction, the underlying law is not affected. He gives the (non-interactionist) examples of quantum fluctuation and quantum tunneling to make a case that the appearance of particles 'out of nowhere' is not in contradiction to the laws – and that is because whatever structure there is in place, whether it be a gravitational or an electromagnetic field, it can accommodate interaction (Koperski 2020, 135). The laws can be said to

[116] According to Koperski (2020, 100), Alvin Plantinga is the first to call this kind of view 'decretalism'.

accommodate interaction in the sense that interaction alters the initial conditions to which the differential equations (which are in turn derived from the laws) apply; laws (and dynamical equations) and initial conditions can and must be distinguished because neither the laws nor the dynamical equations specify the initial conditions:

> Under the Euler procedure, neither force laws nor the laws of motion themselves specify any outcome. (ibd., 144)

Koperski uses the (in my eyes somewhat misleading) term of *adaptation* to describe this triangle between laws, initial conditions and interaction:

> If you want a slogan, it is this: The laws never break; they flow. The laws adapt to change. This was true when we thought that nature was Newtonian, and it remains true in the age of quantum mechanics and relativity. (...) Changes to nonnomic conditions do not violate the laws of nature. Nature allows for change that the laws can seamlessly adapt to. (ibd., 135)

The distinction between initial conditions, dynamical equations and laws is crucial for the possibility of interaction on divine decretalism. But is it sufficient? At closer inspection, the solution is not as smooth as it might seem. Let's therefore turn to this and other challenges for Koperski's theory.

As I pointed out at the end of section 12.1, the initial conditions of one instance of the laws are the outcomes of a previous instance of the laws. Thus, by changing the initial conditions of one instance, interaction changes the outcome of a previous instance, and thereby falsifies a dynamical equation. I also argued, however, that this does not signify genuine law-breaking yet; for that, it takes the obliteration of the underlying metaphysical reality that constitutes the laws. We saw further in section 12.2 that law-breaking need not be assumed even on necessitarian accounts as long as that which constitutes the modal force is granted continued existence. This, however, might not be as easily realizable on divine decretalism as on the other necessitarian accounts. *Divine* interaction would most certainly mean a 'repeal' of the decree by God Himself, which one might consider a 'breach of law', but in another way it is very different therefrom, because the repeal has been effected by none other than the lawgiver. *Mental* interaction, though, is a more complicated case. Here, a *human* choice interferes with a divine choice, and divine choices are *ipso facto* effective due to divine omnipotence, which (*per impossibile*) would mean that a human free action thwarts a divine choice and therefore breaks a law of nature. We will come back to this problem in section 14.2 where I explore the possibilities of remedying Koperski's account.

Another objection has been voiced by Robert Larmer (Larmer 2017). He argues that Koperski's decretalism either collapses into occasionalism or requires the intermediaries between God and nature it seeks to avoid. Larmer grounds his critique in the following question: how does God implement his choices concerning the proceeding of physical events? Remember that Koperski rejects all intrinsic causal powers (à la dispositionalism) but also the construal of laws as reified intermediaries. So what does he think is the relationship between laws as divine decrees and their instantiations? Koperski holds that "one could take God's declaration of law as being prior to creation. Particles, bodies, and fields forevermore act accordingly" and that "there is no need for God to continually step in and repeatedly assert his will, as if nature forgets from one moment to the next" (2017, 98). Larmer finds this unconvincing. He argues that talk of remembering with respect to particles and fields is anthropomorphic and merely metaphoric (Larmer 2017, 446), and that Koperski is thus either committed to occasionalism (God directly bringing about token physical events) or to some sort of intermediaries (e.g. his 'once-and-for-all' divine decrees).

But since Koperski rejects intermediaries, he seems committed to occasionalism. I agree with Larmer's analysis, although I would like to refine it: Koperski is committed to either occasionalism or intermediaries as long as he doesn't tell us what the generative force of the motion of matter is. In that case, either divine decrees (occasionalism) or intermediaries have to take on that role.
In fact, Koperski makes an interesting statement which, unfortunately, he does not elaborate on:

> God's nomological decrees are necessary but not sufficient conditions for almost all physical events, much as the generator is necessary but not sufficient for the motion of the iron. (2017, 99)

If indeed God's decrees are *not* sufficient conditions for physical events, then occasionalism can be eschewed. But again, *something* must constitute the sufficient condition for physical events. What could that be if dispositions and reified laws are rejected? Koperski's position on this is not entirely clear. On the one hand, he ascribes sufficiency (if not explicitly in those terms) to forces and energy:

> Laws *never* make things happen. That way of thinking, again, is rooted in a causal powers intuition. What, then, does account for change? There is no reason to turn to metaphysics for an answer. Forces and energy are responsible for moving systems from state to state, including forces that come into play at higher levels. Physics has solved the question of why particles, fields, and the systems composed of them are active. There is no need for causal powers, Aristotelian natures, or laws to answer that question. (2020, 103)

I am not so sure that there is "no need to turn to metaphysics". For example, the very existence of forces has been contested (e.g. (Landry et al. 2017; Lazarovici 2018; Esfeld 2020)) and the ontology of energy is a notoriously vexed issue. The only undisputed thing about forces and energy is that they are useful and proven theoretical concepts. However, to explain the motion of matter, we need more than concepts.
Interestingly, Koperski also mentions another principle in this context that does give us some direction: Murray Gell-Mann's *totalitarian principle* (Gell-Mann 1956, 859). According to this principle, anything that is not forbidden may happen. If we take it seriously, then the fact that *not* everything happens is due to the fact that some things *have* been forbidden, either by logical impossibility or by God's decrees. In other words, the crucial *explanandum* is no longer what takes a physical system from state A to state B, but why it ends up in state B rather than in state C. To put it yet more colloquially, if we take Gell-Mann's principle seriously, then we do not need to care about the car's engine, for it is present and running anyway; what we should focus on is the steering wheel and, perhaps, the driver. But of course, the claim that "anything that is not forbidden may happen" is a stunning idea that stands in need of explication. How can we metaphysically conceptualize that, were it not for divine constraints, a physical system could behave in a (quasi-)infinite number of ways? Koperski doesn't say. I will attempt to fill in this gap in the next section.
Larmer, at any rate, concludes that given the diagnosed dilemma of Koperski's account, we should reject it, because it does not yield any benefit over and above a view that merely distinguishes between laws and initial conditions and remains non-committal otherwise. In a prior work, Larmer points out:

> This basic distinction between the laws of nature and the stuff of nature suggests that miracles can occur without violating any laws of nature. If God creates or annihilates a unit of mass/energy, or simply causes some of these units to occupy a different position, then He changes the material conditions to which the laws of nature apply. He thereby produces an event that nature would not have produced on its own but breaks no laws of nature. (Larmer 2014, 39)

Thus, for Larmer, Koperski's decretalism creates additional problems without solving any that have not already been solved by the laws/initial conditions distinction. However, *pace* Larmer, a metaphysically non-committal view is far from giving us 'all we want'. As argued in chapter 11, the distinction may mask actual law-breaking in the metaphysical substructure, at least for all non-Humean accounts (and I presume that Larmer won't be satisfied with Humeanism either), so specification of the metaphysical details is unavoidable. Waiving the metaphysics makes the laws some kind of 'black box' which yield outcomes when being fed with initial conditions. What happens inside the box is significant, though. Apart from the question of law-breaking, one difficulty most of the generative necessitarian theories have is how to establish the connection between the laws and their instances. Hence, there is a doubly vested interest to spell out what is happening 'inside the box'. Conversely, if we can find a satisfactory metaphysical account, then that's certainly preferable to a quietist or agnostic stance.

We saw that the pressing challenge for Koperski's theory is to spell out what is the sufficient condition of the motion of matter. Without an answer to that question, he cannot avoid the charge of occasionalism, or of having reintroduced (against his stated intent) laws as intermediaries between God and nature. Koperski mentions Gell-Mann's totalitarian principle as an answer to that question, without going into more detail. So, the question is how that principle can be cashed out metaphysically? Or are there perhaps quite different ways to reach supplement his theory?

Let's call the view derived from the totalitarian principle the 'totalitarian' view. In line with the totalitarian principle, the totalitarian view claims that a physical system may, barring further qualifications, behave in any way whatsoever. This can hardly mean that it behaves in all those ways *at once*; it should rather be interpreted as "*could* behave in any way whatsoever", in other words, we are talking about *possibilities* or *potentialities*, metaphysical concepts one can work with. I hence suggest expressing the totalitarian view as follows:

Physical systems possess (quasi-)infinite possibilities or potentialities.

This, it must be stressed again, describes the state of affairs prior to any constraint.

The totalitarian view at once raises a deep metaphysical question: what is the ontological status of possibilities? Do they exist, and if so, do they exist in the same way as actualities exist? Several positions have been taken on this matter; on one extreme, there is Megarian actualism which denies that possibilities exist in any way whatsoever and affirms the existence only of actual states of affairs. On the other hand, there is the view that possibilities do exist in the actual world. This view is for example defended by Alexander Bird, who spells it out for the special case of dispositions (Bird 2007, 109). (Dispositions can be seen as possibilities already preselected according to a stimulus-manifestation scheme; possibilities *tout court* are what may happen regardless of any scheme). At any rate, I will work with the following axiom of real possibilities:

(RP) Unactualized possibilities exist in the actual world.

RP is of course driven to its extreme in the totalitarian view: not only do possibilities exist, but *all* possibilities exist.

An objection that can be mounted against RP is that worlds in which a set of possibilities is unactualized are indistinguishable from worlds in which those possibilities don't exist. However, as I've argued in section 13.2.3, this problem persists only if (1) the nature of properties is exhausted by their causal effects and (2) if there is no one to have insight into the nature of 'dormant' causal properties. There are good grounds to reject (1), and (2) is resolved *ex hypothesi* on theism. I therefore take it that RP gives us what we need as sufficient condition for the motion of matter within a divine decretalist framework. In that framework, divine decrees play the role of actualizing some possibilities while leaving the others unactualized.

Now, does the totalitarian view in combination with RP constitute a sufficient condition for the motion of matter? I cannot see that it does. For the existence of possibilities does not yet specify what actualizes some of them. Koperski insists that divine decrees are merely constraints on an unspecified 'striving' towards actualization, but not the 'actualizing force'. Thus, unless one is prepared the rather outlandish assumption that matter strives towards all possibilities at once, the totalitarian view has not taken us one step further. I take this to be a heavy blow against the totalitarian version of decretalism.

14.2.2 Dispositionalism and theism

Because of the failure of the totalitarian version, I suggest turning again to dispositionalism to fill the 'sufficient condition gap', albeit without abandoning the decretalist framework. I call the account I am offering *dispositional divine decretalism*. On dispositional divine decretalism, God endows natural substances with dispositions/powers on which the laws of nature supervene; but He also decrees limits to the manifestations of those powers in the form of meta-laws. The dispositionalist part of the theory gives a straightforward answer to what the sufficient condition is, namely the dispositions that physical objects possess. In other words, on a dispositionalist framework the role of 'engine' would be filled by dispositions. Divine decrees would then be constraints on the actualizations of possibilities by the dispositions. However, the metaphysical problems discussed in section 13.2.3 are inherited also by a divine decretalist version of dispositionalism and therefore require a solution, one which I believe theism can provide. I wish to focus on the three challenges posed at the end of section 13.2.3: the problem of how to establish an ontology that combines dispositional with categorical aspects of properties, the question of how to avoid Meinongianism, and finally the worry that dispositionalism cannot account for central concepts of modern physics.

We have already seen that dispositional monism creates severe difficulties especially for a dispositionalist account of the laws of nature, and this is why a reintroduction of some categorical aspect to properties is of interest for the dispositionalist. Since this is not a monograph about the metaphysics of properties, I can but gesture towards a solution that involves theism. The most promising route to take seems to me to assume an identity to the effect that "the qualitative and dispositional are identical with one another and with the unitary intrinsic property itself" (Martin 2008, 65). However, as already mentioned, there is the intuition that the categorical and dispositional natures of a property are either distinct from one another (such that a property would mereologically consist of those natures), or even incompatible with each other (such that a property could either be purely categorical or purely dispositional). The logic behind this seems to be that if something is individuated by its causal powers, then it can no longer be individuated by its quiddity, and vice versa. But how about individuating properties by their unitary natures which express themselves causally in certain circumstances but could still be picked out as *those* properties even if their causal powers never come to fruition? The question that arises here is who could ascertain the identity of a 'causally dormant' property, a question somewhat parallel to the question of the persistence of objects in idealism. The answer I suggest is the same given in Berkeleyan idealism: God. Since God is omniscient, he knows even the identities of qualitative-dispositional properties inaccessible to beings who depend on a causal manifestation of those properties to identify them.

Let's turn to the challenge of Meinongianism. I already sketched two possible answers in section 13.2.3: either stick to the counterfactual analysis of dispositions and consider possible states couched in the consequent of the counterfactual as existing or abandon the counterfactual analysis and view dispositions as (existing) directednesses. The question is whether theism can buttress at least one of those solutions. Both employ an ontology which contains possibilities (or something that points to them) as existing. Those possibilities may be considered a surplus to the ontology. In fact, one common objection to including possibilities in one's ontology is that a world containing them (as unactualized) is indistinguishable from a world where they don't exist. If one additionally embraces the principle of the identity of indistinguishables, one gets the result that those two worlds are really one and the same. Here is where God comes in: the indistinguishability applies only from the human perspective; God knows full well if there are any unactualized but existing states, and so even if one adopts the identity of indistinguishables, the distinguishability of two such worlds is secured. Thus, theism would support the insinuated ontology.

Now, finally, to the worry that dispositionalism is unable to integrate central concepts of physics like the conservation principles or the principle of least action. As indicated in section 13.2.3, Alan Chalmers, who himself defends a powers ontology, advocates a pluralist solution to the question of the nature of fundamental laws of nature[117]:

> I do not see why all the generalisations fundamental to science should be of the same kind. I do not see why we should hold an essentialist view of laws. (1999, 15)

I propose a theistic version of the pluralist solution: laws of motion (dynamical equations) or laws from which laws of motion can be derived have a dispositionalist underpinning, while the constraining principles (meta-laws) applied to derive dynamical equations from laws (such as the Least Action Principle, or Noether's theorem) are conceived of as divine constraints. For this to work one must assume that the dispositions could realize more possibilities *sans* divine constraints than *cum* divine constraints. For example, a system could spontaneously gain or lose energy *even without interaction*; what keeps it from doing so is the divine decree reflected in the Noether theorem. This is admittedly at odds with classical contemporary dispositionalism. According to it, (barring non-physical interactions) the pattern of (dispositional) properties p_1 in the world at t_1 entails the pattern of (dispositional) properties p_2 at t_2, and so the possible 'world state' at t_2, from the perspective of t_1, is the one that is eventually being actualized. It seems at odds with dispositionalism to hold that there was an unactualized possibility p_3 that, according to divine decretalism, was not actualized because of a divine constraint. If all dispositions at a certain time were taken into account, then the next world-state must follow with necessity. However, I can *prima facie* see no compelling reason why one could not think beyond this framework and posit that there are (sometimes, or always) more possibilities than the dispositions actualize.

[117] Albeit one emphatically not along the lines of Nancy Cartwright's view (Chalmers 1999, 8-10).

One more piece of work remains to be done, namely to spell out the interplay between the divine decretalism about laws and mental interaction. It is first important to recall that the outcome of one instance of a law constraint is the initial condition for another instance (S1/IC2 in fig. 14.1). An interaction changes that physical state such that a different one (IC2*) comes about. The laws then 'process' that altered state in the usual manner.

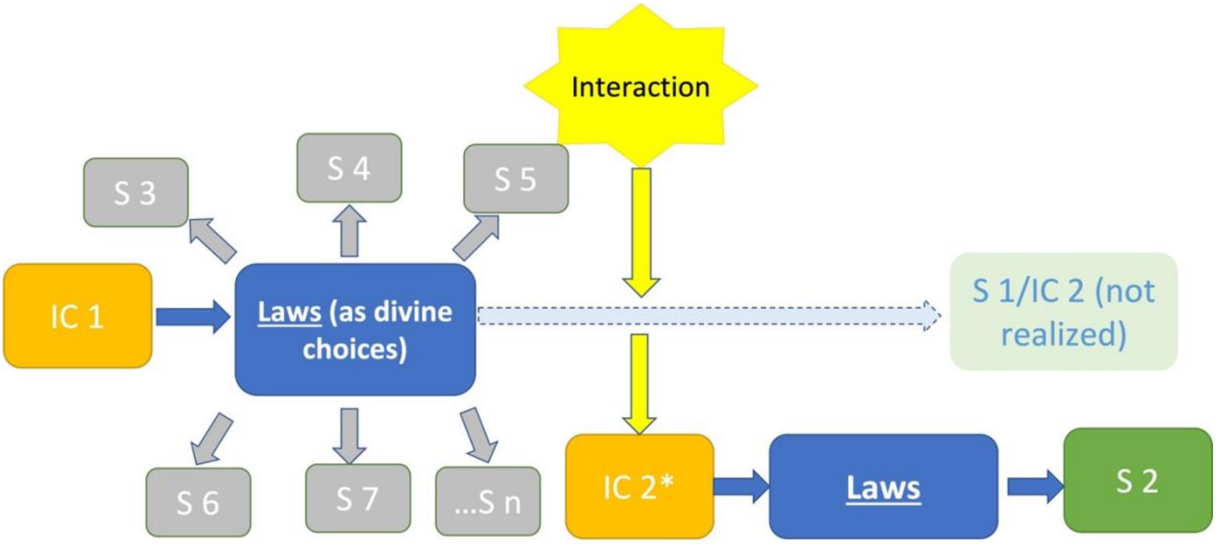

Figure 14.1: Interaction on divine decretalism. "IC" stands for "initial conditions", S for "state".

What is inevitable in this context is that a divine choice is 'thwarted' by mental interaction, at least to the degree that a certain outcome – relative to initial conditions – is constitutive of divine choices (if the degree is zero, then there is *a fortiori* no problem, because the law remains unscathed). If this is considered a bullet to be bitten, then the bullet one must bite. But as argued above, it is a distinctive advantage of a theistic account to make sense of such 'thwarted choices'. For God as the creator can (and likely does) have good reasons to allow such deviations from the regular order of things; on a naturalist atheist no such response is possible. Also, recall that the Noether theorem construes conservation as conditional, and so would the divine choice reflected by the theorem. According to this construal, the divine decrees are essentially conditional, and thus naturally make room for mental interaction. The picture just drawn of the interplay between divine meta-law choices and human interactions becomes thus a metaphysical mirror of the physical conditionality response of chapter 7.

Before turning to a theological/worldview analysis of the presented combination of dispositionalism and divine decretalism (call it 'dispositional decretalism'), let us first, in conclusion of the metaphysical appraisal, put it to the test of Jaag & Schrenk's ten desiderata.

I. Objectivity

Dispositional decretalism is objective in the sense that the laws of nature do not depend on subjects of knowledge. God's decrees are objective from our standpoint, as are dispositions.

II. Inference

The inference from laws to regularities is unproblematic whenever divine decrees are involved in the case of meta-laws, for from God's willing something to be the case follows that it be so. However, dispositionalist decretalism faces the familiar challenges of any dispositionalist metaphysics (see section 13.2.3), namely that on a purely relational/structuralist construal of dispositions, it is hard to see how to get from the fact that it is a law that a property realizes a certain manifestation upon the presence of a stimulus to the regularity that the property reacts with a manifestation in the presence of a stimulus. I argued that abandonment of dispositional monism – which works best on theism – can salvage the account in this respect.

III. Universality/invariance

On divine decretalism, laws constituted by divine decrees are universal and invariant to the extent to which God abides by them. Since God is expected to be utterly reliable and faithful, universality and invariance are secured for divine decrees. Strictly speaking, the dispositionalist decretalist benefits from this only as with respect to the decretal part of the theory. However, the underlying dispositions may also be regarded as invariant and universally valid, because they are plausibly upheld by God. This also overcomes the difficulty Lange charges dispositionalism with, namely that there is no guarantee that the dispositional essences will not change over time.

IV. Contingency

I include this here for the sake of completeness. Dispositionalist decretalism can be considered contingent, to the extent to which the laws depend on divine choices, for in principle God could alter those choices.

V. Counterfactual support

Because of the strong necessity involved due to the divine decrees, dispositionalist decretalism generally has no problem with supporting counterfactuals. However, Lange's criticism (Lange 2004, 229) that dispositionalism supports some counterfactuals trivially because it rules out certain worlds as metaphysically impossible must be answered. For example, the statement that if the speed of light in vacuum deviated only minimally from *c*, then our universe would look radically different seems to be true in a non-trivial way, but dispositionalism rules out such a world because the antecedent is never fulfilled; and this seems to be in tension with our intuition that that counterfactual must be true non-trivially. However, complemented by theism, dispositional can meet that challenge, for God could have made it the case that the dispositions – or the constraining divine decrees – are different, and so the counterfactual becomes non-trivially true.

VI. Explanatory power

That laws explain their instances seems unproblematic as far as divine decrees are concerned, for they make it the case that the instances come about, even if not productively. However, dispositionalism again comes under fire. Bird's *prima facie* reasonable tenet is that laws explain because the underlying potentialities explain; in other words, the laws are 'representatives' of the potentialities (Bird 2007, 197). But what ties a potentiality to its manifestation? Bird's answer is that it is a second-order stimulus-manifestation relation. We already discussed the regress problem this raises. However, this regress may be avoided by having God ensure the connection, and/or by abandoning the stimulus-manifestation model and adopting something like a directedness model of dispositions.

VII. Governance/independence

Divine decrees are clearly independent from the non-nomic facts they are responsible for, which is why laws, to the degree to which they are constituted by divine decrees, do govern their instances. However, dispositionalist decretalism *ex hypothesi* denies the (dynamical) laws governance and instead ascribes it to the dispositions on which the laws supervene (see Bird 2007, 82). That should not be regarded as a drawback, however, because what seems crucial is that there is *some* necessitarian government, and the dispositionalist version can tick *that* box.

VIII. Inner connection

On dispositionalism, the relation of a potentiality *P* to its manifestation property *M* belongs to the nature of *P* and thus holds with metaphysical necessity. Thus, the inner connection between instances of laws is as tight as it can be.

IX. Induction

Dispositionalist decretalism entails that, strictly speaking, the laws don't admit of induction to future regularities, because the laws supervene on regularities. However, as in the case of explanation, a rationalization of induction is available on dispositionalism, for example like this:

> All observed Fs have produced Gs. The best explanation of this is that the Fs are members of a natural kind K, whose essence is or includes the disposition to produce Gs. Hence all Fs (by virtue of membership of kind K) produce Gs. (Beebee 2011, 520)

Beebee argues against this, objecting that it is not guaranteed that natural kinds will always have the same dispositional powers, and that even if were, the presently existing kinds could be replaced by others in the future. Again, theism is of help here, grounding the existence of natural kinds ultimately in God's creative activity, and thereby guaranteeing as much constancy as can be had.

X. Significance for the sciences

We've seen that accounts which construe laws as constraints generally harmonize better with the natural sciences. Dispositionalism, as discussed, faces the charge that it cannot account for meta-laws like the principle of least action or conservation principles. This problem is at once mended by the identification of those principles with divine decrees. That said, dispositionalism also has some inherent amenability to scientific practice. Here is why: the intrinsic nature of an essential dispositional property alone determines its characteristic behavior and the corresponding law. These properties exhibit behavior 'in pure form' only in controlled, ideal experimental situations in which there are no confounding factors (see Cartwright 1999). However, the properties are *de facto* almost never present in isolation, but are instantiated in physical situations where numerous potential interferences are present. Therefore, in such situations they show their characteristic behavior at best approximately. Nevertheless, the presence of one and the same dispositional property in ideal and non-ideal circumstances explains why their behavior in idealized circumstances can be approximately extrapolated to non-ideal circumstances (and vice versa) – and that corresponds to proven scientific methodology.

14.3 A theological assessment of divine decretalism

Divine decretalism puts God at the center of its theoretical framework. A full appraisal therefore requires some theological assessment. It will turn out that the theological analysis, on balance, further discourages embracing the totalitarian metaphysics. The assessment takes place in the context of the Christian tradition, but I presume that the similarities with the other two major monotheistic religions are sufficient. My approach is as follows: I will mainly examine whether the decretalist framework is in line with biblical theology, and investigate the relative merits of the dispositional and totalitarian versions where appropriate. In the quest for a more detailed metaphysical picture, we can begin right with the very first verses of Scripture:

> In the beginning, God created the heavens and the earth. And the earth was without form and void [Hb. *tohu wa bohu*], and darkness was over the face of the deep. (Genesis 1:1-2, ESV)

After being called into existence, creation is formless and void and stands in need of ordering, which is exactly what happens afterward: the story of the six creation days is a story of *differentiation* and *delineation* (of light from darkness, day from night, seas from dry land etc.). This sounds a lot like the limiting that divine choices do on decretalism. At least conceptually, decretalism seems to be there in the Bible. However, its scope is as yet undetermined. Genesis 1 speaks of the 'one-off' act of creation. Were divine decrees uttered just as part of that original creation act, or are they in some sense continually declared as the laws of nature? Also, the question is what divine decrees apply to: an infinite space of possibilities (the totalitarian view) or a finite, small space of possibilities (the dispositionalist version)?

An answer to the latter question can be given by considering the transition between the initial *tohu wa bohu*-state and the more structured creation we see beginning with Genesis 1:3. Even if we granted that Genesis 1:2 reflects a state of as yet infinite possibilities – which is implausible, given that it already contains a differentiation between the heavens (*ha'shamayim*) and the earth (*ha'arets*) – we would still have to acknowledge that with the differentiations throughout the rest of Genesis 1 nature is brought down to a state of less than infinite possibilities. Thus, laws of nature construed as divine decrees would apply to a nature that already possesses an intrinsic structure with only a finite number of possibilities how to evolve. This reflects one important tenet of traditional Christian theology, namely that creation is distinct from God in crucial respects, for example in having a beginning (and possibly an end), and in creatures being limited as regards their nature. In a nutshell, creation's distinctive hallmark is *finitude*, both temporally and in the scope of its being. If we consider creation as depicted in Genesis 1 as a unique, one-time act, then it follows that the way things in nature are – I shall henceforth use the term 'essences' of things – are something *intrinsic* to the creatures, not (constantly) imposed on them by God.

This said, there is in Christian theology also the concept of *creatio continua*, meaning God's continual involvement in (among others) upholding creation. For example, the New Testament contends that Christ "upholds the universe by the word of his power" (Hebrews 1:3, ESV). However, it is not immediately clear what the upholding consists in. It could in principle be that God continually utters decrees that limit the scope of being of creation; that is encouraged by the use of "word" (*logos*) in the text. Still, God's upholding activity consisting in speech acts is compatible with it preserving dispositional essences, or making choices that constrain an excess of possibilities, or both. Since we've already amassed a number of arguments against the totalitarian view, I urge that the most adequate reading is a dispositional decretalist one (corresponding to the last option just listed).

At any rate, the decretalist picture gets more support from Scripture than the verses just cited. For example, God Himself asks Job:

> Or who shut in the sea with doors when it burst out from the womb, when I made clouds its garment and thick darkness its swaddling band, and prescribed limits for it and set bars and doors, and said, 'Thus far shall you come, and no farther, and here shall your proud waves be stayed'? Have you

> commanded the morning since your days began, and caused the dawn to know its place (…)? (Job 38:8-12, ESV)

Of course, and this holds especially for the Old Testament due to its Hebraistic nature (the New Testament, written in Greek and within a broadly Hellenistic culture is much more 'lucid'), one must be careful to draw any metaphysical conclusions from such poetic and florid texts; but this does not mean that one cannot draw *any* conclusions from them. Another hermeneutical issue: one could say (against decretalism) that God's restricting the sea refers only to particular instances, for example to deliver swathes of land or other creatures from the sea's destructive power. However, the greater context of the Job passage is not about a particular instance, but about archetypal creation. With this in mind, the language of "shutting in" the sea, or of "prescribing limits" for it, does suggest something along the lines of a decretalist metaphysics[118]; at least the emphasis on this kind of restrictive action is clear (repeating it three times, one of which is a direct quote of God 'rebuking' the sea). Sure, we are dealing with a metaphor here, but every metaphor is a metaphor *of something*. We could also say that *somehow* God is limiting the being of nature. But *how* exactly? It seems to overstretch textual interpretation to try to draw out such a level of metaphysical detail. If anything, the passage supports the essentialist view: for what is "shut in" is not some indefinite prime matter, but the "sea".

Taking this as a provisional point of departure, it is easy to see and relatively uncontroversial that what God does is to order creation by drawing boundaries. Support again comes from the book of Job, which, even if not metaphysically, deals with many fundamental issues in connection with creation:

> He has *inscribed* a circle on the face of the waters at the *boundary* between light and darkness. (Job 26:10, ESV; italics mine)

What is further relevant in this context is the fact that (to my knowledge) there is no passage in the Bible which renders God as being the active *moving force* of creation. His activity, as pointed out above, is rather depicted in terms of *upholding* and *restricting*. That may be taken as further support for the view that creatures possess dispositional essences: it is their intrinsic powers that make them move, not God in the sense of an efficient cause.

For what it is worth, a totalitarian metaphysics also has unpalatable implications that run counter to deeply entrenched intuitions concerning nature. It seems to many that especially living things have stable natures that seem to be *intrinsic* to them. Those with said intuition are in good company: it was his observations about living things that strongly motivated Aristotle to develop his hylomorphic theory of substance. If, however, the behavior of atoms, say, in a lion's body is due exclusively to divine constraints imposed on otherwise chaotic matter, then the stability of, say 'lionhood' consists solely in those *extrinsic* choices. 'Lionhood' would then be something that exists only in the mind of God, and it would also take some theoretical acrobatics to explain how we get to have the impression that there is something like 'lionhood'.

One final but important criterion is the respective continuity of the two accounts with the metaphysics of personal (agent) causation. I argued section 13.2.3 that dispositionalism offers a framework within which a powers account of personal causation, along the lines of Richard Swinburne's SPL theory, can be

---

[118] Jeffrey Koperski pointed out to me (personal communication) that he often received very positive reactions from Old Testament scholars to this alignment of decretalist metaphysics and OT language.

seamlessly embedded. Such an account is better suited to capture what is going on when persons cause physical events. If one adopts the totalitarian view, by contrast, one is faced with a choice: either modify the metaphysical mechanism mental causation such that it also becomes a reduction of possibilities, or give up the quest for a unified causational picture. To my mind, the overall dispositional picture is preferable; but nothing hinges on that preference, because we have already encountered independent weighty arguments against the totalitarian metaphysics.

I conclude that a dispositional essentialist divine decretalism is the preferable position. It offers a robust metaphysics of laws that also incorporates those meta-principles a secular dispositionalism has a hard time integrating. Moreover, it satisfies our deep-seated intuitions about the essences of things and causality in nature. Its main challenges are on the one hand to spell out an ontology of properties that satisfactorily combines the advantages of categoricalism and dispositionalism, something I argued looks promising on theism. On the other hand, it would be desirable to know in more detail what work divine constraints do on the intrinsically dispositional nature, a gap I did not attempt to fill this gap.

The totalitarian version of divine decretalism, by contrast, has some significant drawbacks. Metaphysically, it fails to provide a sufficient condition for the motion of matter. Theologically, it is slightly in tension with God's nature as absolute Being, but more importantly it gets too little support from Holy Scripture; at least all pertinent instances can be better made sense of with the dispositionalist version. It also runs counter to our deep-seated intuitions about living things and nature in general.

Summary and conclusion

Interactive dualism is one straightforward way to account for our intuitive understanding of ourselves as free agents. That seeming, I argued in chapter 1, is robust and should only be questioned in the presence of sufficient evidence to the contrary.

In chapters 2 to 9 I dealt with such putative counterarguments. The only purely philosophical one, the causal heterogeneity objection, relies on too narrow a conception of causality which to adopt there is no cogent reason; to the contrary, we have strong reasons to believe in a larger conception of causality, not least because we know causality first-hand from ourselves. The objection from causal closure, if construed as something more sophisticated than a metaphysical doctrine, has two significant problems: first, the notion of the physical makes trouble, being either too vague or collapsing into either an uncontroversial causal thesis or again some unwarranted dogma. Second, the reliance of its inductive thesis on conservation principles backfires, since those principles are conditional in nature and make room for non-physical interaction.

It is those conservation principles I dwelt at length in chapters 4 to 7. I showed how the widespread objection from energy conservation is often muddle-headed and uses false conceptions (chapter 4), which replies should be avoided (chapter 5), why quantum mechanics is of no help (chapter 6) and why the conditionality response is the most solid rejoinder dualists can give. Chapter 8 dealt with the question whether quantum mechanics can at least facilitate mental interaction, if not in a conservation-friendly manner; my conclusion was in the negative.

Chapter 9 is the attempt of a 'counterattack': physicalists say there is plenty of empirical evidence for mental causation having a complete physical explanation and no evidence for anomalies putatively caused by non-physical minds. Well, a look at what is available as empirical evidence shows that the question is never really addressed, but an in-principle analysis of the known mechanisms of neuronal activation suggest rather that a full explanation of volitional actions isn't even *possible* without recourse to something non-physical.

With all this said, there was still the legitimate desire to dig deeper. Surely, nature is somehow governed by laws of nature; how does a defense of interactive dualism along the lines of the conditionality response tie in with that? In chapter 10 I argued that what physicalists should be wielding against interactive dualism (if anything) are the laws of nature. Chapter 11 then served the purpose of at least partly putting away historical myopia regarding the concept of laws of nature. Its most important lesson is that the notion of laws with respect to nature has distinctively theistic origins but then was 'hijacked' by naturalists, thereby leaving us with a weirdly sacrosanct, rigid conception of 'laws without a lawmaker' (something that partly explains people's reluctance to accept interaction even if it is philosophically rehabilitated). Chapter 13 does the hard work of exploring which metaphysics of laws is best suited to account both for, broadly speaking, the regularity and stability of the world as well as the possibility of interaction. One important conclusion I drew is that, independent of the exact metaphysics, there is no need to consider laws broken by interaction. Regarding the different contemporary theories, an application of Jaag & Schrenk's ten-point-catalogue of desiderata yielded that dispositionalism (with some modifications) and Chen & Goldstein's minimal primitivism (MinP) stand out, albeit for very different reasons. Dispositionalism can explain why the laws are what they are, grounding them in the dispositional essences of things; MinP is best in accord with the way laws are used in physics, namely, roughly speaking as constraints on the histories of the world, without the invocation of either causality or an arrow of time. However, all extant accounts, at some point or another suffer from problems beyond the reach of their metaphysical resources.

The main drive behind chapter 13 is to go to theism for help. I argued that theism is able to satisfy our most basic desiderata about the stability/regularity of nature *and* the possibility of human interaction. Not all theories square with theism equally well, though; indeed, among the secular theories, dispositionalism and MinP again looked the most compatible. I also discussed Jeffrey Koperski's divine decretalism as a distinctively theistic approach (most similar to MinP), finding it useful in many ways, but leaving some questions unanswered. In section 14.3, I therefore suggested a synthesis of divine decretalism with either dispositionalism or what I call the 'totalitarian view' of possibilities. On balance, I found dispositionalism better suited to meet the demands of metaphysical theorizing as well as large-scale worldview and theological questions. Not least, it provides a unified picture of causation: for both in the case of natural objects and in the case of persons causation consists in the actualization of potentialities, conceived of as directednesses.

References

Adlam, Emily. 2022. "Laws of Nature as Constraints." *Foundations of Physics* 52 (1): 1–41.

Akkal, D., R. P. Dum, and P. L. Strick. 2007. "Supplementary Motor Area and Presupplementary Motor Area: Targets of Basal Ganglia and Cerebellar Output." *Journal of Neuroscience* 27: 10659–73.

Albert, David, and Barry Loewer. 1988. "Interpreting the Many Worlds Interpretation." *Synthese* 77 (2): 195–213. https://doi.org/10.1007/BF00869434.

Albert, David Z. 1992. *Quantum Mechanics and Experience*. First Printing. Cambridge, Mass: Harvard University Press.

Al-Khalili, Jim, and Johnjoe McFadden. 2015. *Life on the Edge: The Coming of Age of Quantum Biology*. 1st ed. Black Swan.

Alston, William. 1994. "Divine Action: Shadow or Substance?" In *The God Who Acts: Philosophical and Theological Explorations*, 41–62. University Park: Pennsylvania State University Press.

Anscombe, Elizabeth. 2018. "Causality and Determination." In *Agency and Responsibility*, 57–73. Routledge.

Anstey, Peter R. 2002. *The Philosophy of Robert Boyle*. Routledge.

Armstrong, David Malet. 2016. *What Is a Law of Nature?* Cambridge University Press.

Askenasy, Jean, and Joseph Lehmann. 2013. "Consciousness, Brain, Neuroplasticity." *Frontiers in Psychology* 4: 412.

Aspect, Alain, Jean Dalibard, and Gérard Roger. 1982. "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers." *Physical Review Letters* 49 (25): 1804.

Aspect, Alain, Philippe Grangier, and Gérard Roger. 1981. "Experimental Tests of Realistic Local Theories via Bell's Theorem." *Physical Review Letters* 47 (7): 460.

———. 1982. "Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities." *Physical Review Letters* 49 (2): 91.

Averill, Edward, and B.F. Keating. 1981. "Does Interactionism Violate a Law of Classical Physics?" *Mind*, no. XC: 102–7.

Baker, Lynne Rudder. 1993. "Metaphysics and Mental Causation." *Mental Causation* 75: 95.

Barrow, Isaac. 1885. "Maker of Heaven and Earth (Sermon XII)." In *Theological Works*. Vol. 2. Cambridge: Cambridge University Press.

Bassi, Angelo, Kinjalk Lochan, Seema Satin, Tejinder P. Singh, and Hendrik Ulbricht. 2013. "Models of Wave-Function Collapse, Underlying Theories, and Experimental Tests." *Reviews of Modern Physics* 85 (2): 471.

Battaglia, Demian, and David Hansel. 2011. "Synchronous Chaos and Broad Band Gamma Rhythm in a Minimal Multi-Layer Model of Primary Visual Cortex." *PLoS Computational Biology* 7 (10).

Beauregard, Mario, Johanne Lévesque, and Pierre Bourgouin. 2001. "Neural Correlates of Conscious Self-Regulation of Emotion." *Journal of Neuroscience* 21 (18): RC165–RC165.

Beck, Friedrich. 1996. "Can Quantum Processes Control Synaptic Emission?" *International Journal of Neural Systems* 07 (04): 343–53. https://doi.org/10.1142/S0129065796000300.

———. 2001. "Quantum Brain Dynamics and Consciousness." In *The Physical Nature of Consciousness*, edited by Philip Van Loocke, 83–116. John Benjamins Publishing.

———. 2008. “Synaptic Quantum Tunnelling in Brain Activity.” *NeuroQuantology* 6 (2).

Beck, Friedrich, and John C. Eccles. 1992. “Quantum Aspects of Brain Activity and the Role of Consciousness.” *Proceedings of the National Academy of Science USA* 89: 11357–61.

Beebee, Helen. 2011. “Necessary Connections and the Problem of Induction.” *Noûs* 45 (3): 504–27.

Bell, J.S. 1964. “On the Einstein-Podolsky-Rosen Paradox.” *Physics*, no. 1: 195–200.

———. 1966. “On the Problem of Hidden Variables in Quantum Mechanics.” *Reviews of Modern Physics*, no. 38: 447–52.

Bird, Alexander. 2007. *Nature's Metaphysics: Laws and Properties*. OUP Oxford.

Bishop, Robert C. 2008. “What Could Be Worse than the Butterfly Effect?” *Canadian Journal of Philosophy* 38 (4): 519–47.

———. 2011. “Free Will and the Causal Closure of Physics.” In *Visions of Discovery : New Light on Physics, Cosmology, and Consciousness*, edited by Raymond Chiao, 601–11. New York: Cambridge University Press.

Bloom, Paul. 2004. *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. Cambridge, MA: Basic Books.

Bourget, David. 2004. “Quantum Leaps in Philosophy of Mind: A Critique of Stapp's Theory.” *Journal of Consciousness Studies* 11 (12): 17–42.

Boyle, Robert. 1725. *The Philosophical Works of the Honourable Robert Boyle Esq: Abridged, Methodized, and Disposed Under the General Heads of Physics, Stat- Ics, Pneumatics, Natural History, Chymistry, and Medicine*. Edited by Pater Shaw. Vol. 2. London: W. and J. Innys.

Bricmont, Jean. 2016a. *Making Sense of Quantum Mechanics*. 1st ed. 2016. New York, NY: Springer.

Brinkman, C., and R. Porter. 1979. “Supplementary Motor Area in the Monkey: Activity of Neurons during Performance of a Learned Motor Task.” *Journal of Neurophysiology* 42 (3): 681–709. https://doi.org/10.1152/jn.1979.42.3.681.

Broad, C.D. 1937. *The Mind and Its Place in Nature*. London: Kegan Paul, Trench, Trubner and Co.

Brooke, John H. 1992. “Natural Law in the Natural Sciences: The Origins of Modern Atheism?” *Science and Christian Belief* 4 (2): 83–103.

Brooke, John Hedley. 1991. *Science and Religion: Some Historical Perspectives*. Cambridge University Press.

Brüntrup, Godehard. 2014. “Quantum Mechanics and Intentionality.” *Quantum Physics Meets the Philosophy of Mind: New Essays on the Mind-Body Relation in Quantum-Theoretical Perspective. Ed. Antonella Corradini and Uwe Meixner. Berlin/Boston: De Gruyter*, 35–49.

Bunge, Mario Augusto. 1980. *The Mind-Body Problem: A Psychobiological Approach*. Pergamon Press.

Carroll, Sean M., and Jackie Lodman. 2021. “Energy Non-Conservation in Quantum Mechanics.” *Foundations of Physics* 51 (4): 1–15.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

———. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.

Chakravartty, Anjan. 2017. “Scientific Realism.” In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/sum2017/entries/scientific-realism.

Chalmers, Alan. 1999. “Making Sense of Laws of Physics.” In *Causation and Laws of Nature*, edited by H. Sankey, 3–16. Dordrecht: Kluwer.

Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.

———. 2000. "What Is a Neural Correlate of Consciousness." In *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, edited by Thomas Metzinger, 17–39. MIT Press.

Chalmers, David, and Kelvin McQueen. forthcoming. "Consciousness and the Collapse of the Wave Function." In *Quantum Mechanics and Consciousness*, edited by S. Gao.

Chen, Eddy Keming, and Sheldon Goldstein. 2022. "Governing without a Fundamental Direction of Time: Minimal Primitivism about Laws of Nature." In *Rethinking the Concept of Law of Nature*, 21–64. Springer.

Chisholm, Roderick. 1976. "The Agent as Cause." In *Action Theory*, edited by M. Brand and D. Walton, 199–211. Dordrecht: Reidel.

———. 2014. *Person and Object: A Metaphysical Study*. Routledge.

Choi, Sungho, and Michael Fara. 2021. "Dispositions." In *Stanford Encyclopedia of Philosophy*. Stanford: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/dispositions/.

Clarke, Peter G. H. 2014. "Neuroscience, Quantum Indeterminism and the Cartesian Soul." *Brain and Cognition* 84 (1): 109–17. https://doi.org/10.1016/j.bandc.2013.11.008.

Clarke, Randolph. 2000. "Modest Libertarianism." *Philosophical Perspectives* 14: 21–45.

———. 2011. "Alternatives for Libertarians." In *The Oxford Handbook of Free Will*, edited by Robert Kane, 2nd ed. Oxford University Press.

Coleman, William, and William R. Coleman. 1977. *Biology in the Nineteenth Century: Problems of Form, Function and Transformation*. Vol. 1. Cambridge University Press.

Collins, Robin. 2008. "Modern Physics And The Energy-Conservation Objection To Mind-Body Dualism." *American Philosophical Quarterly* 45 (1): 31–42.

———. 2011a. "A Scientific Case for the Soul." In *The Soul Hypothesis*, edited by Marc Baker and Stewart Goetz, 222–46. continuum.

———. 2011b. "The Energy of the Soul." In *The Soul Hypothesis*, edited by Marc Baker and Stewart Goetz, 123–33. continuum.

Crane, Tim. 1991. "Why Indeed? Papineau on Supervenience." *Analysis* 51 (1): 32–37.

Crick, Francis, and Christof Koch. 1990. "Towards a Neurobiological Theory of Consciousness." In *Seminars in the Neurosciences*, 2:263–75. Saunders Scientific Publications.

Cucu, Alin C., and J. Brian Pitts. 2019. "How Dualists Should (Not) Respond to the Objection from Energy Conservation." *Mind and Matter* 17 (1): 95–121.

Davydov, Alexander S. 1973. "The Theory of Contraction of Proteins under Their Excitation." *Journal of Theoretical Biology* 38 (3): 559–69.

———. 1977. "Solitons and Energy Transfer along Protein Molecules." *Journal of Theoretical Biology* 66 (2): 379–87.

Deckert, Dirk-André, Michael Esfeld, and Andrea Oldofredi. 2020. "A Persistent Particle Ontology for Quantum Field Theory in Terms of the Dirac Sea." *The British Journal for the Philosophy of Science*.

Deecke, L., and H. H. Kornhuber. 1978. "An Electrical Sign of Participation of the Mesial 'supplementary' Motor Cortex in Human Voluntary Finger Movement." *Brain Research* 159 (2): 473–76. https://doi.org/10.1016/0006-8993(78)90561-9.

Dennett, Daniel C. 1991. *Consciousness Explained*. Penguin Books.

Descartes, René. 1841. "Discours de La Méthode, Cinquième Partie." In *Œuvres de Descartes*, edited by Jules Simon. Paris.

Desmond, A.J. 1997. *Huxley: From Devil's Disciple to Evolution's High Priest*. Reading: Addison-Wesley.

Dilley, F.B. 2004. "Taking Consciousness Seriously: A Defense of Cartesian Dualism." *International Journal for Philosophy of Religion* 55: 135–53.

Dolbeault, Joël. 2017. "Laws of Nature or Panpsychism?" *Journal of Consciousness Studies* 24 (1–2): 87–110.

Ducasse, C. 1960. "In Defense of Dualism." In *Dimensions of Mind: A Symposium*, edited by S. Hook, 85–90. New York: New York University Press.

Earman, John, John T. Roberts, and Sheldon Smith. 2002. "Ceteris Paribus Post." *Erkenntnis* 57 (3): 281–301.

Eccles, John C. 1994. *How the Self Controls Its Brain*. Springer.

Esfeld, Michael. 2019. "Why Determinism in Physics Has No Implications for Free Will." *Organon F* 26 (1): 62–85.

———. 2020. *Science and Human Freedom*. Springer.

———. 2022. "Super-Humeanism and Mental Causation." In *Powers, Time and Free Will*, 257–72. Springer.

Everett, Hugh. 1957. "'Relative State' Formulation of Quantum Mechanics." *Reviews of Modern Physics* 29 (3): 454–62. https://doi.org/10.1103/RevModPhys.29.454.

Fair, D. 1979. "Causation and the Flow of Energy." *Erkenntnis*, no. 14: 219–50.

Fales, Evan. 2010. *Divine Intervention: Metaphysical and Epistemological Puzzles*. Routledge.

Feldmann, William, and Roderich Tumulka. 2012. "Parameter Diagrams of the GRW and CSL Theories of Wavefunction Collapse." *Journal of Physics A: Mathematical and Theoretical* 45 (6): 065304.

Fine, Kit. 2002. "Varieties of Necessity." In *Conceivability and Possibility*, edited by Tamar Szábo Gendler and John Hawthorne, 253–81. Oxford: Oxford University Press.

Fisher, P. 2014. "Pierre Gassendi." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Stanford University. http://plato.stanford.edu/archives/spr2014/entries/gassendi/.

Flanagan, Owen J. 1991. *The Science of the Mind*. MIT Press.

Fodor, Jerry A. 1983. "Making Mind Matter More." In *Mental Causation*, edited by John Heil and Alfred R. Mele. Oxford: Clarendon Press.

Foster, John. 1991. *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. London ; New York: Routledge.

———. 2001. "VII—Regularities, Laws of Nature, and the Existence of God." In *Proceedings of the Aristotelian Society*, 101:145–61. Oxford University Press Oxford, UK.

Frankfurt, Harry G. 1969. "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66 (23): 829–39.

Fried, Itzhak, Roy Mukamel, and Gabriel Kreiman. 2011. "Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition." *Neuron* 69: 548–62.

Gao, Shan. 2017. *The Meaning of the Wave Function*. Cambridge University Press.

———, ed. forthcoming. *Quantum Mechanics and Consciousness*. Oxford University Press.

Gasser, Georg. 2015. "Eine Analyse Der Ontologischen Voraussetzungen Der Agenskausalität." Ph.D. dissertation, Innsbruck: University of Innsbruck.

Gehlert, Thomas. 2019. "System-Aufstellungen Als Instrument Zur Unternehmensführung Im Rahmen Komplexer Entscheidungsprozesse Sowie Ein Naturwissenschaftlich Begründetes Erklärungsmodell Für Den Dahinterliegenden Prozess." Chemnitz: TU Chemnitz.

Gell-Mann, Murray. 1956. "The Interpretation of the New Particles as Displaced Charge Multiplets." *Il Nuovo Cimento (1955-1965)* 4 (2): 848–66.

Georgiev, Danko, and James F. Glazebrook. 2014. "Quantum Interactive Dualism: From Beck and Eccles Tunneling Model of Exocytosis to Molecular Biology of SNARE Zipping." *Biomedical Reviews* 25 (0): 15–24. https://doi.org/10.14748/bmr.v25.1038.

Gibb, Sophie. 2010. "Closure Principles and the Laws of Conservation of Energy and Momentum." *Dialectica* 64: 363–84.

Ginet, Carl. 2007. "An Action Can Be Both Uncaused and Up to the Agent." In *Intentionality, Deliberation, and Autonomy*, edited by Lumer, 243–55. Ashgate.

Goetz, Stewart. 1997. "Libertarian Choice." *Faith and Philosophy* 14 (2): 195–211.

Goetz, Stewart, and Charles Taliaferro. 2011. *A Brief History of the Soul*. 1st ed. Malden, MA: Wiley-Blackwell.

Goldstein, Herbert. 1980. *Classical Mechanics*. 2nd edition. Addison-Wesley.

Haggard, Patrick. 2008. "Human Volition: Towards a Neuroscience of Will." *Nature Reviews Neuroscience* 9 (12): 934–46. https://doi.org/10.1038/nrn2497.

Hall, Ned. 2009. "Humean Reductionism about Laws of Nature." http://philpapers.org/rec/HALHRA.

Halvorson, Hans. 2011. "The Measure of All Things: Quantum Mechanics and the Soul." In *The Soul Hypothesis*, edited by Marc Baker and Stewart Goetz. continuum.

Hameroff, Stuart, and Roger Penrose. 2014. "Consciousness in the Universe: A Review of the 'Orch OR' Theory." *Physics of Life Reviews* 11 (1): 39–78. https://doi.org/10.1016/j.plrev.2013.08.002.

Hammond, Constance. 2015. *Cellular and Molecular Neurophysiology*. 4th edition. Amsterdam ; Boston: Academic Press.

Harrison, Peter. 2013. "Laws of Nature in Seventeenth-Century England." *The Divine Order, the Human Order, and the Order of Nature: Historical Perspectives. Oxford Univeristy Press, New York*, 127–48.

Hart, W.D. 1994. "Dualism." In *A Companion to the Philosophy of Mind*, edited by S. Guttenplan, Blackwell Companions to Philosophy, 265–69. Oxford: Blackwell.

Hasker, William. 1999. *The Emergent Self*. Ithaca, N.Y: Cornell University Press.

Heil, John, and John Fergusson Heil. 2003. *From an Ontological Point of View*. Oxford University Press on Demand.

Henry, John. 2004. "Metaphysics and the Origins of Modern Science: Descartes and the Importance of Laws of Nature." *Early Science and Medicine* 9 (2): 73–114.

———. 2009. "Voluntarist Theology at the Origins of Modern Science: A Response to Peter Harrison." *History of Science* 47 (1): 79–113.

Herschel, William. 1851. *A Preliminary Discourse on the Study of Natural Philosophy*. New. London: Longman, Brown, Green & Longmans.

———. 2013. *The Scientific Papers of Sir William Herschel*. Cambridge University Press.

Hille, Bertil. 2001. *Ion Channels of Excitable Membranes*. 3rd ed. Sunderland, MA: Sinauer Associates.

Hirsh, Ira J., and Carl E. Sherrick Jr. 1961. "Perceived Order in Different Sense Modalities." *Journal of Experimental Psychology* 62 (5): 423.

Hobbs, Jesse. 1991. "Chaos and Indeterminism." *Canadian Journal of Philosophy* 21 (2): 141–64.

Hong, Felix T. 2003. "Towards Physical Dynamic Tolerance: An Approach to Resolve the Conflict between Free Will and Physical Determinism." *Biosystems* 68 (2–3): 85–105.

Hořava, Petr. 2009. "Quantum Gravity at a Lifshitz Point." *Physical Review D* 79 (8): 084008. https://doi.org/10.1103/PhysRevD.79.084008.

Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham, Md: Rowman & Littlefield Publishers.

Hurley, Patrick J., and Lori Watson. 2018. *A Concise Introduction to Logic*. 13th ed. Cengage Learning.

Huxley, Thomas Henry. 1874. *Address Delivered before the British Association Assembled at Belfast*. New York: D. Appleton.

———. 1875. *History of the Conflict Between Religion and Science*. The International Scientific Series, XII. D. Appleton.

———. 1890. *Introductory [Science Primer]*. American Book Company.

———. 1902. "Science and Pseudo-Science." In *Collected Essays, Volume V: Science and Christian Tradition*. New York: D. Appleton.

Ismael, Jenann. 2016. *How Physics Makes Us Free*. Oxford University Press.

Jacobs, Jonathan D. 2011. "Powerful Qualities, Not Pure Powers." *The Monist* 94 (1): 81–102.

Jahanshahi, Marjan, I. Harri Jenkins, Richard G. Brown, C. David Marsden, Richard E. Passingham, and David J. Brooks. 1995. "Self-Initiated versus Externally Triggered MovementsI. An Investigation Using Measurement of Regional Cerebral Blood Flow with PET and Movement-Related Potentials in Normal and Parkinson's Disease Subjects." *Brain* 118 (4): 913–33. https://doi.org/10.1093/brain/118.4.913.

Jahn, Reinhard, and Dirk Fasshauer. 2012. "Molecular Machines Governing Exocytosis of Synaptic Vesicles." *Nature* 490 (7419): 201–7.

Jedlicka, Peter. 2017. "Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)Biology?" *Frontiers in Molecular Neuroscience* 10 (November).

Kajamies, Timo. 2009. "De Novo Creat : Descartes on Action, Interaction, and Continuous Creation." In *The World as Active Power: Studies in the History of European Reason*, edited by Juhani Pietarinen and Valtteri Viljanen. Brill.

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

———. 2011. "Rethinking Free Will: New Perspectives On An Ancient Problem." In *The Oxford Handbook Of Free Will*, edited by Robert Kane, 2nd ed. Oxford University Press.

Katzav, Joel. 2004. "Dispositions and the Principle of Least Action." *Analysis* 64 (3): 206–14.

Kedar, Yael, and Giora Hon. 2017. "'Natures' and 'Laws': The Making of the Concept of Law of

Nature–Robert Grosseteste (c. 1168–1253) and Roger Bacon (1214/1220–1292).” *Studies in History and Philosophy of Science Part A* 61: 21–31.

King, Chris C. 1991. “Fractal and Chaotic Dynamics in Nervous Systems.” *Progress in Neurobiology* 36 (4): 279–308.

Koons, Robert C., and Timothy Pickavance. 2015. *Metaphysics: The Fundamentals*. John Wiley & Sons.

Koperski, Jeffrey. 2000. “God, Chaos, and the Quantum Dice.” *Zygon®* 35 (3): 545–59.

———. 2017. “Breaking Laws of Nature.” *Philosophia Christi* 19 (1): 83–101.

———. 2020. *Divine Action, Determinism and the Laws of Nature*. Taylor & Francis.

———. 2021. “How the Laws of Nature Were Naturalised.” *Science & Christian Belief* 33 (2).

Korn, Henri, and Philippe Faure. 2003. “Is There Chaos in the Brain? II. Experimental Evidence and Related Models.” *Comptes Rendus Biologies* 326 (9): 787–840.

Kristofferson, Alfred B. 1967. “Successiveness Discrimination as a Two-State, Quantal Process.” *Science* 158 (3806): 1337–39.

Ladyman, James, Don Ross, John Collier, Don Spurrett, David Spurrett, and John Gordon Collier. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press on Demand.

Landry, Elaine, Dean Rickles, Michael Esfeld, Dirk-André Deckert, Dustin Lazarovici, Andrea Oldofredi, and Antonio Vassallo. 2017. *A Minimalist Ontology of the Natural World*. Routledge.

Lange, Marc. 2004. “A Note on Scientific Essentialism, Laws of Nature, and Counterfactual Conditionals.” *Australasian Journal of Philosophy* 82 (2): 227–41.

———. 2005. “Laws and Their Stability.” *Synthese* 144 (3): 415–32.

———. 2009. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford University Press.

Larmer, Robert. 1986. “Mind-Body Interaction and the Conservation of Energy.” International Philosophical Quarterly. August 1, 1986. https://doi.org/10.5840/ipq198626316.

———. 2008. “Miracles, Physicalism, and the Laws of Nature.” *Religious Studies* 44 (2): 149–59.

———. 2014. “Divine Intervention and the Conservation of Energy: A Reply to Evan Fales.” *International Journal for Philosophy of Religion* 75 (1): 27–38.

———. 2017. “Decretalism and the Laws of Nature: A Response to Jeffrey Koperski.” *Philosophia Christi* 19 (2): 439–47.

Lazarovici, Dustin. 2018. “Against Fields.” *European Journal for Philosophy of Science* 8 (2): 145–70.

Leftow, Brian. 1998. “Voluntarism.” In *Routledge Encyclopedia of Philosophy*. https://www.rep.routledge.com/articles/thematic/voluntarism/v-1.

Leibniz, G.W. 1973. *Leibniz: Discourse on Metaphysics/Correspondence with Arnauld/Monadology*. Translated by G.R. Montgomery. La Salle, Illinois: Open Court.

Lévesque, Johanne, Fanny Eugene, Yves Joanette, Vincent Paquette, Boualem Mensour, Gilles Beaudoin, Jean-Maxime Leroux, Pierre Bourgouin, and Mario Beauregard. 2003. “Neural Circuitry Underlying Voluntary Suppression of Sadness.” *Biological Psychiatry* 53 (6): 502–10.

Lewis, C. S. 1960. *Miracles, Revised Edition*. New York: Macmillan.

———. 2000. *Essay Collection & Other Short Pieces*. Kindle.

Lewis, David. 1973. *Counterfactuals*. Oxford: Blackwell.

———. 1986. *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.

———. 1994. "Humean Supervenience Debugged." *Mind* 103 (412): 473–90.

Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8 (4): 529–39. https://doi.org/10.1017/S0140525X00044903.

———. 1994. "A Testable Field Theory of Mind-Brain Interaction." *Journal of Consciousness Studies* 1 (1): 119–26.

———. 1999. "Do We Have Free Will?" *Journal of Consciousness Studies* 6 (8–9): 47–57.

Libet, Benjamin, C. A. Gleason, E. W. Wright, and D. K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential). The Unconscious Initiation of a Freely Voluntary Act." *Brain : A Journal of Neurology* 106 (Pt 3) (September): 623–42. https://doi.org/10.1093/brain/106.3.623.

Libet, Benjamin, E. Wright, and C. Gleason. 1983. "Preparation – Or Intention-to-Act, in Relation to Pre-Event Potentials Recorded at the Vertex." *Electroenceph. And Clin. Nerophysiology* 56: 367–72.

Lichtenstein, Eli I. 2022. "Revaluing Laws of Nature in Secularized Science." In *Rethinking the Concept of Law of Nature*, 347–77. Springer.

Lowe, Jonathan. 1992. "The Problem Of Psychophysical Causation." *Australasian Journal of Philosophy* 70 (3): 263–76.

———. 1996. *Subjects of Experience*. Cambridge: Cambridge University Press.

———. 2006. "Non-Cartesian Substance Dualism and the Problem of Mental Causation." *Erkenntnis* 65: 5–23.

Lycan, William G. 2018. "Redressing Substance Dualism." In *The Blackwell Companion to Substance Dualism*, edited by Jonathan J. Loose and Angus J.L. Menuge, 22–40. Hoboken, NJ: Wiley Blackwell.

Lyell, Charles. 1889. *Principles of Geology*. 11th ed. London: D. Appleton.

Malebranche, Nicolas. 1980. *Nicolas Malebranche: The Search after Truth*. Translated by T.M. Lennon and P.J. Olscamp. Columbus, OH: Ohio State University Press.

Martin, Charles Burton. 2010. *The Mind in Nature*. OUP Oxford.

Maudlin, Tim. 2007. *The Metaphysics Within Physics*. Oxford: Clarendon Press.

Maudlin, Tim, Elias Okon, and Daniel Sudarsky. 2020. "On the Status of Conservation Laws in Physics: Implications for Semiclassical Gravity." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 69: 67–81.

Maxwell, Nicholas. 2000. "The Mind–Body Problem and Explanatory Dualism." *Philosophy* 75 (1): 49–71. https://doi.org/10.1017/S003181910000005X.

Mcginn, Colin. 2000. *The Mysterious Flame: Conscious Minds in a Material World*. New Ed. New York, NY: Basic Books.

McQueen, Kelvin. 2015. "Four Tails Problems for Dynamical Collapse Theories." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 49: 10–18.

———. n.d. "Consciousness and the Collapse of the Wave Function." https://kelvinmcqueen.files.wordpress.com/2015/08/consc-coll-advcd.pdf.

Meixner, Uwe. 2004. *The Two Sides of Being: A Reassessment of Psycho- Physical Dualism*. Paderborn: Mentis.

———. 2008. "New Perspectives for a Dualistic Conception of Mental Causation." *Journal of*

Consciousness Studies 15 (1): 17–38.

———. 2010. “Materialism Does Not Save the Phenomena - and the Alternative Which Does.” In *The Waning of Materialism*, edited by Robert Koons and George Bealer. Oxford: Oxford University Press.

———. 2019. “Elements of a Theory of Nonphysical Agents in the Physical World.” *Organon F* 26 (1): 104–21. https://doi.org/10.31577/orgf.2019.26107.

Mele, Alfred R. 2014. *Free: Why Science Hasn’t Disproved Free Will*. Oxford, New York: Oxford University Press.

Menuge, Angus. 2009. “Is Downward Causation Possible?: How the Mind Can Make a Physical Difference.” *Philosophia Christi* 11 (1): 93–110.

Meselson, Matthew, and Franklin W. Stahl. 1958. “The Replication of DNA in Escherichia Coli.” *Proceedings of the National Academy of Sciences* 44 (7): 671–82.

Mills, Eugene. 1996. “Interactionism and Overdetermination.” *American Philosophical Quarterly* 33 (1): 105–17.

Milton, John. 2005. *Paradise Lost*. Dover Publications.

Mohrhoff, Ulrich. 1997. “Interactionism, Energy Conservation, and the Violation of Physical Laws.” *Physics Essays* 10 (4).

Molnar, George. 2003. *Powers: A Study in Metaphysics*. Clarendon Press.

Montero, Barbara Gail, and David Papineau. 2016. “Naturalism and Physicalism.” In *The Blackwell Companion to Naturalism*, edited by Kelly James Clark, 1st ed., Ch. 13. John Wiley & Sons.

Moreland, J.P. 2018a. “In Defense of a Thomistic-like Dualism.” In *The Blackwell Companion to Substance Dualism*, edited by Jonathan J. Loose and Angus J.L. Menuge, 102–22. Hoboken, NJ: Wiley Blackwell.

———. 2018b. “Substance Dualism and the Unity of Consciousness.” In *The Blackwell Companion to Substance Dualism*, edited by Jonathan J. Loose and Angus J.L. Menuge. Hoboken, NJ: Wiley Blackwell.

Moreland, J.P., and Scott B. Rae. 2000. *Body & Soul - Human Nature & the Crisis in Ethics*. Downers Grove, Illinois: InterVarsity Press.

Mumford, Stephen. 2004. *Laws in Nature*. Routledge.

———. 2011. “Dispositions.” In *Routledge Encyclopedia of Philosophy*. Routledge. https://www.rep.routledge.com/articles/thematic/dispositions/v-2.

Musso, Mariacristina, Cornelius Weiller, Stefan Kiebel, Stephan P. Müller, Peter Bülau, and Michel Rijntjes. 1999. “Training-Induced Brain Plasticity in Aphasia.” *Brain* 122 (9): 1781–90.

Needham, Joseph. 1951. “Human Laws and Laws of Nature in China and the West (I).” *Journal of the History of Ideas*, 3–30.

Nester, J.M., Lau Loi So, and T. Vargas. 2008. “On the Energy of Homogeneous Cosmologies.” *Physical Review D* 78 (044035).

Nida-Rümelin, Martine. 2010. “An Argument from Transtemporal Identity for Subject Body Dualism.” *The Waning of Materialism*, 191–211.

Noether, E. 1918. “Invariante Variationsprobleme.” *Nachrichten von Der Gesellschaft Der Wissenschaften Zu Göttingen, Mathematisch-Physikalische Klasse*, 235–57.

Numbers, Ronald. 2008. “Science without God: Natural Laws and Christian Beliefs.” In *When Science & Christianity Meet*, edited by D.C. Lindberg and R.L. Numbers, 265–85. Chicago: University of Chicago

Press.

Oakley, Francis. 1961. "Christian Theology and the Newtonian Science: The Rise of the Concept of the Laws of Nature." *Church History* 30 (4): 433–57.

———. 2019. "The Rise of the Concept of Laws of Nature Revisited." *Early Science and Medicine* 24 (1): 1–32.

Ochsner, Kevin N., Silvia A. Bunge, James J. Gross, and John DE Gabrieli. 2002. "Rethinking Feelings: An FMRI Study of the Cognitive Regulation of Emotion." *Journal of Cognitive Neuroscience* 14 (8): 1215–29.

O'Connor, Timothy. 2011. "Agent-Causal Theories Of Freedom." In *The Oxford Handbook Of Free Will*, edited by Robert Kane, 2nd ed. Oxford University Press.

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. 2014. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0." *PLoS Computational Biology* 10 (5): e1003588.

Oldofredi, Andrea, Dustin Lazarovici, Dirk-André Deckert, and Michael Esfeld. 2016. "From the Universe to Subsystems: Why Quantum Mechanics Appears More Stochastic than Classical Mechanics." *Fluctuation and Noise Letters* 15 (03): 1640002.

Papineau, David. 1990. "Why Supervenience?" *Analysis* 50 (2): 66–71.

———. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 3–36. Cambridge: Cambridge University Press.

———. 2009. "The Causal Closure of the Physical and Naturalism." In *The Oxford Handbook of Philosophy of Mind*.

Paquette, Vincent, Johanne Lévesque, Boualem Mensour, Jean-Maxime Leroux, Gilles Beaudoin, Pierre Bourgouin, and Mario Beauregard. 2003. "'Change the Mind and You Change the Brain': Effects of Cognitive-Behavioral Therapy on the Neural Correlates of Spider Phobia." *Neuroimage* 18 (2): 401–9.

Peebles, P.J.E. 1993. *Principles of Physical Cosmology*. Princeton: Princeton University Press.

Peirce, C.S. 1891. "The Architecture of Theories." In *Collected Papers of Charles Sanders Peirce*, VI:6.7-6.34. Cambridge, MA: Harvard University Press.

Penrose, Roger. 1995. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New Ed. London: Vintage Books.

Pessiglione, M. et al. 2006. "Dopamine-Dependent Prediction Errors Underpin Reward-Seeking Behaviour in Humans." *Nature* 442: 1042–45.

Picard, N., and P. L. Strick. 1996. "Motor Areas of the Medial Wall: A Review of Their Location and Functional Activation." *Cereb. Cortex*, no. 6: 342–53.

Pietarinen, Juhani. 2009. "Causal Power in Descartes' Mind-Body Union." In *The World as Active Power*, edited by Juhani Pietarinen and Valtteri Viljanen, 131–62. Leiden/Boston: Brill.

Pietarinen, Juhani, and Valtteri Viljanen. 2009. *The World as Active Power: Studies in the History of European Reason*. Vol. 180. Brill.

Pitts, J. Brian. 2004. "Has Robert Gentry Refuted Big Bang Cosmology? On Energy Conservation and Cosmic Expansion." *Perspectives on Science and Christian Faith* 56 (4): 260–65.

———. 2019a. "Conservation Laws and the Philosophy of Mind: Opening the Black Box, Finding a Mirror." *Philosophia*, 1–35.

———. 2019b. "Mental Causation, Conservation Laws, and General Relativity." *Submitted to European Journal for Philosophy of Science*.

Plantinga, Alvin. 2007. "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter Van Inwagen and Dean Zimmerman, 99–141. Oxford University Press.

———. 2008. "What Is 'Intervention'?" *Theology and Science* 6 (4): 369–401. https://doi.org/10.1080/14746700802396106.

Pollock, John L. 1989. *How to Build a Person: A Prolegomenon*. MIT Press.

Porro, Pasquale. 2014. "Henry of Ghent." In *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/fall2014/entries/henry-ghent.

Powell, Baden. 1838. *The Connexion of Natural and Divine Truth: Or, The Study of the Inductive Philosophy, Considered as Subservient to Theology*. JW Parker.

Rang, H. P. 2003. *Pharmacology*. Edinburgh: Churchill Livingstone.

Robb, David, and John Heil. 2019. "Mental Causation." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2019. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2019/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/.

Roland, P. E. 1981. "Somatotopical Tuning of Postcentral Gyrus during Focal Attention in Man. A Regional Cerebral Blood Flow Study." *Journal of Neurophysiology* 46 (4): 744–54. https://doi.org/10.1152/jn.1981.46.4.744.

Roland, P. E., B. Larsen, N. A. Lassen, and E. Skinhøj. 1980. "Supplementary Motor Area and Other Cortical Areas in Organization of Voluntary Movements in Man." *Journal of Neurophysiology* 43 (1): 118–36. https://doi.org/10.1152/jn.1980.43.1.118.

Ruby, Jane E. 1986. "The Origins of Scientific" Law"." *Journal of the History of Ideas* 47 (3): 341–59.

Ruse, Michael. 1975. "The Relationship between Science and Religion in Britain, 1830–1870." *Church History* 44 (4): 505–22.

Russell, Bertrand. 1912. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 13: 1–26.

Ryle, Gilbert. 1973. *The Concept of Mind*. N. i. edition. Harmondsworth, Eng.: Penguin.

Salari, V., J. Tuszynski, M. Rahnama, and G. Bernroider. 2011. "Plausibility of Quantum Coherent States in Biological Systems." In *Journal of Physics: Conference Series*, 306:012075. IOP Publishing.

Salari, Vahid, Narges Moradi, M. Sajadi, F. Fazileh, and F. Shahbazi. 2015. "Quantum Decoherence Time Scales for Ionic Superposition States in Ion Channels." *Physical Review E* 91 (3): 032704.

Schwartz, Jeffrey M. 1998. "Neuroanatomical Aspects of Cognitive-Behavioural Therapy Response in Obsessive-Compulsive Disorder: An Evolving Perspective on Brain and Behaviour." *The British Journal of Psychiatry* 173 (S35): 38–44.

Schwartz, Jeffrey M., and Sharon Begley. 2009. *The Mind and the Brain*. Springer Science & Business Media.

Schwartz, Jeffrey M., Henry P. Stapp, and Mario Beauregard. 2005. "Quantum Physics in Neuroscience and Psychology: A Neurophysical Model of Mind–brain Interaction." *Philosophical Transactions of the Royal Society Biological Sciences*, no. 360: 1309–27.

Schwartz, Jeffrey M., Paula W. Stoessel, Lewis R. Baxter, Karron M. Martin, and Michael E. Phelps. 1996. “Systematic Changes in Cerebral Glucose Metabolic Rate after Successful Behavior Modification Treatment of Obsessive-Compulsive Disorder.” *Archives of General Psychiatry* 53 (2): 109–13.

Searle, John R. 1984. *Minds, Brains and Science*. Cambridge, MA: Harvard University Press.

———. 2004. *Mind: A Brief Introduction*. Oxford University Press.

Sellars, Wilfrid S. 1962. “Philosophy and the Scientific Image of Man.” In *Science, Perception, and Reality*, edited by Robert Colodny, 35–78. Humanities Press/Ridgeview.

Shapiro, Lisa, ed. 2007. *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*. University of Chicago Press.

Shoemaker, Sydney. 1980. “Causality and Properties.” In *Time and Cause*, 109–35. Springer.

Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. 2008. “Unconscious Determinants of Free Decisions in the Human Brain.” *Nature Neuroscience* 11 (5): 543–45. https://doi.org/10.1038/nn.2112.

Stanley, M. 2019. “God and the Uniformity of Nature: The Case of Nineteenth-Century Physics.” In *Science without God? Rethinking the History of Scientific Naturalism. Ian Ramsey Centre Studies in Science and Religion*, edited by P. Harrison and J.H. Roberts. Oxford: Oxford University Press.

Stapp, Henry P. 1999. “Attention, Intention, and Will in Quantum Physics.” *Journal of Consciousness Studies* 6 (8–9): 143–143.

———. 2000. “The Importance of Quantum Decoherence in Brain Processes.” *ArXiv Preprint Quant-Ph/0010029*.

———. 2004. “Quantum Leaps in Philosophy of Mind: Reply to Bourget’s Critique.” *Journal of Consciousness Studies* 11 (12).

———. 2017. *Quantum Theory and Free Will: How Mental Intentions Translate into Bodily Actions*. 1st ed. 2017. New York, NY: Springer.

Stein, P S G. 1978. “Motor Systems, with Specific Reference to the Control of Locomotion.” *Annual Review of Neuroscience* 1 (1): 61–81. https://doi.org/10.1146/annurev.ne.01.030178.000425.

Stevens, C. F. 1993. “Quantal Release of Neurotransmitter and Long-Term Potentiation.” *Cell* 72 Suppl (January): 55–63. https://doi.org/10.1016/S0092-8674(05)80028-5.

Südhof, Thomas C. 1995. “The Synaptic Vesicle Cycle: A Cascade of Protein–Protein Interactions.” *Nature* 375 (6533): 645–53. https://doi.org/10.1038/375645a0.

Summhammer, Johann, Vahid Salari, and Gustav Bernroider. 2012. “A Quantum-Mechanical Description of Ion Motion within the Confining Potentials of Voltage-Gated Ion Channels.” *Journal of Integrative Neuroscience* 11 (02): 123–35.

Summhammer, Johann, Georg Sulyok, and Gustav Bernroider. 2018. “Quantum Dynamics and Non-Local Effects behind Ion Transition States during Permeation in Membrane Channel Proteins.” *Entropy* 20 (8): 558.

Swinburne, Richard. 1996. “Dualism Intact.” *Faith and Philosophy* 13 (1): 68–77.

———. 1997. *The Evolution of the Soul*. Rev. Oxford: Oxford University Press.

———. 2013. *Mind, Brain, and Free Will*. Oxford: Oxford University Press.

———. 2019. “The Implausibility of the Causal Closure of the Physical.” *Organon F* 26 (1): 25–39.

Swoyer, Chris. 1982. "The Nature of Natural Laws." *Australasian Journal of Philosophy* 60 (3): 203–23.

Taliaferro, Charles. 1994. *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.

Tanji, J., and K. Kurata. 1982. "Comparison of Movement-Related Activity in Two Cortical Motor Areas of Primates." *Journal of Neurophysiology* 48 (3): 633–53. https://doi.org/10.1152/jn.1982.48.3.633.

Tanji, Jun, and Shima Keisetsu. 1994. "Role for Supplementary Motor Area Cells in Planning Several Moevements Ahead." *Nature* 371: 413–16.

Tegmark, Max. 2000. "Importance of Quantum Decoherence in Brain Processes." *Physical Review E* 61 (4): 4194–4206.

Tononi, Giulio, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. "Integrated Information Theory: From Consciousness to Its Physical Substrate." *Nature Reviews Neuroscience* 17 (7): 450–61.

Tsuda, Ichiro, and Hiroshi Fujii. 2007. "Chaos Reality in the Brain." *Journal of Integrative Neuroscience* 6 (02): 309–26.

Tuszynski, Jack A. 2006. *The Emerging Physics of Consciousness*. Springer Science & Business Media.

Ushkaryov, Yuri A., Alexis Rohou, and Shuzo Sugita. 2008. "α-Latrotoxin and Its Receptors." In *Pharmacology of Neurotransmitter Release*, edited by Thomas C. Südhof and Klaus Starke, 171–206. Handbook of Experimental Pharmacology. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74805-2_7.

Van Fraassen, Bas C. 1989. *Laws and Symmetry*. Clarendon Press.

Van Inwagen, Peter. 1983. *An Essay on Free Will*. Clarendon Press.

Varela, F.J. 1997. "The Specious Present: A Neurophenomenology of Time Consciousness." In *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford: Stanford University Press.

Von Neumann, John. 1932. *Mathematische Grundlagen Der Quantenmechanik*. Berlin: Springer.

Von Wachter, Daniel. 2001. "A World of Fields." In *Things, Facts and Events*, 305–25. Brill.

———. 2003. "Free Agents as Cause." In *On Human Persons*, edited by K. Petrus, 183–94. Frankfurt/Lancaster: Ontos Verlag.

———. 2009a. *Die Kausale Struktur Der Welt*. Alber.

———. 2009b. *Die Kausale Struktur Der Welt: Eine Philosophische Untersuchung Über Verursachung, Naturgesetze, Freie Handlungen, Möglichkeit Und Gottes Kausale Rolle in Der Welt*. Alber.

———. 2012. "Kein Gehirnereignis Kann Ein Späteres Festlegen." *Zeitschrift Für Philosophische Forschung* 66 (3): 393–408.

———. 2015. "Miracles Are Not Violations Of The Laws Of Nature Because The Laws Do Not Entail Regularities." *European Journal for Philosophy of Religion* 7 (4): 37–60.

———. 2016. "Readiness Potentials Do Not Cause Our Actions."

———. 2019. "The Principle of the Causal Openness of the Physical." *Organon F* 26 (1): 40–61.

Wald, Robert M. 1984. *General Relativity, Chicago, Usa: Univ*. Pr.

Watkins, Eric. 2013. *The Divine Order, the Human Order, and the Order of Nature: Historical Perspectives*. Oxford University Press.

Westphal, Jonathan. 2016. *The Mind–Body Problem*. Cambridge, Massachusetts: MIT Press.

Whewell, William. 1856. *Astronomy and General Physics: Considered with Reference to Natural Theology*. New York: Harper and Brothers.

Wick, Gian Carlo, Arthur Strong Wightman, and Eugene Paul Wigner. 1997. "The Intrinsic Parity of Elementary Particles." In *Part I: Particles and Fields. Part II: Foundations of Quantum Mechanics*, 102–6. Springer.

Wigner, Eugene. 1967a. "Remarks on the Mind-Body Question." In *Symmetries and Reflections*, 171–84. Bloomington & London: Indiana University Press.

———. 1967b. *Symmetries and Reflections*. Indiana University Press.

Williamson, Timothy. 2013a. "The Unclarity of Naturalism." In *Philosophical Methodology: The Armchair or the Laboratory?*, 36–38. Routledge.

———. 2013b. "What Is Naturalism?" In *Philosophical Methodology: The Armchair or the Laboratory?*, 39–41. Routledge.

Wilson, David. 1999. "Mind–Brain Interaction and Violation of Physical Laws." *Journal of Consciousness Studies* 6 (8–9): 185–200.

Zilsel, Edgar. 1942. "The Genesis of the Concept of Physical Law." *The Philosophical Review* 51 (3): 245–79.